What is a Character Trait?

Natalie Gold  
Duke University  
November 2005

Abstract. Recently, Gilbert Harman and John Doris have used evidence from social psychology to argue that character traits do not exist and hence that virtue ethics is a misguided enterprise. Neither of them analyses what a character trait is, both lean on a dispositional account and claim that observations of behavior are enough to assess the propriety of trait attribution. In this paper I distinguish two functions of trait attribution, a descriptive one and an explanatory one. I argue that, whilst a descriptive attribution of a trait is made purely on the basis of behavior, an explanatory attribution at least implies an underlying causal mechanism involving motivations. Virtue ethics presupposes the propriety of explanatory trait attributions. For reasons I explain, the boundary between the two types of attribution may be fuzzy. However, I argue that it is important to keep the two functions distinct when assessing the existence of character traits and that previous work in philosophy and psychology has not always done this. I show that the evidence from social psychology gives insights into descriptive trait attributions but not into explanatory attributions and hence that the arguments brought to bear have not always had relevance to virtue ethics. I present some evidence from experimental economics that supports the existence of personality differences. I suggest a causal mechanism that may underlie explanatory trait attributions, show how this is supported by behavioral evidence and explore some implications of it for virtue ethics.

Keywords. Character traits, virtue ethics, social psychology, experimental economics

Gilbert Harman and John Doris have argued that character traits, of the sort that are needed to ground virtue ethics, do not exist.¹ A common core to both of their arguments is the idea that a character trait is a disposition and that behavioral evidence from social psychology shows that such dispositions do not exist. Neither of them engages in any further analysis of what a character trait is. Indeed, Doris explicitly claims that such analysis is not necessary. However, the citing of evidence assumes, at least implicitly, a null hypothesis that is being disproved. This leaves both Doris and Harman open to the charge that the theory of traits that is implicit in their critique does not actually capture what is meant by a character trait. Many people have supplied responses to their arguments that basically take this form.² In this paper, I take a more from the ground-up approach, asking what exactly a character trait is and considering what the evidence does show.

I will argue that trait attribution serves two different functions: a descriptive function, summarizing past behavior, and an explanatory function, giving a causal explanation of behavior. A pre-requisite for virtue ethics is that character traits of the explanatory sort exist. One expectation is that these would show up in the form of individual differences in behavior. I will argue that, by its nature, the psychology evidence that is cited in this debate conflates the two functions. Further, there is a evidence from experimental economics that seems to support the existence of character traits. This leads to the questions of what is a character trait and what is shown by each of these two bodies of evidence.

Building on Aristotle’s definition of virtue as “a state concerned with choice”³ and work by modern virtue ethicists which emphasizes that virtue consists in acting for reasons, I look for causes of individual differences in the decision making process. I identify two loci in the decision making process where individuals might differ. I argue that the reason for the success of two of the personality instruments used by economists to predict behavior is that they each

---


capture a single one of these, and that they are then tested in situations where behavior can be expected to differ along this locus. If this is correct, then they provide evidence for the existence of individual differences in the form of character, and also for what these traits consist of. I discuss some of the implications for character traits and for virtue ethics.

In section 1, I present Doris and Harman’s arguments against the existence of character traits. In section 2, I distinguish two functions of trait attribution. In section 3, I argue that the psychological evidence conflates these two functions. In section 4, I present evidence from laboratory experiments in economics that personality measures do correlate with behavior. In section 5, I offer an explanatory account of character traits that is consistent with this evidence. In section 6, I develop some of the implications for character traits and virtue ethics.

1. Lack of Analysis: Dispositions, Traits and Behavior

It is standard to assume that at least part of what it is to have a character trait is to have a behavioral disposition. Both Harman and Doris lean heavily on this idea. Harman defines a character trait as a “relatively stable and long-term disposition to act in distinctive ways”.5 Doris goes through various possible formulations before settling on: “If a person possesses a trait, that person will engage in trait-relevant behaviors in trait-relevant eliciting conditions with markedly above chance probability p”.6 Both present experimental evidence and argue that it shows that behavioral variation is caused by situational differences rather than behavioral dispositions that differ between persons, but they each use the evidence to develop rather different attacks on the existence of character traits. It will be useful to briefly sketch the form that each of their arguments takes. Their implicit assumptions about character traits turn out to be similar to the commitments made by some psychologists and, in the next section, I will go on to show that these confound two different functions of trait attributions. I won’t rehearse the psychological evidence they cite in great detail or the criticisms of the way the data was gathered. I do not want to haggle over data or interpretations here.5 My interest is in conceptual analysis.

Harman and Doris both draw on a body of experiments from social psychology on the situational determinants of behavior. Psychology is traditionally the study of individual differences but in the twentieth century experimental evidence emerged that situations were more important and individual differences less important than psychologists (and folk psychology) had assumed. There was a shift from studying the relation between individual differences and behavior, to investigating the common effects of situations on behaviour and there was a backlash against personality psychology in the form of situationism, the view that human behavior is largely determined by the characteristics of the situation rather than personal predispositions.7

Doris rests his case mainly on evidence about behavioral consistency. He defines a behavioral conception of character that is at issue, which he calls “globalism”, and which he says maintains the following thesis, called consistency: “Character and personality traits are reliably manifested in trait-relevant behavior across a diversity of trait-relevant eliciting conditions that may vary widely in their conduciveness to the manifestation of the trait in question.”8 He then makes a case that experimental evidence shows that this consistency does not exist and that the globalist conception of character is therefore incorrect.

Doris gives most attention to Hartshorne and May’s 1928 study of deceptive behavior in children.9 Hartshorne and May went into schools and set up situations where the children might be tempted to cheat, lie or steal. They scored the children according to their behavior in each situation and calculated the correlation coefficient between the different situations. For instance, they

---

4 Harman “Moral Philosophy Meets Social Psychology” p.2
5 Doris Lack of Character p.19, italics in original.
6 I have gone into further detail elsewhere, in Framing and Decision Making: A reason-based approach (Oxford: Unpublished D.Phil, 2005), as have other philosophers (see references in footnote 2).
observed how much children would cheat on a variety of tests. In their copying tests the children wrote answers to questions that were collected, copies of the paper were made and then the papers returned to pupils who were each allowed to correct their own. A child’s cheating score consisted of the number of changes made on her paper. On the speed and athletic tests, the children were given two or three “practice” or “warm-up” tests, whose purpose was in fact to record the child’s performance. Children were then left to record their own scores on a final test. The cheating score was defined as the amount of increase in score on the last trial. Hartshorne and May found that cross-situational consistency, or the mean correlation between different pairs of types of situations presenting opportunities for deception or honesty was 0.23. This figure was considered by psychologists at the time to be low. This contrasts with the correlations between behavior in the same type of situation at different times, known as temporal consistency, which are undisputedly high. For instance, in Hartshorne and May, copying from an answer key on one test had a correlation of 0.696 with copying from a key on another test. This leads Doris to conclude that there are only very local, situation specific traits and that these “narrow” character traits cannot support virtue ethics.

One might reconstruct the logical form of Doris’s argument as follows: Premise 1 (consistency): if traits exist then individuals will display cross-situationally consistent behavior Premise 2 (psychological evidence): there is no cross-situationally consistent behavior Conclusion: traits do not exist.

Doris concentrates his argument on the second premise, that there is no cross-situationally consistent behavior, but he does not give us any reason to accept the first premise, which is an implicit part of the valid argument for his conclusion. His rebuttal of globalism could be construed in two ways: either it could be an argument against character traits per se, or it could be an argument against some person or group of people’s conception of a character trait. If it is supposed to be an argument about character traits per se, then it requires an analysis of character traits that supports the global consistency premise. Otherwise the evidence cited may not be relevant. If it is an argument against some people’s conception of character, then we need to identify the butt of the argument: some person or group of persons who maintains the thesis of consistency as part of their conception of character.

