

# THE CASE FOR NIETZSCHEAN MORAL PSYCHOLOGY<sup>#</sup>

Joshua Knobe\* and Brian Leiter\*\*

Draft of September 29, 2005

Forthcoming in Brian Leiter and Neil Sinhababu (eds.). *Nietzsche and Morality*  
(Oxford: Oxford University Press, 2007).

Comments welcome: [bleiter@mail.law.utexas.edu](mailto:bleiter@mail.law.utexas.edu)

*Please do not cite or quote without permission.*

## I. Introduction

Moral psychology is the branch of ethics directly concerned with the psychology of the kind of agency we exercise in acting morally. Moral psychology asks whether such agency is psychologically possible, what motivations it requires, what the source of those motivations might be, and what the emotive and cognitive mechanisms are by which they translate into actions. Until fairly recently (e.g., Doris 2002; Nichols 2004; Prinz forthcoming), Anglophone philosophers writing about moral psychology have tended to approach these questions “from the armchair,” and without regard to pertinent empirical findings about human psychology.<sup>1</sup>

Indifference to empirical findings is probably not unrelated to a second striking feature of the moral psychology literature, namely, that it has been dominated by just two major historical figures, Aristotle and Kant. From Aristotle has come to us the tradition

---

<sup>#</sup> This was a fully collaborative project; authors are listed alphabetically. We are deeply grateful to John Doris both for the stimulus of his published work and for many hours of conversation on these topics. We are also grateful to comments and suggestions from Gilbert Harman and C.D.C. Reeve.

\* Department of Philosophy, University of North Carolina at Chapel Hill.

\*\* School of Law and Department of Philosophy, University of Texas at Austin.

<sup>1</sup> Moral psychologists influenced by Freud, like Deigh (1996), are also an exception to the inattention to empirical psychology, though even Deigh does not spend time investigating the empirical

of virtue ethics,<sup>2</sup> which emphasizes the importance of stable characterological dispositions to act in morally appropriate ways, dispositions which it is the task of a sound moral education to inculcate in children. From Kant, by contrast, has come the rationalist tradition in moral psychology,<sup>3</sup> according to which reason is the source of moral motivation, and the mechanism for moral action is one in which rational agents legislate for themselves certain principles on the basis of which they consciously act.

Our goal in this essay is to add a third figure to this debate, namely Nietzsche, and to show that a fair reading of the relevant empirical sciences strongly favors many aspects of his moral psychology as against the Aristotelian and Kantian traditions. We shall largely follow the account of Nietzsche's moral psychology in Leiter (2002), which makes the interpretive case for the reading relied on here. Our primary concern in this paper is not interpretive, but philosophical: to show that neglect of Nietzsche in moral psychology is no longer an option for those philosophers who recognize that moral psychology must be based on *real* psychology.

## **II. Three Views in Moral Psychology**

The Aristotelian and Kantian traditions in moral psychology are historically complex and philosophically rich. Our ambition, plainly, is not to do justice to the history or even all the philosophical permutations. Rather, we want to extract certain *core* and *distinctive* elements of these traditions, ones that are, on almost any rendering, important to the views so named, and which, at the same time, involve psychological claims that admit of empirical evaluation. Just as there are a multitude of "Humean"

---

evidence for Freudian moral psychology. But the Freudian theory is an empirical one, and support does exist (e.g., Westen [1998]).

<sup>2</sup> See, e.g., \_\_\_\_\_.

views in ethics and action theory that are traceable to Hume, but do not necessarily have the full texture of Hume's actual views, so too, we claim, there are Kantian and Aristotelian views in moral psychology that are traceable to their distinguished historical forebears, but which we do not claim are Kant's or Aristotle's *precise* views. What we do claim, in each case, is that the views in question are important views in moral psychology *to the present* and that these views do not fare well when compared to the, hitherto, under-appreciated "Nietzschean" approach to moral psychology.

#### A. Aristotle

In the Aristotelian tradition of moral psychology, moral agents are *virtuous* agents, that is, agents possessed of stable dispositions to act in morally appropriate ways as different situations require. The agent who acts morally, according to Aristotle, has three attributes: "he must act knowingly, next he must choose the actions, and choose them for themselves, and thirdly he must act from a firm and unalterable character" (*NE* 1105a29-33).

But Aristotle does not merely suggest that moral action stems from a certain type of character; he also advances a series of specific hypotheses about the nature and origin of that type of character. In particular, he claims that good character consists in certain *habits* (*NE* 1103a25), that these habits are acquired during *childhood* (*NE* 1103b25) and that the key to their acquisition is proper *upbringing* (*NE* 1095b5-10). Ultimately, then, we are left with a definite picture of how virtuous character is acquired. This picture says that people are encouraged to perform certain virtuous behaviors during childhood and

---

<sup>3</sup> See, e.g., \_\_\_\_\_.

that they gradually come to acquire the corresponding dispositions, leading eventually to a full-fledged possession of the relevant virtue.

Richard Kraut (2001) provides a more nuanced discussion of this hypothesis:

All free males [according to Aristotle] are born with the potential to become ethically virtuous and practically wise, but to achieve these goals they must go through two stages: during their childhood, they must develop the proper habits; and then, when their reason is fully developed, they must acquire practical wisdom (*phronêsis*). This does not mean that first we fully acquire the ethical virtues, and then, at a later stage, add on practical wisdom. Ethical virtue is fully developed only when it is combined with practical wisdom (1144b14-17). A low-grade form of ethical virtue emerges in us during childhood as we are repeatedly placed in situations that call for appropriate actions and emotions; but as we rely less on others and become capable of doing more of our own thinking, we learn to develop a larger picture of human life, our deliberative skills improve, and our emotional responses are perfected. Like anyone who has developed a skill in performing a complex and difficult activity, the virtuous person takes pleasure in exercising his intellectual skills. Furthermore, when he has decided what to do, he does not have to contend with internal pressures to act otherwise. He does not long to do something that he regards as shameful; and he is not greatly distressed at having to give up a pleasure that he realizes he should forego.

“To keep such destructive inner forces [or pressures] at bay,” notes Kraut, “we need to develop the proper habits and emotional responses when we are children, and to reflect intelligently on our aims when we are adults.”

This “process of training” through which a virtuous agent is produced is not, as John Cooper emphasizes, “purely mechanical”:

Aristotle holds that we become just (etc.) by being repeatedly made to act justly (etc.).... [S]ince he emphasizes that the outcome of the training is the disposition to act in certain ways, knowing what one is doing and choosing to act that way, the habituation must involve also...the training of the mind. As the trainee becomes gradually used to acting in certain ways, he comes gradually to understand what he is doing and why he is doing it: he comes, to put it vaguely, to see the point of the moral policies which he is being trained to follow, and does not just follow them blindly. (Cooper 1975: 8; citations omitted)

Of particular importance for our purposes are two features of Aristotle’s moral psychology of the virtuous agent: first, the moral agent, properly raised, must have “a firm and unalterable character”; second, this type of character is typically the product of *childhood upbringing*.<sup>4</sup> Although there has been a great deal of excellent work on the proper interpretation of Aristotle’s account of the origin of virtue, there has been

---

<sup>4</sup>Although these two themes have been central to the 'Aristotelian' tradition within contemporary moral psychology, Aristotle himself appears to have had a more complex and multi-facetted view. He attributes the development of character to a broad process of 'acculturation' (trophē) which includes more than just treatment from one's caregivers, and he mentions at a number of points that there are innate differences between individuals in their capacity for virtue, even to the point of suggesting that women and slaves are not capable of true virtue regardless of their childhood experiences. Since modern philosophers working in the tradition of Aristotelian moral psychology have no reason to accept Aristotle’s view about who has the potential to be virtuous, we may assume that a credible modern Aristotelian moral psychology must be committed to the proposition that everyone is potentially “brought up” properly such that they can become virtuous agents.

surprisingly little discussion of the question as to whether or not Aristotle's views are actually correct. Our concern here will be with this latter question. We want to know whether there actually is any evidence for the view that people's dispositions are shaped primarily by childhood upbringing or whether people's dispositions might arise through some other process entirely.

#### B. *Kant*

In the Kantian tradition of moral psychology, moral obligations are grounded in principles that each agent consciously chooses. But it is not enough for an agent simply to perform behaviors that happen to accord with these moral principles. If an agent's behavior is merely the product of emotion or habit, then no matter how well that behavior fits with her moral principles, she can never truly be acting morally. Genuine moral action must actually be chosen *because* it is morally right. Or, as Kant famously puts it, genuine moral action is not merely *in accordance with* duty; it is done *out of* duty.

Here is how J.B. Schneewind usefully summarizes the Kantian view:

At the center of Kant's ethical theory is the claim that normal adults are capable of being fully self-governing in moral matters. In Kant's terminology, we are "autonomous." Autonomy involves two components. The first is that no authority external to ourselves is needed to constitute or inform us of the demands of morality. We can each know without being told what we ought to do because moral requirements are requirements we impose on ourselves. The second is that in self-government we can effectively control ourselves. The obligations we impose upon ourselves override all other calls for action, and

frequently run counter to our desires. We nonetheless always have a sufficient motive to act as we ought. (Schneewind 1992: 309)

So on the Kantian view of moral psychology, (1) agents impose moral requirements on themselves, and (2) these self-imposed requirements are motivationally effective. In order for the self-imposition of moral requirements to be genuinely autonomous it must presumably be a conscious process of self-imposition. And for these consciously imposed principles to be motivationally effective it must be the case that conscious moral principles are motivationally effective.<sup>5</sup>

### C. *Nietzsche*

The Nietzschean account of moral psychology differs from the Aristotelian and Kantian accounts along almost every dimension. What is decisive is not upbringing, particular habits, or conscious choice; what matters is heritable psychological and physiological traits.

Of course, Nietzsche would not deny that people have habits and conscious moral principles. The only question is about whether these factors actually play any important role in the etiology of people's moral behavior. So, for example, Nietzsche would say that conscious moral principles don't actually lead people to perform moral behaviors. Instead, people *first* perform certain behaviors and *then* develop principles that serve to justify the behaviors they have already performed. The most important factors in the origin of moral behavior are people's basic psychological and physiological traits; the

---

<sup>5</sup> We take Schneewind's summary, and the points we emphasize, to comport reasonably well with more elaborate treatments of Kantian ethics and moral psychology, such as that in Korsgaard (1996). So, e.g., Korsgaard says that for Kant, "principles of practical reason" are "principles that govern choice" (xii) and that Kant demonstrates "the reality of moral obligation" in the *Critique of Practical Reason* by appeal to "our consciousness of the moral law and its capacity to motivate us whenever we construct maxims. We

conscious moral principles simply serve as post hoc justifications for behaviors that would have been performed either way.