Doris claims that characterological moral psychology and personality psychology are “typically committed” to globalism. I think that it is fair to accuse the psychologists who collected some of the data Doris cites as making a globalist assumption. Looking back on this work, psychologists Jack Wright and Walter Mischel characterize Hartshorne and May (and T. M. Newcomb who ran a similar sort of study at about that time)14) as having expected that “traits would produce rather broad consistencies in social behavior across situations whose range and limits remained unclear”. However, Doris wants to claim that his argument shows the pointlessness of virtue ethics so he needs to be rebutting a conception of character that is prevalent amongst virtue ethicists, and he does not cite any philosopher who has argued that global consistency is part of what it is to have a character trait. Indeed, in response to Doris’s argument, some virtue ethicists have explicitly claimed that globalism is not part of their conception of character. Another possible butt of Doris’s argument is the ordinary folk psychological conception of a character trait. I will go on to argue that globalism is not the folk psychological conception of a character trait either.

In Harman’s original paper on character traits, he concentrates wholly on evidence that people make the fundamental attribution error, underestimating the effect of situational variables on behavior. The two experiments he emphasizes are Stanley Milgram’s investigations into obedience to authority and John Darley and C. Daniel Batson’s into helping behavior. Milgram told subjects that they were participating in an experimental investigation of the effects of punishment on learning. Their task was to administer electric shocks to a learner (who they were told was another volunteer but who was, in fact, a confederate of the experimenter), whenever that person made an error in a word pairing task. At each wrong answer, the intensity of the putative electric shocks increased and, as the shocks got higher, the learner appeared to be in more and more distress. If the subject wanted to stop the experiment, the experimenter had otherwise the evidence cited may not be relevant. If it is an argument against 10 Harshorne and May’s measures are not beyond criticism. On this see Gold Framing and Decision-Making. 11 Walter Mischel gives the classic statement of this in Personality and Assessment (New York: Wiley, 1968). Whether or not it is low is debatable. See Gold’s Framing and Decision-Making and the references therein and Sabini and Silver’s “Lack of Character? Situationism Critiqued”. 12 These are measured by correlations between two single observations, rather than two composite measures. Correlating single rather than aggregate observations should get smaller coefficients, a fact that has had some attention in the debate over cross-situational coefficients (e.g. Seymour Epstein “The Stability of Behavior: I. On Predicting Most of the People Much of the Time” Journal of Personality and Social Psychology 57 (1979): 1097-1126), so the high inter-temporal coefficients are particularly noteworthy. 13 Lack of Character, p.23 14 T. M Newcomb Consistency of certain extrovert-introvert behavior patterns in 51 problem boys (New York: Columbia University, Teachers College, Bureau of Publications, 1929). 15 Jack Wright and Walter Mischel “A Conditional Approach to Dispositional Constructs: The Local Predictability of Social Behavior” Journal of Personality and Social Psychology 53(6) (1987): 1159-77, p.1160. 16 For instance Rachana Kamtekar “Situationism and Virtue Ethics on the Content of Our Character”. 17 “Behavioral Study of Obedience” Journal of Abnormal and Social Psychology 67 (1963): 371-378. 18 “From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior” Journal of Personality and Social Psychology, 27 (1973): 100-108.
continue: “Please continue”, “The experiment requires that you continue”, “It is absolutely essential that you continue”, and “You have no choice; you must go on”. After these had been exhausted, if the subject persisted in wanting to stop then he or she was excused. The device used to administer the “shocks” had intensity labels and Milgram had expected that few subjects would go beyond the designation “Very Strong Shock” (150 volts). Not only did most subjects go past there, a majority went through “Intense Shock”, “Extreme Intensity Shock” and “Danger: Severe Shock” to “XXX” (450 volts). Milgram ran several versions of his experiment and there was variation in behavior depending on the details of the scenario but, in this basic version, 68% of subjects followed through to the end. Milgram had underestimated the pressures of the situational factors in his experiment on subjects’ behavior. He was not alone in this; when he asked people to predict how far subjects would go, they also underestimated the extent of subjects’ obedience.21

Darley and Batson told their subjects, who were students at Princeton Theological Seminary, that they were participating in a study of the vocational careers of seminary students. When subjects arrived at the experiment, they were told that they were scheduled to give a talk in another building. On the way to that building, the experimenters had positioned a confederate slumped, coughing and groaning in a doorway. They investigated what factors were important in determining who stopped to help. They found that neither subjects’ attitudes to their religion (as measured by answers to a pre-experimental questionnaire) nor the subject that they were scheduled to talk about (which was either the Good Samaritan parable or vocational alternatives to the ministry) was significantly related to helping behavior. The only variable they tested that made a significant difference was how much of a hurry the subjects were in. Some subjects were told they were late for their talk, some that they had just enough time to get there and some that they were early. Only 10% of subjects who were late stopped, compared to 45% of those who were on time and 63% of those that were early. Harman concludes that standard versions of the Good Samaritan parable overlook situational factors.

The results were not just surprising to the experimenters. In a further experiment, when a similar Good Samaritan situation was described to subjects and they were asked to predict the experimental results, subjects thought that students’ religious orientations would affect how likely they were to stop and that whether they were in a hurry would not make a difference.20 These subjects committed the fundamental attribution error. Harman says that: “Since it is possible to explain our ordinary belief in character traits as deriving from certain illusions, we must conclude that there is no empirical basis for the existence of character traits.”21

We can state Harman’s argument in the form of two premises and a conclusion. Premise 1 (attribution): If traits exist, then we attribute them correctly Premise 2 (evidence): we make the fundamental attribution error Conclusion: traits do not exist.

Harman never states the first premise but he relies on it in order to derive the conclusion that the fundamental attribution error shows that traits do not exist. However it is not clear why we should accept this premise. The fact that people make errors is not usually held to show that underlying variables do not exist. For instance, people are also subject to a false consensus effect,22 overestimating the degree to which their own behavior, attitudes and beliefs are shared by other people, but no-one has ever to my knowledge used this to argue that no consensus exists.

Maybe some people do think that the incorrect attribution of traits imply their non-existence. As we will see below, there are some psychologists who do rely on folk psychological attributions being correct and could be identified as the butt of this argument. Harman also says that his argument is directed at “ordinary character traits of the sort people think there are”23 and “character traits as ordinarily conceived”.24 It may be that we should take the folk psychological conception of character to be the butt of his argument and replace premise 1 with premise 1′, “If character traits as ordinarily conceived exist, then we attribute them correctly”, and the conclusion with conclusion′ “character traits as ordinarily conceived so not exist”. However, Harman explicitly targets his argument against virtue ethics so, if we grant him this move, he relies on another implicit argument (equivalent to the original premise 1) that the incorrectness of the folk psychological conception of character trait has implications for virtue ethics. I will argue that it does not.25

---

23 Harman “Moral Philosophy meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error” p.316
25 We should also note that, even in the experiments that Harman cites, there were individual differences in behavior. Not all subjects went to 450 volts in Milgram’s study and, in Darley and Batson’s study, if 10% of those in a hurry stopped to help then they behaved differently from the 90% of people in that situation who did not stop. Harman in “Virtue Ethics Without Character Traits” in A. Byrne, R. Stalnaker, and R. Wedgewood, eds., Fact and Value (Cambridge, MA: MIT Press, 2001), makes the further claim that “What a person with a seemingly ideal moral character will do in a particular situation is pretty much what anyone else will do in exactly that situation, allowing for random variation.” However, he does not cite any evidence to support this claim and, in
Whilst Doris does not share Harman’s focus on the fundamental attribution error, he also cites these experiments in making his case.26 Given that evidence of mistakes in trait attribution does not necessarily have any implications for the existence of traits, it is interesting that they both find it to be relevant. Their common starting point, the dispositional analysis of traits, obscures the fact that there are two different functions of trait attribution. These are often run together in the psychological literature, but only one of them has implications for virtue ethics.