Under the influence of several decades of postmodern readings, these central aspects of Nietzsche's moral psychology have often been neglected. In consequence, they deserves a bit more exposition than we have accorded to the Aristotelian and Kantian views, whose broad outlines are widely recognized. To begin, we should remember that Nietzsche was very much influenced by the idea, popular among German Materialists in the 1850s and after, that human beings are fundamentally bodily organisms, creatures whose physiology explains most or all of their conscious life and behavior (see generally Leiter 2002: 63-71). Nietzsche adds to this Materialist doctrine the proto-Freudian idea that the unconscious psychic life of the person is also of paramount importance in the causal determination of conscious life and behavior.<sup>6</sup> Thus, Nietzsche accepts what we may call a "Doctrine of Types" (Leiter 2002: 8), according to which,

Each person has a fixed psycho-physical constitution, which defines him as a particular *type* of person.

These "type-facts", for Nietzsche, are either *physiological* facts about the person, or facts about the person's unconscious drives or affects. The claim, then, is that each person has certain largely immutable physiological and psychic traits that constitute the "type" of person he or she is. While this is not, of course, Nietzsche's precise terminology, the ideas are familiar enough from his writings.

---

are conscious of the law not only in the sense that it tells us what to do, but in the sense that we know we *can* do what it tells us, no matter how strong the opposing motives" (26).

A typical Nietzschean form of argument, for example, runs as follows: a person's theoretical beliefs are best explained in terms of his moral beliefs; and his moral beliefs are best explained in terms of natural facts about the type of person he is (i.e., in terms of type-facts). So Nietzsche says, “every great philosophy so far has been...the personal confession of its author and a kind of involuntary and unconscious memoir”; thus, to really grasp this philosophy, one must ask “at what morality does all this (does *he*) aim” (BGE 6)? But the “morality” that a philosopher embraces simply bears “decisive witness to *who he is*” — i.e., who he *essentially* is — that is, to the “innermost drives of his nature” (BGE 6). Indeed, this explanation of a person's moral beliefs in terms of psychophysical facts about the person is a recurring theme in Nietzsche. “[M]oralities are...merely a sign language of the affects” (BGE 187), he says. “Answers to the questions about the *value* of existence...may always be considered first of all as the symptoms of certain bodies” (GS P:2). “Moral judgments,” he says are, “symptoms and sign languages which betray the process of physiological prosperity or failure” (WP 258). “[O]ur moral judgments and evaluations...are only images and fantasies based on a physiological process unknown to us” (D 119), so that “it is always necessary to draw forth...the *physiological* phenomenon behind the moral predispositions and prejudices” (D 542). A “morality of sympathy,” he claims is “just another expression of ... physiological overexcitability” (TI IX:37). *Ressentiment* — and the morality that grows out of it — he attributes to an “actual physiological cause [*Ursache*]” (GM I:15). Nietzsche sums up the idea well in the preface to the *Genealogy*: “our thoughts, values, every ‘yes,’ ‘no,’ ‘if’ and ‘but’ grow from us with the same inevitability as fruits borne

---

<sup>6</sup> Nietzsche’s “official” view seems to be that physiology is primary, but he mostly concentrate on

on the tree — all related and each with an affinity to each, and evidence of one will, one health, one earth, one sun” (GM P:2).

We can see Nietzsche’s Doctrine of Types clearly at work in his discussion of the “error of confusing cause and effect,” in sections 1 and 2 of “The Four Great Errors” section of *Twilight of the Idols*, which is devoted to debunking the idea of free will. The crux of this first error can be summarized simply: given two regularly correlated effects E1 and E2 which have the same “deep cause,” we confuse cause and effect when we construe E1 as the cause of E2, missing altogether the existence of the deep cause that *really* explains them both. Let us call this error “Cornarism” after the (now) famed example Nietzsche invokes:

Everybody knows the book of the famous Cornaro in which he recommends his slender diet as a recipe for a long and happy life...I do not doubt that scarcely any book (except the Bible, as is meet) has done as much harm... The reason: the mistaking of the effect for the cause. The worthy Italian thought his diet was the *cause* of his long life, whereas the precondition for a long life, the extraordinary slowness of his metabolism, the consumption of so little, was the cause of his slender diet. He was not free to eat little *or* much; his frugality was not a matter of “free will”: he became sick when he ate more. (TI IV: 1)

In other words, what explains Cornaro’s slender diet *and* his long life is the same underlying fact about his metabolism. Cornaro’s mistake was to prescribe his diet for all

---

psychological claims, most obviously because he is no physiologist!

without regard for how individuals differed metabolically, metabolism being the relevant type-fact in this context.

Even if we grant Nietzsche all the facts as he presents them, this would not suffice for a *general* attack on free will unless the error involved in Cornarism extended beyond cases such as diet and longevity. But that is exactly Nietzsche's contention, since in the very next section he saddles morality and religion quite generally with Cornarism.

According to Nietzsche, the basic "formula on which every religion and morality is founded is: 'Do this and that, refrain from that and that—then you will be happy! Otherwise....'" Cornaro recommended a slender diet for a long life; morality and religion prescribe and proscribe certain conduct for a happy life. But, says Nietzsche,

[A] well-turned out human being...*must* perform certain actions and shrinks instinctively from other actions; he carries the order, which he represents physiologically, into his relations with other human beings and things.

So morality and religion are guilty of Cornarism: the conduct they prescribe and proscribe in order to *cause* a "happy life" are, in fact, *effects* of something else, namely the physiological order represented by a particular agent, one who (as Nietzsche says) "*must* perform certain actions," just as Cornaro *must* eat a slender diet (he is "not free to eat little *or* much"). That one performs certain actions *and* that one has a happy life are themselves both effects of the physiological order. If we grant Nietzsche the Doctrine of Types, then there is indeed reason to think that Cornarism is a feature of morality too, since morality fails to recognize the crucial role of type-facts in determining what one does, even what morality one accepts.

The implications, in turn, of such a view for moral psychology should be apparent: individuals are simply born with a certain psycho-physical package of traits (the person's distinctive type-facts); these type-facts play a powerful role in determining one's behavior and values, a far more powerful role than education or upbringing or conscious choice; indeed, a person's crucial conscious choices and values are themselves explicable in terms of these type-facts. *That* one is a "moral" agent is explained by one's biological inheritance, the type-facts; *that* one is not a moral agent is similarly explained.

#### D. *Three Views in Moral Psychology: Summing Up*

Thus far, we have been presenting three rival views in moral psychology. Our goal now is to figure out which of these three views provides the best account of how people actually come to perform moral behaviors. In addressing this question, we make use of an extremely straightforward methodology. We simply turn to studies that directly measure the extent to which different factors appear to be influencing behavior.

Of course, it might be suggested here that the issue before us now is not really an empirical one. Thus, someone might say: 'Kant's theory is not intended as a psychological hypothesis. It should be understood rather as a statement of the conditions of possibility of moral agency. Hence, if we find that no one actually meets the conditions set out by the theory, we should not conclude that the theory itself was mistaken. Instead, we should conclude that no one ever truly is a moral agent.' Let us call philosophers who adopt such a posture *Above-the-Fray Kantians*. Such philosophers are indeed invulnerable to the empirical results, but they are also, in our view, decidedly uninteresting. We will assume, with most moral philosophers (including many Kantians), that there are agents who perform morally valuable acts, and thus the question

for moral psychology is *not* merely a question about the *possibility conditions* for such psychology, but how this psychology *actually works*.

Yet, though we do regard the issue as an empirical one, we should also emphasize that it is not the kind of issue that could ever be resolved by a single crucial experiment. In essence, the problem here is that none of the three views can be refuted by a single isolated case. Virtue ethicists in the Aristotelian tradition do not typically claim that *everything* about a person's character was determined by the way in which he or she was brought up. Nor does Nietzsche need to say that *everything* about a person's character is determined at birth. The three positions differ primarily in their understanding of what *typically* happens in cases where a person performs a behavior deemed valuable. Our question is whether the existing empirical evidence favors one of these positions over the others.

### **III. The Empirical Evidence in Moral Psychology**

To address this question, we turn to the literature in empirical psychology. We will proceed by reviewing psychological research that will enable us to assess the plausibility of the Aristotelian, Kantian, and Nietzschean assumptions about what people are like. The evidence strongly suggests, we shall argue, that the Nietzschean view is far more likely to be correct than either of the others.

We should emphasize that the empirical results we will be discussing here are not those of a few maverick scientists drawing on some small number of scattered experiments. Rather, we will be focusing on some of the major lessons of personality and social psychology, replicated in numerous experiments using a wide variety of methodologies and subject pools. Occasionally, we will describe a specific experiment

and report its results, but the importance of these specific experiments is not that they themselves provide evidence for the theories discussed but rather that they serve as *examples*—giving the reader a sense for the kinds of techniques and results to be found in the relevant literatures. In addition to descriptions of specific experiments, we therefore rely heavily on reviews that summarize large numbers of relevant studies. Thus, to take just one example, we briefly mention a paper by Feingold (1992) on the impact of attractiveness on personality. That paper is a review of more than ninety studies including a total of more than *fifteen thousand* subjects. What makes Feingold's theory convincing is the fact that such a wide variety of studies have converged on a single basic result. The same could be said of each of the other theories we discuss.

#### **A. *Type-Facts and Heredity***

As we have seen, Nietzsche puts forward the view that a person's traits are determined, to a great extent, by factors (type-facts) that are fixed at birth. This view has gone more or less unexplored in the contemporary philosophical literature on moral psychology. (No one suggests, e.g., that the secret to becoming a compassionate person might lie in part in inheriting a genetic propensity of compassion.) And yet, although the Nietzschean view has not found much favor among philosophers, it is receiving an ever-growing mountain of support from empirical studies.

The most important evidence here comes from studies in behavioral genetics. Typically, these studies are conducted either by looking at twins (comparing monozygotic to dizygotic) or by looking at adopted children. The results of such studies are as consistent as they are shocking. Almost every personality trait that has been studied by behavioral geneticists has turned out to be heritable to a surprising degree. So,

for example, a recent review of five studies in five different countries (comprising a total sample size of 24,000 twins) estimates that genetic factors explain 60% of the variance in extraversion and 50% of the variance in neuroticism (Loehlin 1992).

It is difficult to convey just how astoundingly high these numbers are, but perhaps one can get a better sense for the issue by considering the effect sizes obtained in some classic social psychology experiments. The Festinger and Carlsmith (1959) study that launched the investigation into cognitive dissonance found an effect that explained 13% of the behavioral variance; the Darley and Batson (1968) study of bystander intervention and the diffusion of responsibility found an effect that explained 14% of the behavioral variance; the Milgram (1975) study of obedience and proximity showed an effect that explained 13% of the behavioral variance.<sup>7</sup> These are among the most influential and important experiments in all of social psychology. In each case, the fact that researchers were able to explain 13-14% of the variance led to a veritable revolution in our understanding of the relevant phenomena. Now consider, by contrast, the fact that behavioral geneticists routinely find effects that explain *fifty percent* of the variance in trait measures. Effect sizes of this magnitude are beyond the range that would previously have been considered possible.