2. Two Functions of Trait Attributions

In the analysis of character, it is standard to start with the idea that a trait is a disposition.27 Many things have been characterized as dispositions, including beliefs,28 power,29 and rule following,30 as well as chemical properties such as solubility or fragility. One thing all dispositions have in common is that they can be characterized by a set of counter-factual claims about their manifestation. However, the things on this list are very different types of beast. One dimension on which they differ is the mechanism by which the observed behavior is manifested. The idea that appeal to a disposition in a causal explanation must involve more than just a set of counter-factual claims is widely, though not universally held.31 It is important for virtue ethics that character exists in this causal explanatory sense.

the areas of the two experiments that he cites in “Moral Philosophy Meets Social Psychology” research has shown that there are individual differences that are consistent across situations. On obedience to authority see for instance A. Elms and Stanley Milgram “Personality Characteristics Associated with Obedience and Defiance toward Authoritative Command” Journal of Experimental Research in Personality 2 (1966): 282-89; John Sabini and Maury Silver “Dispositional vs Situational Interpretations of Milgram Obedience Experiments: The fundamental attributional error” Journal for the Theory of Social Behaviour 13, (1983): 147-154; T. Blass “Understanding Behavior in the Milgram Obedience Experiment: The role of personality, situations, and their interactions” Journal of Personality and Social Psychology 60, (1991): 398-413; R. B., Cialdini, and N. Goldstein “Social Influence: Compliance and conformity” Annual Review of Psychology 55, (2004): 591-621. Some of the evidence relevant to helping behavior is cited in the discussion of social value orientation below. This further statement of Harman’s also conflicts with his conclusion in “Moral Philosophy meets Social Psychology” that people may “differ in their goals, strategies, neuroses, optimism etc.”, which one presumes are non-random sources of behavioral difference, even if they are not differences that Harman claims would be recognized as a character traits.

26 Lack of Character ch.3.


29 Peter Morris Power: a philosophical analysis (Manchester: Manchester University Press, 1987)


31 For example: John Mackie Truth Probability and Paradox; Robert Shope “The Conditional Fallacy in Contemporary Philosophy” The Journal of Philosophy 75 (1978): 397-413; Elizabeth Traits differ from the other examples of dispositional properties in a further manner. As well as this explanatory function, as a causal explanation of behavior, trait attribution can also have a descriptive function, conveying information about how a person has acted in the past.32 These are both encompassed in a dispositional analysis as both support inferences about behavior in counter-factual situations. The inferences that are made on the basis of a descriptive attribution will be inductive whilst those made on the basis of a causal explanation may be deductive. A crucial difference between the two is that a descriptive trait attribution implies no particular mechanism by which the behavior was caused, whereas an explanatory attribution carries a commitment to the existence of some causal mechanism. Another distinction is that, whilst analysis of the folks’ explanatory attribution does not necessarily tell us how a trait causes behavior, analysis of how the folk descriptively attribute traits is also analysis of what a descriptive attribution is.

If we do not maintain the distinction between the descriptive and explanatory function of traits, we risk circularity in our trait-explanations of behavior, analogous to the mistake made by the doctor in Moliere’s La Malade Imaginaire who explains that the reason opium puts people to sleep is because of its dormative powers, rather than citing a causal mechanism such as the chemical properties of opium and their effect on the brain. As I will explain in the next section, in folk psychological trait attribution the boundary between the two functions is fuzzy. It is natural for people to start out making descriptive attributions and, from there, to move to explanatory attributions. This may be a valid way of developing an explanatory attribution, as descriptive data can have a role to play in the development and application of a causal theory. However, it means that when psychologists base their personality theories on character traits as they are ordinarily used, they do not maintain the distinction either.

For instance, take the dominant “five factor model” of personality.33 According to the model, there are five basic dimensions of personality: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. People are scored on each of these according to their answers to a pen and paper questionnaire about their behaviors, beliefs and attitudes. One can criticize the classification for relying on people’s self-descriptions and one can also dispute


32 I do not claim that these categories are exhaustive. For instance once we attribute one trait to a person, we might use this to infer other traits that we can attribute to her, provided we have a suitable theory of the relation between traits. Another possible function of character traits is to attribute praise and blame.

whether it is a good predictor of behavior. However, we might wonder why anyone would expect the results of the test to correlate with behavior in the first place. The big five factors were discovered by analyzing the language of personality. Subjects were asked to rate themselves on a large number of trait terms. Then the ratings were “factor analysed” to see if the multiple terms clustered on a smaller number of dimensions. So the “big five” are the five dimensions that underlie how we talk about character traits. Their origins are in the folk psychological use of trait terms.

To move from the observation that the big five are the five dimensions that underlie our descriptive attributions to the claim that they are also the explanatory five dimensions that underlie behavior we need an additional premise, which Raymond Cattell called the sedimentation hypothesis, that all important aspects of human psychology become sedimented into language as trait-descriptive adjectives. The psychologists who propose the five factor model are aware of the reliance of their work on the sedimentation hypothesis, which they accept. If the sedimentation hypothesis is true, then we might expect that the five factor classification, based on the linguistic analysis of traits, would also correlate with behavior. The sedimentation hypothesis implies that folk psychological trait attributions track explanatory variables. If it is correct, then it would be appropriate to measure character traits with scales that have a descriptive basis, without reference to their causes. Frankly, I think the sedimentation hypothesis is implausible. Further, I think there is good evidence about how the folk make trait attributions, evidence which implies that they do not necessarily track the causes of individual differences in behavior. In the next section I will argue that, because much of the work in personality psychology relies on folk psychological conceptions of character, without differentiating between descriptive and explanatory trait attributions, the evidence does not implicate explanatory trait attribution and, hence, is not relevant for virtue ethics.

34 Both of which are done by Doris in Lack of Character p.67-71.
35 To be precise, factor analysis takes an n-dimensional vector and seeks an m-dimensional vector of common factors, where m < n, such that each of the n entries in the original vector can be reconstructed by an appropriate linear combination of the m common factors.
37 For example Lewis Goldberg in “Language and individual differences: The search for universals in personality lexicons” in L. Wheeler (Ed.), Review of Personality and Social Psychology: Vol. 2 (Beverly Hills, CA: Sage, 1981) or “From Ace to Zombie: Some explorations in the language of personality” in C. D. Spielberger and J. N. Butcher (eds.) Advances in Personality Assessment Vol. 1 (Hillsdale: New Jersey: Erlbaum, 1982). Which is not to claim that all psychologists accept it, or think that the five factor model is a good one. For example, S. Briggs “The optimal level of measurement for personality constructs” in David Buss and Nancy Cantor (Eds). Personality psychology: recent trends and emerging directions (NY: Springer-Verlag, 1989) complains that the original studies leading to the five-factor model “prompted no a priori predictions as to what factors should emerge, and a coherent and falsifiable explanation for the five factors has yet to be put forward” p.249.

38 Hampshire, in “Dispositions”, implies that this is the only function of a character trait and, hence, he argues that character traits should not be considered as dispositions because they are unlike the standard chemical examples of dispositional properties.
40 Though note that the prototype is not supposed to refer to a specific category member.
and other types of owls. Similarly, bird is a member of the super-ordinate category of living things. The more abstract a category, the more objects it includes. We can construct a taxonomy that relates categories to each other by means of class inclusion, as in figure 1. As class inclusion increases (i.e. at “higher”, more abstract levels), the less properties the class members will have in common. As class inclusion decreases (i.e. at “lower”, more specific levels), the more the attributes of class members are shared with members of other categories. People show a preference for a basic level that may be considered to provide an optimal trade-off between the similarity within the category and distinctness from other categories. In this case, the basic level is “bird”. However, the category level used will depend on the context. An ornithologist might use a sub-ordinate category, as would someone standing in a pet shop or a butcher’s shop with the intention of buying one of the birds on display.

![Figure 1: Categorization of Birds](image)

A lot of current psychological research on character is at the inter-face of personality psychology and cognition, and the interplay between character traits and categorization theory plays a part in a variety of psychological approaches. Of particular interest here is the work of Sarah Hampson and her colleagues, which investigates the parallels between trait use and other categorizations. People are attributed traits on the basis of their behaviors, so we can set up similar taxonomies for character traits with behaviors at the lowest level, working up through increasingly broader traits at higher levels. In experiments, Hampson and others have identified a basic level that is generally preferred when describing friends and acquaintances. For instance, in the four level hierarchy charitable-generous-kind-good, they found kind to be basic. In the hierarchy tactful-polite-considerate-nice, considerate is preferred. The experimenters also tested the trait opposites, stingy-selfish-unkind-bad and tactless-impolite-inconsiderate-dislikeable and found the same level preferred. But the level used depends on context and background. As people get to know each other, their use of simple behavior terms decreases and their use of trait terms shows a corresponding increase. They move to more abstract categorizations.