Having said that, we should emphasize that it would be a mistake to attach too much importance to the exact percentages obtained in these studies. On one hand, adoption studies generally yield lower heritabilities than twin studies do, and one might therefore suspect that the true heritabilities are lower than those reported here. On the

---

<sup>7</sup> Note for the statistically inclined: No effect sizes are reported in the original papers, but Funder and Ozer (1983) have shown that it is possible to compute additional analyses based on information that the authors do report. (All effect sizes given here are calculated by taking the square of the relevant correlation coefficient.)

other hand, our ability to measure traits is quite limited, and one might therefore suspect that we would obtain even higher heritabilities if we could develop a more accurate trait measure. Whatever the resolution of these various difficulties, it seems clear that most traits have extremely high heritabilities.

Here we should pause to avert a potential misunderstanding of what it means for a trait to be 'heritable.' When we say that a trait is heritable, we do not mean that it is produced entirely by a person's genes, without any intervention from the environment. All we mean is that the differences between different people's scores on this trait can be explained in part by differences in those people's genetic material. This effect may not be direct. Differences in people's genes might lead to differences in their environments, which in turn lead to differences in their scores on certain traits. Often the result will be a self-reinforcing cycle in which early behaviors that express a given trait lead the person to possess that trait to ever greater degrees. For example, a person's initial extraverted behavior might leave her with a reputation for extraversion, which in turn makes her even more extraverted.

At least in principle, then, it is possible that heritable differences in personality are caused by heritable differences in some non-psychological characteristic. For example, it might turn out that heritable differences in physical appearance lead to differences in treatment by parents and peers, which in turn lead to differences in personality traits (Hoffman 1991). In actual fact, however, it is highly unlikely that any substantial portion of the variance in personality traits can be explained in this way. To take one striking example, physical attractiveness appears to have almost no impact at all

on personality: it explains around 2% of the variance in dominance, 0% of the variance in sociability, 2% of the variance in self-esteem, and so forth (Feingold 1992).

Of course, the impact of genetics is not confined to morally-neutral traits like extraversion and neuroticism; it also extends to traits that lie at the heart of moral psychology. Consider the tendency to use violence (what psychologists sometimes call ‘aggressive antisocial behavior’). A number of studies have examined the causes of violent behavior among children, and all show a strong influence of genetics. One recent study using 1,523 pairs of twins found a heritability of 70% (Eley, Lichtenstein & Stevenson 1999). Other studies yield percentages that are lower but still surprisingly high — 60% (Edelbrock, Rende, Plomin & Thompson 1995) 49% (Deater-Deckard & Plomin 1999) and 60% (Schmitz, Fulker & Mrazek 1995). These huge effect sizes cannot plausibly be ascribed to experimental artifacts or measurement error. Clearly, genetic factors are playing a substantial role in the etiology of certain kinds of violence.

Studies like these confirm the commonsense view that morally-relevant traits, like most other traits, are the product of not only environmental factors but also of heredity. This is the view we find assumed (commonsensically enough) in the works of Nietzsche, where it is enmeshed in a complex fabric of philosophical reflection. Subsequent philosophical work, in both the Aristotelian and Kantian traditions, has more or less ignored the role of heredity, focusing either on environmental factors like culture and upbringing, or ignoring questions about the genesis of motivation altogether. Yet all available evidence points to the view that heredity plays a major role in the development of morally-relevant traits, and if we want our moral psychology to be defensible and empirically sound, we need to grapple seriously with the philosophical issues this

evidence raises. Of the three great figures in the history of moral psychology, only Nietzsche has come to terms with the issue.

### **B. *Type-Facts and Fatalism***

Thus far, we have been concerned with questions about how people come to have certain traits rather than others. But Nietzsche also makes very strong claims about the *importance* that these traits — however they are acquired — actually have in people's lives. A person's character, he seems to suggest, has a substantial and pervasive impact on the whole course of that person's life. This claim may seem so banal and obviously correct as not even to be worthy of discussion. In actual fact, however, aspects of it have in fact been the object of a long-standing controversy within social and personality psychology.

Personality psychologists have performed numerous studies in which subjects first engage in some task designed to measure their personality traits (typically, filling out a questionnaire) and then are given an opportunity to perform a behavior that ought to be influenced by those traits. One surprising result of such studies is that correlations between a trait measure and an actual behavior rarely exceed .30. In other words, the trait measure rarely allows us to explain more than 9% of the variance in the behavior.<sup>8</sup> This is an extremely important finding, and it has been discussed in detail by both personality and social psychologists.

In his groundbreaking discussion of the phenomenon, Mischel (1968) suggested that perhaps broad traits do not really exist at all. The suggestion was that it might be

---

<sup>8</sup> Note on statistics: Although results in behavioral genetics are normally reported as percentages of variance, results in personality psychology are normally reported as correlation coefficients. For the sake of consistency, we therefore transform each correlation coefficient ( $r$ ) into a coefficient of determination

more accurate to posit only extremely narrow traits (e.g., a tendency to cheat on exams by copying other people's answers) and stop looking for broad traits like 'extraversion' and 'neuroticism.' This suggestion spurred a great deal of debate throughout the 1970's (e.g., Bem & Allen 1974; Jones & Nisbett 1972), but that debate is now over. Almost all psychologists now believe that broad traits do exist.<sup>9</sup> The key question is how important they are — whether they actually have a large impact on people's behavior or whether they turn out to be far less powerful than certain subtle situational forces.