42 The basic level can be considered to be a special case of this, as the level of abstraction that is appropriate in most situations.
45 We might wonder if “good” is a character trait, rather than a character evaluation, but that does not affect the argument here.
It is easy to see that the behaviors observed in the classic Hartshorne and May study of dishonesty would fit neatly into the sort of taxonomy that we should expect on the basis of cognitive psychology, as represented in figure 2. For instance, cheating on a test and cheating on the athletics field are both types of cheating, and cheating is a type of dishonesty. There are three different possible levels of categorization here. Other intermediate-level traits Hartshorne and May were interested in were lying and stealing.

If traits are attributed descriptively in this manner, what implications do they have for the consistency of behavior? Following categorization theory, we might expect that someone who is attributed a trait would be more likely to exhibit those behaviors that are most prototypical of it. However, if trait attribution is based on the performance of prototypical behaviors, no particular behavior is necessary or sufficient and we cannot infer (as globalism does) than someone who is attributed a basic level trait will have performed all the behaviors in the sub-ordinate categories. Hence we should not expect global consistency between behaviors, though there might be some consistency between prototypical behaviors.

The idea that there is consistency between prototypical behaviors arguably gets support from Walter Mischel and P. K. Peake’s study on conscientiousness amongst students at Carleton college. They monitored nineteen behaviors that were supplied by the students themselves in pre-tests in order “to obtain referents for the trait constructs as perceived by the students” and “in contrast to many studies in which the references are selected exclusively by the assessors”. These behaviors included: class attendance, study-session attendance, assignment neatness, assignment punctuality, room neatness, reserve-reading punctuality for course sessions and personal-appearance neatness. For each measure, repeated observations (ranging in number from two to twelve) were obtained and the observations were aggregated into a composite measure for each subject on each behavior. The experimenters did not find global consistency. However, behaviors that we might consider to be more prototypical of conscientiousness did show cross-situational consistency with each other. If we start with the premise that a conscientious student will tend to turn up to class, we can note that class attendance was significantly positively correlated (correlation coefficients in brackets) with: appointment attendance (0.67), assignment punctuality (0.53), completion of class readings (0.58) and amount of time spent studying (0.31). This compares to the lack of any correlation between class attendance and: class-note thoroughness, punctuality to lectures and assignment neatness. The latter three, whilst not unconnected to conscientiousness, are arguably more peripheral behavioral manifestations of the trait.

Traditionally, personality psychologists assumed that the way we talk about traits also tracks the relevant underlying individual differences that cause behaviour: they accepted the sedimentation hypothesis. In the experiments on cross-situational consistency, the researchers used their (or, in Mischel and Peake’s case, their subjects’) folk psychological understanding of trait

---

47 Incidentally, since honesty is not one of the “big five” factors underlying behaviour, by the lights of current personality psychology we should not expect to find cross-situational consistency in Hartshorne and May’s study.

48 Walter Mischel and P. K. Peake “Beyond Déjà Vu in the Search for Cross-Situational Consistency” Psychological Review 89 (1982): 730-755. This was specifically designed to avoid some of the criticisms made of Hartshorne and May.

49 Ibid., p.734. The authors do not say exactly how they obtained the behaviors to be monitored in the pre-tests or exactly how they measured them.

50 They report that, of the 171 correlation coefficients between the composite behavioral measures, 20% were significant at the 5% level. The average correlation coefficient was 0.13. (But note that the authors report the average over all the coefficients, including the non-significant ones, and the full set of coefficients is not given, so it is not possible to perform other calculations using the data.) The authors conclude that “broad cross-situational consistencies remain elusive” (p.735).

51 The classic statement of this is provided by Gordon Allport in Personality (London: Constable and Company Ltd, 1934).
attribution in order to choose the behaviors that were monitored, so it should not be surprising to find that their results track descriptive trait attributions.

Some psychologists now think that the only function of traits is the descriptive one of categorizing behavior. For example, Hampson asserts that personality is entirely socially constructed and that traits only exist between rather than within people, implying that there is no explanatory function of traits. However, her own theory gives us grounds to suspect that this is not the case. Despite the analogy to the categorization of objects, there are two key differences. Firstly, in principle categories are governed by a strict class inclusion relation, that each category in the hierarchy is entirely included in one and only one other category (unless it is the highest level category). In other words, once we know that an owl is a type of bird we know that all types of owls are types of bird and that owl does not belong to any other category at the basic level, for instance it is not a type of fish. Strict class inclusion does not hold for traits. It holds neither “horizontally” nor “vertically” in the taxonomy. Horizontally, some behaviors may belong in two categories, for instance getting one’s friend out of trouble by telling a lie is both a deceptive and a loyal behavior. This plays back into vertical inclusion; someone who tells a lie to protect a friend may be deceptive or loyal or both. The possibility that they are only deceptive or only loyal violates vertical strict class inclusion.

The violation of strict class inclusion should make higher level categories less useful for predictions of behavior. This is where the second disanalogy comes in. In the categorization of objects, experts tend to use narrower, sub-ordinate classifications rather than the usual basic level. However, with character traits, as people become “experts” in each other, they move to more abstract categorizations, from the sub-ordinate level of behaviors to the basic level. I think that this is related to folk psychological explanatory attributions. At the very lowest level of the trait taxonomies are simple behaviours. Not only do higher categories encompass a variety of behaviors, they also introduce a new feature: they imply that the behavior was caused by a particular type of motivation. A motivation is a causal factor in behavior, so use of a higher level trait category may also be an explanatory attribution.

In folk psychology, there is no clear boundary between the descriptive and explanatory functions of traits. Descriptive summaries of past behavior provide the data from which an explanatory attribution is made. Explanatory attribution is made difficult because of the violation of strict class inclusion; any one behavior could have been produced by a multitude of motivations. This would explain why people move upwards from simple behaviors to the basic level. When they see a target performing a variety of behaviors that might have a common motivational cause, they begin to attribute that motivation. However, her results track descriptive trait attributions. A folk psychological trait attribution might be either descriptive or explanatory.

Can we expect descriptive attributions to track causal explanations of behavior? The prima facie answer is surely no. If descriptive attributions are pre-theoretical, then there is no reason to expect them to track the underlying variables of the theory. (This is granting that a folk psychological theory where character consists of acting for motivations is correct. I think it is near enough, as will become clear below.) On my account, they are made on the basis of similarities and categorization, but this is not enough for an explanatory attribution. If we base the academic study of character on folk psychology, we are on shaky ground.

4. Moral Philosophy Meets Experimental Economics

51 (1986) 37-54 use the “X is a type of Y” test, modified to “X is a way of Y” and found that the strict asymmetry observed by Rosch in “Cognitive Reference Points” does not hold. But, even with objects, class inclusion may not be strict. Consider Duchamp’s “Fountain” where, by taking a urinal out of its usual functional context, he made it into a work of art. Does this violate strict class inclusion? Possibly not, as cognitive psychologists allow that concepts can be included in multiple categories in different hierarchies and that the hierarchy used depends on the domain of study. Maybe they would view Duchamp’s urinal as an object that can be placed in two different hierarchies, rather than as a violation of strict class inclusion.

52 Sarah Hampson “The Social Construction of Personality” in Han Bonarious, Guus Van Heck and Nico Smid (eds.) Personality Psychology in Europe: Theoretical and Empirical Developments Lisse: Swets and Zeitlinger 1982. Act frequentists and proponents of the social cognitivist approach also acknowledge that their theories have no causal status – see Jack Wright and Walter Mischel “A Conditional Approach to Dispositional Constructs: The Local Predictability of Social Behavior”. (Unlike Hampson’s claim, this position does not necessarily rule out there being some causal explanatory traits that do not track the theory, though it may be the case that these psychologists believe that traits can never serve an explanatory function.)

53 Eleanor Rosch et al “Basic Objects in Natural Categories”.

54 Sarah Hampson, Oliver John and Lewis Goldberg “Category Breadth and Hierarchical Structure in Personality: Studies of Asymmetries in Judgements of Trait Implications” Journal of Personality and Social Psychology 51 (1986) 37-54 use the “X is a type of Y” test, modified to “X is a way of Y” and found that the strict asymmetry observed by Rosch in “Cognitive Reference Points” does not hold. But, even with objects, class inclusion may not be strict. Consider Duchamp’s “Fountain” where, by taking a urinal out of its usual functional context, he made it into a work of art. Does this violate strict class inclusion? Possibly not, as cognitive psychologists allow that concepts can be included in multiple categories in different hierarchies and that the hierarchy used depends on the domain of study. Maybe they would view Duchamp’s urinal as an object that can be placed in two different hierarchies, rather than as a violation of strict class inclusion.

55 My account of how the folk move from descriptive to explanatory trait attributions bears some similarity to the speculative account given by Adam Morton in Frames of Mind (Clarendon Press: Oxford, 1980) of how a child may develop an understanding of character traits.

56 Another factor which might promote explanatory trait attribution is observations of behaviour in situations where motivation can be unambiguously inferred. Each of explanatory attributions based on variety of behaviors and attributions based on unambiguous behaviors may be more relevant for different traits. For instance courageous behaviors may be rarely seen but, when they are, the motivation is obvious, whereas honesty may be displayed often but be very ambiguous. Further, we might speculate that unambiguous behavior is also that which is most prototypical, and that the more peripheral manifestations of the trait are those that may have a variety of motivations and hence be more ambiguous.
The situationist view that Harman and Doris attribute to psychologists is a staple of much traditional economics, where the representative agent paradigm dominates: everyone is assumed to have the same utility function (i.e. same preferences) and differences in behavior are caused by differences in agents’ budget constraints (i.e. different situations). Nobel Laureates George Stigler and Gary Becker notoriously argued that economic explanations that appeal to differences in the utility function are “ad hoc arguments that disguise analytical failures”.57 However, since Stigler and Becker wrote in 1977, experimental economics has burgeoned. Researchers regularly find that subjects who are put in the same experimental situation exhibit different behavior, and this has prompted a nascent research program into the relationship between individual differences and behavior.

In the psychology experiments that are cited in the personality debate, subjects’ behavior is observed in a naturalistic setting. In contrast, the experimental economics paradigm uses a laboratory setting where situations are stripped down to isolate particular strategic features of real-world situations.58 These experiments are very carefully controlled so that subjects are all placed in the same situation. Then, in different conditions of the experiment, a parameter can be altered in the knowledge that the only situational variable that differs between groups is the variable that is being tested. However, researchers regularly find that there are differences in behavior within groups, differences between subjects who are placed in the same condition.59 This has led behavioral economists to formulate and test theoretical models that include heterogeneous agents.60 Recently, there have been some experiments that investigate the relation between typologies taken from personality psychology and individual differences in behavior in experimental games. Experimenters have concluded that this work “clearly demonstrates that the personality of the player matters” and that “future research should therefore focus on differences instead of similarities between individuals in order to increase our understanding of real-life.”61

In this section I present two types of experiment that are run in economics: public goods games and trust games. Each takes a strategic situation from game theory, with the outcomes in the payoff matrix realized in terms of money. Each is commonly thought to have the potential for eliciting trait-relevant behavior: cooperativeness in public goods games and trustworthiness in trust games. In each of these types of experiment, a relationship has been found between subjects’ behavior and a their classification using a personality typology taken from the psychological literature. This is suggestive that some individual differences in character which would serve an explanatory function do exist. In the next section, I will offer a theory of how individual differences in decision making lead to different behavior that would explain why the particular classifications are successful in these games.

### a. Cooperativeness: Social Value Orientations and Contributions in Public Goods Games

Public goods have benefits that accrue to all. If they are provided privately, the costs accrue to the individual. Standard examples in economics are national defense, lighthouses and public broadcasting. These things are often provided by the state as, left to themselves, individuals would have no incentive to pay for them. If the benefits are sufficiently dispersed then, for any individual, the amount of benefit she gets from contributing is outweighed by the costs. However, if every individual contributed, each would be better off than if no-one contributed. In its two person form, a public goods game is the notorious prisoner’s dilemma, and contributing is equivalent to making the so-called cooperative move.62

Experimental public goods games operationalize this strategic structure. Each subject is given an endowment of money. She must then choose whether or not to make a contribution to a group project (which could be a binary choice to contribute her endowment or not, or a level of contribution out of her endowment). This contribution is multiplied up by some factor $k$ and shared out equally between everyone in the group. The factor $k$ is chosen such that, with $n$ players,

\[ k \gt 1 \]


58 It may be wrong to think of this as a distinction between economics and psychology per se. The contrast I made between the traditional concerns of personality psychology being individual differences, and the concern of economics being situations, could been also been drawn as a distinction between personality psychology and social psychology. Although there are differences between the experimental protocols, as detailed by Ralph Hertwig and Andreas Ortmann “Experimental Practices in Economics. A Methodological Challenge for Psychologists” *Behavioral and Brain Sciences* (2001) 24: 383-451, there is a tradition in social psychology of running laboratory experiments that pre-dates that in economics and the literatures do intersect, for instance in the study of public goods games. But nothing of importance hangs on exactly how distinct the distinction is drawn.


60 For instance Ernst Fehr and K Schmidt’s “A Theory of Fairness, Competition, and Cooperation” *Quarterly Journal of Economics* 114(3) (1999), 817-68.


62 With “step level” public goods, such as lighthouses, which are either provided or not, then the circumstance of interest is the individual’s probability of being the marginal contributor who enables the public good to be provided. If there are already enough contributions without her, it will be provided whether she contributes or not. If there are not enough contributions even including hers, it will not be provided despite her contribution. The player’s expected benefit from contributing is thus the probability of her contribution being necessary to the good’s provision multiplied by her benefit from getting the good. If this is less than the cost, then her incentive is not to contribute.
subjects, $k/n$ is less than one, so the benefit received is less than the cost of contribution and hence the individual incentive is not to contribute. For instance if there are four subjects ($n = 4$) and each contribution is doubled ($k = 2$) then, for every dollar a subject contributes, she gets fifty cents back. But if all four subjects contribute a dollar, then each will get two dollars in total, fifty cents from her contribution and fifty cents from each of the other three players’ contributions.

When public goods games are played in the experimental laboratory, the overall contribution rate is usually between 40 and 60 per cent. Individual subjects’ contributions range between 0 and 100 per cent. This is similar to experiments in which people play the prisoner’s dilemma for money with a binary choice to contribute or not, where the proportion of participants choosing to cooperate is typically between 40 and 50 per cent. These experiments are typically characterized as testing cooperation and there are differences in behavior between players. Economists and psychologists have found that these differences in behavior correlate with an independent measure of cooperativeness, a personality classification called “Social Value Orientation”.

Social Value Orientation (SVO) was invented by two psychologists, David Messick and C. McClintock, who noted that when subjects are confronted with experimental games, subjects pursuing different goals may make the same choice. In games where each player has a choice between two actions, since there are more than two possible motives underlying behavior, some motivations will be observationally equivalent. SVO was developed in order to provide a classification of subjects in terms of their motivations, which could then be related to their behavior in subsequent games and in other situations outside of the laboratory.

Messick and McClintock assumed that there are three main motivational orientations with corresponding goals: cooperation, competition and individualism. Subjects are classified according to a series of choices that they make between payoff allocations for themselves and another player. Co-operative types are seen as maximizing the sum of the player’s monetary payoffs, individualists as maximizing their own monetary payoff and competitive types as maximizing the difference between their payoff and that of the other player. There is no particular theoretical reason for thinking that the types actually have the goals attributed to them by the psychologists. Rather their behavior is consistent with having that goal. I will come back to the interpretation of SVO and the causes of the differences between the types in the next section. For now, the important point is simply that SVO is an independent measure of individual differences that is related to differences in cooperative behavior.