This issue is surprisingly complex. Ross and Nisbett (1991) have offered sophisticated arguments for the view that traits have only a small impact on behavior, but Funder and Ozer (1983) and Epstein (1979) have offered arguments of equal sophistication for the view that traits can have quite large impacts on behavior.

To get a sense for the complexity of the issue, consider what would happen if we tried to predict a basketball player's performance using some measure of his or her ability. Clearly, our predictive power would depend in part on how much of the player's behavior we were trying to predict. If we tried to predict the player's success in getting one particular randomly-selected rebound, our measure of ability would give us only very limited predictive power. (The most important factor would be the difficulty of that particular rebound.) On the other hand, if we were trying to predict the quality of the

---

( $r^2$ ), which is equal to the percentage of variance explained. The reader can obtain correlation coefficients by taking the square root of each percentage of variance given in the text.

<sup>9</sup> By 'broad traits,' we simply mean traits that produce a wide variety of different types of behavior. Belief in the existence of broad traits should be carefully distinguished from what Doris (2002) has called *globalism* — namely, belief in the existence of traits that are stable, evaluatively integrated, and yield consistent behavior. (A trait that explains, say, 9% of the variance in a wide range of morally-relevant behaviors could be extremely broad but would not yield consistent behavior and would therefore provide no evidence at all for globalism.) When we say that the existence of broad traits is no longer a matter of controversy in social and personality psychology, we certainly don't mean to imply that all psychologists are globalists. Far from it: as we shall see, trait-relevant behaviors are often surprisingly inconsistent.

player's overall performance across the course of an entire season of play — including numerous different kinds of tasks performed in a wide variety of situations — our ability measure would probably prove extremely useful. So should we say that ability has only a small impact on performance or that it has a very large impact? Ultimately, our answer will depend on the precise nature of our concern: whether we are concerned with success on one particular occasion or with success over the course of a whole season.

As Epstein (1979) has argued, a similar conundrum arises in the domain of moral psychology. For example, suppose we wanted to know whether a broad trait of 'honesty' can be used to predict the degree to which children will engage in a broad array of different kinds of honesty-related behaviors. If we try to predict just *one* such behavior on the basis of one other behavior, we obtain a correlation that explains only 5% of the behavioral variance. However, if we look at the overall honesty that a child shows across a whole battery of tests and then try to predict the honesty that the same child will show in another battery of tests, we obtain a much higher correlation — this time, explaining a full 81% of the variance (Hartshorne & May 1928).

So should we say that traits have only a small impact on behavior or that they have a very large impact? Here again, the answer will depend on the nature of our concern: whether we are concerned with one particular behavior or with a long sequence of behaviors performed over the course of many years. Doris (2002) and Harman (1999) have argued that traditional virtue ethics can only be tenable if we have some way to predict specific behaviors on the basis of broad personality traits. This is a powerful argument — and one for which we have considerable sympathy — but the issue remains controversial. A number of philosophers have argued that virtue ethics can still be viable

even in the face of the Doris-Harman critique (see, e.g., Kamtekar 2004; Merritt 2000; Sabini & Silver 2005).

Our aim here is not to resolve this controversy but rather to emphasize that the problem Doris and Harman have identified for virtue ethics does not also apply to Nietzsche's account. Since Nietzsche is interested in the *structure of a life*, and not in isolated, particular instances of conduct, it would seem that Epstein's approach offers strong support. What matters for Nietzsche is that heritable traits structure the *course of a life*, not that they enable one to predict any particular instance of conduct in that life. As we shall see in a moment, though heritable traits may not predict people's behavior on any individual occasion, a wide variety of studies show that they do have a quite substantial impact on the long-run path of an individual's life.

### ***C. The Role of Upbringing***

In contrast to Nietzsche, philosophers working in the Aristotelian tradition tend to assume that upbringing plays a major role in the shaping of people's character traits. Here it is essential to distinguish two related claims. First, there is the bland and relatively uncontentious claim that a person's environment has an important influence on his or her character. Second, there is the more specific and largely unsubstantiated claim that character is shaped by *upbringing*, i.e., by the ways in which person is treated by his or her parents or caregivers. This latter claim is usually put forward without argument, but as we shall see, recent empirical research gives us quite substantial reasons to be suspicious of it.

In thinking about this issue, it may be helpful once again to consider what percentage of the variance in personality traits is explained by each of a number of

different factors. We saw above that heredity explains around one-third to two-thirds of the variance in most traits, with the rest presumably explained by environmental factors. Our question now is: Of the variance explained by the environment, how much is explained by upbringing and how much is explained by other environmental factors?

To begin with, we should note that socialization researchers have uncovered numerous correlations between childrearing practices and personality development (e.g., in the classic studies of Baumrind 1967; 1991). In other words, it can be shown that children who have been raised in particular ways tend to have particular personality traits. But the existence of correlations is not in question here; the only question is about whether particular childrearing practices actually *cause* people to have particular personality traits. For example, it is widely assumed that there is a correlation whereby people who are beaten as children tend to be more violent as adults.<sup>10</sup> One possible explanation of this correlation would be that childhood beatings actually cause people to develop more violent personalities. But there are other plausible interpretations. It could be that certain people have more violent personalities even as children and that these people are more likely to misbehave and then to be beaten by their parents. Alternatively, it could be that a genetic propensity for violence is passed down from parents to children and that, since violent people are especially likely to have violent parents, such people are especially likely to be beaten as children.

The key contribution of behavioral genetics to this question has been in distinguishing between variance explained by the *shared environment* and variance

---

<sup>10</sup> Widom (1989) reviews dozens of studies on the etiology of violence and concludes that there is actually surprisingly little empirical support for this assumption. Still, the balance of evidence does seem to suggest a correlation between being beaten as a child and being violent as an adult, and we will assume for the sake of argument that the correlation is really there.

explained by the *non-shared environment*. The ‘shared environment’ is made up of those aspects of the environment that are shared by all children growing up in the same family, while the ‘non-shared environment’ is made up of those aspects of the environment that differ even between two children growing up in the same family. Thus, suppose that two children are brought up by the same parents but have different peer groups. The traits of the parents would then be part of the shared environment, while the traits of the peers would be part of the non-shared environment. We can now ask how much of the variance in personality traits is explained by the shared environment. The surprising answer is: *very little* (only 5% to 10% in most studies). This is truly a shocking result, but it has been replicated in an enormous variety of studies and is now the basis of a wide-ranging consensus among researchers (see, e.g., Bouchard 1994; Loehlin 1992; Plomin & Daniels 1987).

To see the force of this finding, it may be helpful to engage in a quick thought experiment. Suppose we know that a given child is going to be adopted by a pair of particularly kind, loving and open parents. What should we predict about the development of this child’s personality? The answer appears to be that our knowledge of the parents gives us almost no predictive power at all. If these parents adopt three different children, those three children will be hardly any more similar than three randomly selected individuals.

As usual, the findings obtained for morally-neutral personality traits hold for morally-relevant personality traits as well. We noted above that one recent study finds that 70% of the variance in children’s aggressiveness is explained by genetic differences. That same study finds that only 5% of the variance is explained by the shared

environment (Eley, Lichtenstein & Stevenson 1999). But as the authors themselves point out, this result is methodologically suspect, since the study had parents themselves assessing the degree to which their children behaved violently. When the violence of children is assessed by their teachers, heredity accounts for 49% of the variance and shared environment has no impact (Deater-Deckard & Plomin 1999). Of course, results like these do not call into question the widespread assumption that there is a correlation whereby violent parents are especially likely to rear violent children — but they do suggest another possible interpretation of that correlation. Perhaps the observed correlation has almost nothing to do with parents serving as ‘bad role models’ or ‘perpetuating a cycle of violence.’ The effect might be almost entirely genetic, the product of genetic similarity between parents and children.

Reading the works of behavioral geneticists, it is easy to get the impression that no study has ever found the shared environment to have a substantial impact on anything of importance. But that is not quite right. Some studies have indicated a substantial impact of shared environment; it’s just that the vast majority of studies have shown no substantial impact, and even when shared environment does have a substantial impact, this impact is usually far smaller than that of either heredity or nonshared environment.

For a case in which shared environment really has sometimes been shown to make a difference, let us consider the study of criminality. As one might expect, there is a correlation whereby criminal parents are more likely to have criminal children. But what explains this correlation — nature or nurture? To find the answer, we can look at studies of adopted children. Our question will be whether criminality in the children is best predicted by criminality in the adoptive parents or by criminality in the biological

parents. A number of early studies using this methodology found that criminality in the biological parents predicted criminality in the children but that criminality in the adoptive parents had no significant impact (Schulsinger 1972; Crowe 1974). Later studies, however, did show that children of criminal adoptive parents had somewhat higher rates of criminality. This is an important victory for the significance of shared environment. Yet, even here, the importance of genetics ends up dwarfing the importance of shared environment. To give one striking example, Cloninger and colleagues showed that children of criminal adoptive parents did have higher rates of criminality, but they also showed that children of criminal biological parents were *twice* as likely to become criminals as were children of criminal adoptive parents (Cloninger, Sigvardsson, Bohman & Knorrning, 1982). Thus, of the total explained variance, 59% was explained by the criminality of the biological parents and only 19% was explained by the criminality of the adoptive parents.

In light of the repeated failure of shared environment to explain a large portion of the variance in personality, we seem forced to choose between three possible views. One view would be that parental treatment has only a very small impact on the development of personality, with other environmental factors playing a much more important role (Harris 1995; 1998). A second view would be that, although the similar treatment received by children raised together has very little impact, the respects in which such children are raised differently actually do have considerable impact (Plomin & Daniels 1987). A third would be that the very same kinds of parental treatment can have radically different impacts on different kinds of children (Maccoby & Martin 1983). The debate

among these three views continues to rage on — with considerable theoretical sophistication (and a fair amount of animosity) being shown on all sides.

In sum, we have overwhelming evidence that heredity plays a major role in the shaping of personality, whereas the claim that upbringing plays a major role is contentious at best. It may somehow be possible to vindicate Aristotle's moral psychology against its Nietzschean rival, but in light of the empirical evidence, there is plainly no reason to think that the Aristotelian account is more plausible.

#### ***D. Conscious Decision and Behavior***

Recall that on the Kantian view, moral agents impose motivationally effective moral requirements upon themselves. This process of rational moral self-legislation is presumably a conscious one, and thus we must presume these consciously imposed moral "laws" to be morally effective. On the Nietzschean view, by contrast, conscious choice plays no such role in moral (or immoral) agency. Consider the case of a professor who devotes a great deal of time to her students. One explanation of the professor's behavior would be that she has a conscious belief about the importance of devoting time to one's students and that she is acting on that belief. This is the type of explanation that Nietzsche wants to reject; it is the mundane analogue of Cornaro's self-understanding, according to which it was his "free" choice to follow a certain kind of diet that explained his long life. A second type of explanation would be that the professor is simply the type of person who feels compelled to help her students and that, although she may have various conscious beliefs about how she ought to live, these beliefs have very little impact on the way she actually treats other people. It is this sort of explanation that one

frequently finds in Nietzsche's works and that we saw illustrated, earlier, in Nietzsche's account of Cornaro

It seems that one way to decide between these two types of explanations would be to see whether there were substantial correlations between certain types of conscious attitudes and certain types of behaviors. For example, we could check whether there was a correlation between the degree to which professors believed they were obligated to spend time with their students and the degree to which those professors actually did spend time with their students. After all, it does appear that we would have a certain kind of *prima facie* evidence that attitudes were influencing behavior if we found a substantial correlation here.