Psychologists have long accepted that SVO is related to behavior. Outside of the laboratory, it is related to personality descriptions made by room-mates and friends and is predictive of helping behavior in non-gaming situations. In social psychology laboratory experiments based on game theory, SVO is a good predictor for subjects’ declared most preferred outcome in the matrix, their choices in the first game of a series and the relative proportion of co-operative choices they make over a series of games, including prisoners’ dilemmas.

When psychologists measure SVO, the choices are hypothetical. Subjects are asked which payoff allocations they prefer but no money is actually allocated. Some of the laboratory experiments in psychology also use such hypothetical choices. In the experimental economics paradigm choices are always backed up by monetary payments to make them “real”. They are incentive compatible, in that they do not rely on truthful revelation of preferences about some allocation to self and other. Each choice is between two vectors in this plane. Summing the individual’s 24 chosen vectors yields an estimate of the individual’s preferred motivational vector. Subjects with vectors between 67.5° and 22.5° are labeled cooperative, those with vectors between 22.5° and 337.5° are labeled individualistic and those with vectors between 337.5° and 292.5° are labeled competitive. This reflects the interpretation of SVO by its creators that “[i]n effect, different individuals assign different weights to their own and to others’ outcomes.”

In the measurement of SVO, subjects are asked to make a set of 24 choices. In each one they must choose one of two possible allocations of money between themself and another player. Their social value orientation is calculated from these choices. Specifically, these choices are derived from the circumference of a circle in own-other outcome plane, with its center at the origin (which has the consequence that the total amount to be allocated is not constant over combinations). A graphical representation of this can be found in the appendix to this paper. On the $x$ axis, defined as 0°, is own outcome and on the $y$ axis, defined as 90°, is the other’s outcome, so each grid reference is an alternative, each of which was dominant with respect to a particular motivation. The old method is also sometimes still used but is less satisfactory, for reasons given in Gold’s Framing and Decision Making, where further criticism of both methods and some suggested improvements can be found.

66 In the measurement of SVO, subjects are asked to make a set of 24 choices. In each one they must choose one of two possible allocations of money between themself and another player. Their social value orientation is calculated from these choices. Specifically, these choices are derived from the circumference of a circle in own-other outcome plane, with its center at the origin (which has the consequence that the total amount to be allocated is not constant over combinations). A graphical representation of this can be found in the appendix to this paper. On the $x$ axis, defined as 0°, is own outcome and on the $y$ axis, defined as 90°, is the other’s outcome, so each grid reference is an alternative.
counterfactual situation. Theo Offerman, Joep Sonnemans and Arthur Schram tested whether a person’s SVO is related to her contribution in public goods games, where the both the classification and the game were incentive compatible. They found that 65% of their subjects could be classified as individualists, 27% as cooperative, 4% as competitive and 1% as aggressive. In the ensuing public goods game, they observed individual contribution rates ranging from 0 to 100%. They then tested to see if there was a relationship between a subject’s being classified as a cooperator or an individualist and her subsequent contribution rate. They found that cooperators contributed more than individualists.

b. Trustworthiness: The Mach Scale and Trust Games

In a “basic trust game” there are two players. The first player has the choice to take some action, which has the potential for her to make a gain if player two fulfils her trust, but exposes her to risk because player two has some temptation to violate her trust, which would leave her worse off than she would have been if she had withheld her trust. This is often operationalized as an “investment game”, popularized by Joyce Berg, J. Dickhaut and Kevin McCabe. The first mover, or sender, has to decide how much of a $10 stake to send to the second player, or receiver. The experimenter then triples this amount and the receiver decides how much of it, if any, to send back. The efficient strategy for the sender is to send all the money, as this maximises the total earnings of the two. However, if the receiver is maximizing money payoffs, then her best strategy is to send nothing back. Knowing this, by backwards induction a money maximizing sender can deduce that she should send nothing. Despite this, experimental subjects both place trust in the other player and respond to trusting behavior. In this basic form of their game, Berg, Dickhaut and McCabe found that, on average, $5.16 was sent, resulting in an average payback from receivers of $4.66. Within this, there was plenty of individual variation: the amounts sent varied from 0 to $10, amounts paid back (given that money had been received)

from 0 to $20. These results are typical for trust experiments. Whether or not a subject sends money back is related to her trustworthiness, so trust experiments show that there are individual differences in trustworthiness. Economists have also found that subjects’ scores on the “Mach Scale”, an independent measure of personality type, are correlated with this behavior.

The Mach scale was developed by R. Christie and F. Geis. Their test was designed using Machiavelli’s _The Prince and Discourses_ and the scale was intended to measure a test-taker’s tendency toward Machiavellian behavior and, in particular, cynicism, a penchant for manipulativeness, and the belief that the end sanctifies the means. People’s score on the Mach scale is measured according to their degree of agreement with various normative statements, such as “all in all, it is better to be humble and honest than important and dishonest” and “anyone who completely trusts anyone else is asking for trouble”. Social psychologists find that individuals’ scores on the scale are stable over time and related to other personality assessment tools.

Anna Gunnthorsdottir, Kevin McCabe and Vernon Smith hypothesized that people who got high scores on the Mach scale (“high Machs”) should reciprocate trust less than those who got low scores (“low Machs”). In their trust experiment, subjects played the trust game explained above but with a binary decision whether or not to trust and respond to trust: senders had to send $10 or nothing, reciprocators could send back $15 (and keep $25) or nothing (and keep $40). (Everyone started with $10.) They found that 54.2% of low-to-average Machs reciprocated trust, whereas 72.2% of high Machs took advantage

---


71 The numbers do not add up to 100% as the remaining subjects showed too low a degree of consistency in their choices to be classified (below 33%). The overall level of consistency was 90%.

72 They used a step-level public goods experiment, so differences in expectation of being critical for the good’s provision might also affect behavior. However, they subjects’ estimated probability of being critical for the provision of the good incentive compatible, with subjects being paid according to their stated estimates according to a payoff schedule that was constructed so that the subjects would maximize their payments by telling the truth, and they found that the differences in contribution could not be explained by differences in subjects’ expectations.

73 See Michael Bacharach, Gerardo Guerra and Daniel Zizzo “Is Trust Self-Fulfilling? An experimental study” University of Oxford Department of Economics Discussion Paper No. 76 (2001) for a more formal statement of this, including mathematical conditions that the payoffs must meet for a game to be a basic trust game.


77 Although we might think that someone who exhibits all these in large quantities would not be very pleasant, the scale is explicitly supposed to measure a trait and not a pathology, so it is specifically a scale to classify “normal” people, as opposed to those with psychological disorders, and even people who come out as “high Mach” have normal social interactions.


79 p.52-3 of Gunnthorsdottir, A., McCabe, K. and Smith, V. (2002) “Using the Machiavellianism Instrument to Predict Trustworthiness in a Bargaining Game” _Journal of Economic Psychology_ 23, 49-66. They thought that the effect of Mach score on trusting would be ambiguous, because one the one hand high Machs are supposed to be more cynical than low Machs (so they would be less likely to expect to get money back) but on the other hand they are sometimes thought to be more risk seeking (so they might be more likely to send money despite a low estimate of the probability of getting any back).
of the trust and did not reciprocate, confirming the authors’ hypothesis about the lower trustworthiness of high Machs.

In both of these cases, cooperativeness in public goods games and trustworthiness in trust experiments, economists have found a correlation between a personality classification and a relevant behavior. They show that the classifications are predictive of individual differences in behavior but they do not offer an analysis of how the differences are caused, other than that the classification measures a trait. Regarding SVO, McClintock and Liebrand say that “[i]n effect, different individuals assign different weights to their own and to others’ outcomes.” So there is no test of the internal causes of the motivations. The types act as if they make the transformations attributed to them and the classifications are name tags for particular behaviors. (In fact, these results are evidence of cross-situational consistency of behavior.) If we don’t refer to the internal causes of behavior, we have a correlation of behaviors rather than a correlation of a causal trait attribution with behavior. In the next section, I identify differences located in the individuals’ decision making processes that, in each case, can explain both differences in the classifieratory tool and the differences in gaming behavior.

5. Explanatory Trait Attributions and Decision Making

Virtue ethics goes back to the Greeks and is particularly associated with Aristotle. In the Nicomachean Ethics, he states that “Excellence [of character], then, is a state concerned with choice.” Acting for appropriate reasons is central to most definitions of traits given by virtue ethicists. For instance, Julia Annas says “A virtue, unlike a mere habit, is a disposition to act for reasons, and so a disposition which is exercised through the agent’s practical reasoning; it is built up by making choices and exercised in the making of further choices. When an honest person decides not to take something to which he is not entitled, this is not the upshot of a causal build-up from previous actions, but a decision, a choice which endorses his disposition to be honest.” In the experiments cited in the previous section, the trait-relevant behaviors each involved making a decision, so we might look for explanatory causes of individual differences in the decision making process.

Aristotle also provides an account of decision making. He characterizes practical reasoning as involving a syllogism with a major, universal premise (e.g. everything sweet ought to be tasted) and a minor, particular premise (e.g. this thing is sweet) that lead to a conclusion (e.g. this thing should be tasted). In order to reach the conclusion of the syllogism, the agent has to “use” the minor premise. In other words, the agent has to notice a particular fact about the situation in order to trigger use of the major, normative premise. The idea that the way the agent construes the situation may affect the decision she makes has been represented in some formal models of decision making by incorporating her “understanding” or her “frame”. Frederick Schick argues that the agent’s “understanding” is the third part of a reason, alongside beliefs and desires.

Natalie Gold and Christian List argue that “framing” may trigger the use of a particular reason, which is composed of a combination of factual and normative propositions.

In these models, there are two potential loci of individual differences in the outcome of decision making. One source is different desires (in Schick’s framework) or the disposition to accept different normative propositions (in the Gold and List framework). This might lead two agents to act differently for any given way of construing the situation. The other source is the disposition to construe or “frame” situations differently. Even if two agents have the same desires or the same normative commitments, a difference in how they frame the situation will lead to a difference in behavior if this causes each to act on different normative commitments that have conflicting implications for the choice.

The idea that construal has an important role to play in behavior is also widely accepted in social psychology. Lee Ross and Richard Nisbett say the principle of constual, that the impact of a stimulus depends upon the “subjective meaning that the actor attaches to that situation”, is one of the three pillars on which psychology is based. They say that “To predict the behavior of a given person successfully, we must be able to appreciate the actor’s construal of the situations” and that a common error is the “failure to appreciate the inherent variability of situational construal”. Although for Ross and Nisbett construal is subjective and they say that there is “substantial variability” and “any two people are likely to interpret the same situation in somewhat different ways”, they do not consider individual differences in construal as a stable source of personality differences. For them, construal is a situational variable. It is obviously related to situational variables and there a lot of evidence that by

---

82 It is generally thought that Aristotle intended that the conclusion is also an action. For a minority dissenting view, that the conclusion of the syllogism is not identical with the action, see David Charles Aristotle’s Philosophy of Action (London: Duckworth, 1984).
83 Making Choices (Cambridge: Cambridge University Press, 1997)
85 For further details and conditions on when this would happen, see Gold Framing and Decision Making and Gold and List “Framing as Path Dependence”.
86 The Person and the Situation p.11-12.
87 Ibid p.68.
manipulating the presentation of options, experimenters can get subjects to exhibit “framing effects”. But they give no reason why construal should not also be a source of personality differences. I will argue that the evidence supports the idea that construal is a stable source of individual differences in behavior. In the next section I will come back to the implications for character traits.

If the principle of construal is taken seriously, then it makes problems for the testing of causal-explanatory individual differences because it is not clear that endorsing a reason and the likelihood of construing the situation so that that reason appears relevant should be related. Evidence suggests that they are not. Baruch Fischhoff gave subjects a choice between two contingency plans for civil emergencies. He then gave them three ways of thinking about the problem and asked “which of these three phrasings most closely captures the way in which you thought about the problem when making your choice between the two options?”. Fischhoff found that there was an absence of any relation between frame preference and option choice, i.e. some subjects who had reported using the same frame made different choices, whilst others made the same choice but reported using different frames. If behavior is caused by the interaction of perceptions and reasons, when we have two observations of different behavior in a situation, it is not clear at which locus the difference in behavior was caused and, even in two observations of the same behavior, the causes may not be the same.

I contend that the reason that SVO and the Mach Scale have been successfully used in economics experiments is that, by happy co- incidence, they have prised the two loci apart - that SVO is a measure of individual differences in construal and the Mach Scale of normative commitments - and that the games in which they are predictive of behavior are situations where the relevant locus can be expected to be the main cause of different choices.

**a. Differences in Construal and Social Value Orientation**

Economics experiments attempt to strip situations of all the normal contextual cues in order to isolate strategic variables. But a subject put in the experimental situation still has to describe the situation to herself. In effect, the experimenter has relinquished any control over the way subjects frame the situation. Subjects must discriminate between the options somehow, in a manner that provokes a reason for choice. I argue that SVO has predictive power because it measures subjects’ propensity to construe the situation in a particular way.

There is direct evidence supporting the idea that types with different SVOs spontaneously bring to mind different concepts when making decisions. In a negotiating experiment, De Dreu and Boles find that cooperative types recall more cooperative than competitive heuristics (and also use them), whilst individualists and competitors recalled more competitive than cooperative heuristics. Other evidence suggests that the different types attend to different information. Camac had his subjects play experimental games but he covered up all the payoffs in the matrix and made players uncover the outcome cells in order of decreasing importance, in order to get the information for their choice. Cooperative types are more likely to want to know the payoffs that they would get if both players make the same choice, whereas non-cooperative types considered the most important outcomes for their decision to be what they got from doing something different from the other player which, in the prisoner’s dilemma, includes the payoff for defecting if the other cooperated.

There is also evidence that a subject’s SVO correlates with the evaluative dimensions which she uses when playing the prisoner’s dilemma. Wim Liebrand and others got subjects to rate four different “partners” that they had played with on various evaluative dimensions. (The “partners” were in fact pre-designated computer strategies.) The experimenters found that cooperative types were more likely to discriminate between cooperative and defecting strategies of the other player along moral dimensions, such as goodness and fairness, whereas individualist types were more likely to discriminate along dimensions of potency such as power and intelligence. They called this the “Might versus Morality Hypothesis”.

Fairness and intelligence are evaluative concepts. On the whole, normal people want to do actions that are good and actions that are intelligent. However, in the
prisoner’s dilemma and public goods games these two principles conflict. The fair or good thing to do may be to contribute and the payoff maximizing thing to do is not to contribute. If all subjects want to be both intelligent and fair then it will matter which reason for acting they see as being appropriate to the situation. Differences in construal that affect which one appears relevant will cause differences in behavior. These same two principles also come into conflict in the SVO outcome allocation task, as a subject cannot distribute the money both fairly and maximize her payoffs. If SVO captures the likelihood that a subject will construe her situation in a certain way and, hence, for which reason she will choose, then the propensity to use these particular frames would be a common cause of subjects’ choices in the SVO outcome allocation task and their contribution choices in the public goods game.

b. Difference in Normative Commitments and The Mach Scale

Although experimental trust games are usually run outside of any contextual setting, by their nature these games offer little potential for different construals. As Gunnthorsdottir, McCabe and Smith point out, whilst simultaneous move games are often very ambiguous, games in which there is a move order are much less ambiguous. In their trust experiment, the payoffs were displayed spatially on a computer screen as a “game tree”, with the decision to send money represented as moving down and the decision not to send represented as moving right. They say that “Player 1 signals to his or her anonymous counterpart through his or her very move. Moving right conveys distrust in Player 2”. Conversely, sending money is obviously placing a trust so, when a receiver gets to make a decision, the situation is clear.

According to my analysis, if their construal of the situation is the same, then differences in subjects’ behavior should reflect differences in normative commitments. This is precisely what the Mach Scale tests, so we could attribute differences in trustworthiness between high and low Machs to differences in the principles they endorse.

To sum up, in this section I have argued that each of SVO and the Mach Scale captures a single source of individual differences in the decision making process: SVO captures differences in construal and the Mach Scale asks about their normative commitments. The reason they have been effective in predicting behavior in experimental games is that they have each been used in experiments where the source of difference they measure can be expected to be critical in causing behavior.

6. On the Existence of Character: Decision Making, Character Traits and Virtue Ethics

I have argued that there are stable, internal sources of individual differences in behavior that are realized in the decision making process. It remains to argue that these are differences that we would recognize as character traits, and to connect them back to traits as employed in virtue ethics.

I identified two sources of individual differences that combine to cause differences in behavior: the reasons a person endorses and her construal of the situation. Ross and Nisbett say that “predictable individual differences are not always mere cognitive illusions” and that their account of individual differences “gives heavy weight to the complex dynamics of social systems and to the role of construal processes”. They recognize the importance of differences in construal for behavior, they just do not consider them to be related to character traits. Harman follows Ross and Nisbett’s position, when he says that people “differ in their situations and their perceptions of their situations….. But character traits do not explain what differences there are.” Much of the following discussion will focus on the implications of the perceptual component of character traits.

Ross and Nisbett’s textbook is not a bible and their interpretations are not sacrosanct. Nor is their position representative of all psychologists. In contrast to Ross and Nisbett, George Kelly, an influential figure in twentieth century psychology, argued that the way a person represents the world amounts to that person’s personality. The evidence of the previous section shows that differences in the propensity to use particular frames is a part of stable personality differences. Some personality traits seem to result mainly from construing the world in a certain way; most obviously optimism, which is to tend to attach a positive valence to features of the situation, but also aggression, which includes a tendency by the agent to interpret people’s actions as hostile towards her.

Nor am I the first philosopher to include construals as a part of character traits. My account of two loci of individual differences has affinities with the analysis of character traits given by Douglas Butler. He argues that there are some two-part traits, which consist of a disposition to represent the situation in a particular way and the disposition to reason practically on that representation, and other three-part traits, which also involve a specific conditional desire. With respect
to the account I have given, two-part traits would correspond to behavior caused by differences in construals of the situation, and three-part ones to the use of differing reasons given the frame. Some contrasts between our two accounts will come out in the discussion below, as I relate these two sources of traits to virtue ethics.

Some virtue ethicists have emphasized the role of perceptions in the virtues. Foremost amongst them is John McDowell, who has argued that “a failure to see reason to act virtuously stems… from the lack of a distinctive way of seeing situations”. Although I have identified two different loci of trait differences, I think that there is reason to expect that propensity to use certain frames is particularly important for character traits. Another function that character traits have, which I have not explored in this paper, is to attribute praise and blame. They have an evaluative component. Normal people tend to want to do evaluatively positive things but perception of a normative principle is necessary before one can act on it. My claim is that, when normal people fail to act virtuously, this is more likely to be from a failure to perceive reason to act virtuously rather than from a deliberate decision not to act on that reason. Perceptions may do much of the work in causing moral behavior.

This analysis has implications for the relation between possessing and not possessing traits, and possessing and not possessing trait opposites. It may help here to consider Butler’s three-part character traits in more detail. For him, being kind and being cruel are three-part character traits. Each involves noticing the same fact of a person in trouble, but kindness involves a desire to help her given this representation of her predicament whereas cruelty involves the desire not to help. Normal people in normal circumstances do not usually have “negative” desires not to help. But this does not imply that everyone is kind. What people do have are “positive” desires to do other things. People fail to be kind because they are doing something else for some other reason, like hurrying to keep an appointment. This tells against the Aristotelian doctrine of the mean because not being kind is not simply an intermediate between being kind and being cruel.

I have only talked about agents who have a disposition to frame a situation in a particular way and who endorse certain reasons. However, it might be the case that some traits involve not having certain construals or not acting on certain reasons, or some combination of having and not having construals and reasons. For instance, Butler says that courage is a two-part trait involving noticing danger and not responding to it, which we might think of as not acting on certain reasons. Assessment of this claim requires further analysis and the relating of that analysis to particular virtues.

A further complication, which I have not tackled here, is that a person may be aware of conflicting reasons and even of conflicting ways of construing the situation. (The second of these relates to the second element of Butler’s two-part traits, the disposition to reason practically on a representation, as it is possible that a person may be aware of a way of framing the situation but not act on it.) Grounds for not considering this as central to the analysis come from the psychology literature. Regarding reasons, there is evidence that people prefer to make choices on the basis of a single reason, to the extent of constructing and selecting choice situations such that there is always a dominant reason for choice. Regarding construals, people are typically not aware that there are different ways of construing their situation. This is a claim of Ross and Nisbett’s that is supported by evidence from framing experiments, which test how frames affect decisions, where Kahneman and Tversky say that subjects do not realize that they would have made a different decision under a different framing of the problem. Given that my project was to provide an analysis that would allow us to assess evidence relating to the propriety of trait attributions, if people do not tend to have multiple construals then they are not central to my task. Further, the fact that people are generally not aware that there are conflicting ways of framing a situation helps to explain why the individual differences in construal are so influential in behavior.

However, although this may be how people actually make decisions, the possibility of multiple construals raises another issue for virtue ethics, namely that true virtue requires certain types of deliberation and not just dispositions to action. If the virtuous agent exercises no influence on her frame, then the possession of a virtue is a case of moral luck. Maybe the character building that virtue ethicists advocate should in fact consist of developing particular ways of seeing situations, or developing the disposition to construe the same situation in multiple ways.

Framing the situation in a particular way seems to be necessary for virtuous action. However, it is not always sufficient. (Whether it is sometimes sufficient relates to the example of courage and not acting on reasons above, where it is arguable that the motivation for the courageous act does not matter in its correct attribution.) The possession of at least some virtues requires acting for an

---

101 These issues about the relations between perceptions, reasons and how they motivate people is related to Aristotle’s discussion of akrasia, on which see my Reasons and Decision Making.


appropriate reason. For instance, one might see that the situation affords possibilities for helping behavior and therefore perform that behavior, but perform it with the motivation of impressing some bystander rather than in order to help the victim. This is arguably not virtuous behavior. It raises questions for psychology and for virtue ethics. Can people perform an action for two motivations at once? If the answer to this first question is “yes”, what are the implications of dual motivations for virtuous actions? These questions are too big to pursue further here.

7. Conclusion

After concluding that the standard dispositional analysis of character traits fails, Doris says: “No bother: I’m doing moral psychology, not metaphysics; my interest is not in conceptual analysis but in the evidential standards regarding trait attribution.” I hope I have shown both that conceptual analysis is necessary in order to evaluate the evidence and that a correct evaluation of it does not lead to Harman and Doris’s conclusions. I provided a preliminary analysis of the role of perceptions and reasons in character traits. I think that the task of moral psychology is precisely to provide further analysis in order to clarify and interpret work in psychology and to de-limit and assess virtue ethics.

I should stress that, although I base my analysis in decision-making, I do not mean to imply that all acting for reasons is connected to character traits. Conversely, although I said that acting for reasons is central to virtue ethics, it would be wrong to imply that virtue ethicists are only concerned with reasons and decision making. For instance Rosalind Hursthouse says that character traits are a complex mindset. According to her this consists of a variety of things, which I think that it is fair to characterize as reasons, emotions and cognitions. I have talked about reasons and cognitions but said nothing about emotions. Emotions may influence the decision making process, in which case it would be possible to include them in a more complicated version of my analysis. But I do not mean to claim that character traits consist only of those factors that are causal in the decision making process. There may also be other aspects of character which do not affect decisions. These may include emotions and affective states. I leave it to virtue ethicists to enumerate these.

---

104 Lack of Character p.16
105 For instance, we sometimes act for reasons that relate to physical states, such as eating because one is hungry, that do not relate to character. A further difference relating to personality traits, but not what we would consider character traits, is that not all (personality) trait explanations are based in reason explanations (for instance those traits that are stylistic, or denote a manner of acting, such as when we say of a person that she is quiet).