It should be clear, however, that this sort of test is not sufficient to settle the question. The mere existence of a correlation plainly does not establish causality. Just as it is possible that people's attitudes influence their behavior, it is possible that people's behavior influences their attitudes. Thus, it might turn out that certain professors just happen to be the kinds of people who spend time with their students (for reasons that have nothing to do with their conscious beliefs) and that these professors then come to have the belief that they have an obligation to spend time with their students as a result of the fact that they are already performing the relevant behavior.

Accordingly, we proceed in two steps – beginning with the question as to whether conscious attitudes are *correlated* with behavior and then asking whether conscious attitudes actually *cause* behavior.

First, let us consider the question of correlation. In the early decades of the twentieth century, most researchers simply assumed that attitudes were highly correlated

with behavior. It was assumed, for example, that any program that decreased racist attitudes would thereby also decrease racist behavior. This initial assumption was called into question by the influential work of LaPiere (1934). LaPiere went on a long car trip with a Chinese couple. Along the way, he took careful notes about how his companions were treated at each of the hotels and restaurants they visited. Despite the widespread prejudice against Chinese people in America at the time, LaPiere found that he and his companions were generally treated quite well and that they were refused service on only one occasion. Later, he wrote to all 250 hotels and restaurants listed in his notes, asking the employees whether or not they would be willing to serve Chinese guests. Over 90% of respondents said that they would not serve Chinese, in spite of the fact that they had just done exactly that. This finding seemed to suggest that attitudes and behavior were not quite as closely linked as had previously been thought.

The ensuing decades saw an enormous profusion of studies testing the degree to which attitudes and behavior were correlated. The results of this initial wave of research were extremely surprising. In almost every domain studied, the correlation between attitudes and behavior was shockingly low. By 1969, Wicker was able to draw on a wide variety of studies for the influential review in which he argued that there was little convincing evidence for a substantial attitude-behavior correlation (Wicker 1969).

Wicker's review served as a challenge to the next generation of researchers. The goal was to find specific circumstances in which attitudes truly were substantially correlated with behavior. As it happened, researchers were quite successful at this task – devising ever more clever ways to create a situation in which attitudes and behavior were correlated. (To give one particularly striking example, it has been shown that behavior is

more highly correlated with attitudes when subjects are looking at themselves in a mirror [Carver 1975].) In a summary of this next generation of research, Kraus reviewed 88 studies and showed that the attitude-behavior correlation was explaining, on average, 14% of the total variance (Kraus 1995).

As might be expected, there has been a fair amount of debate about whether a correlation of this size should be regarded as large or small (e.g., McGuire 1985; Kraus 1995). But the size of the correlation is not our primary concern here. Our concern is with the question as to whether or not attitudes actually *cause* behavior. If we find, for example, that there is a substantial correlation between attitudes toward a given race and actual behavior toward that race, we still cannot be sure whether the attitudes are causing the behavior, the behavior is causing the attitudes, or some third factor is causing both the attitudes and the behavior.

In fact, systematic experiments suggest that a substantial portion of the observed correlation is due to the impact of behavior on attitudes rather than other way around. For a simple example, consider the results reported in Fendrich (1967). Subjects were (a) given a questionnaire regarding their attitudes toward black people and (b) asked to participate in a meeting of the National Association for the Advancement of Colored People (NAACP), a civil rights group advocating for the interests of black people. The key question was whether there would be any correlation between subjects' attitudes (as measured by the questionnaire) and their behavior (actual participation in the meeting). There were two conditions in the experiment. In one condition, subjects were *first* given the questionnaire and *then* asked to participate in the meeting. In this first condition, there was no significant correlation between attitude and behavior – indicating that,

whatever attitude was measured by the questionnaire, that attitude had very little impact on people's actual attendance at NAACP meetings. The second condition was exactly the same as the first, except that the order of the tasks was reversed: subjects first decided whether or not to attend the meeting and then filled out a questionnaire regarding their attitudes toward black people. In this second condition, there was a significant and substantial correlation between attitude and behavior. The overall pattern of the results thus points to a surprising conclusion. In this experiment at least, it appears that attitudes had very little impact on behavior but *that behavior had a substantial impact on attitudes*. In particular, subjects appeared to be modifying their attitudes toward black people in such a way as to justify a prior decision to attend or not attend a meeting of a civil rights organization.

Fendrich's experiment is just one of the many that have demonstrated the surprising impact of behavior on attitudes. In a typical experiment of this type, psychologists find some way to manipulate subjects into performing a behavior that goes against their pre-existing attitudes. The result – as psychologists have found again and again – is that subjects modify their attitudes to fit the behavior they have been manipulated into performing. The examination of this phenomenon has been a major preoccupation of the field of social psychology, and a number of competing theories have been proposed to explain it (Aronson; Bem 1972; Festinger 1957; Steele 1988). Although there is no clear consensus as yet, the dominant view seems to be that people are motivated to believe that their own behaviors are justified and that they therefore tend to adopt attitudes that justify the behaviors they have already performed.<sup>11</sup>

---

<sup>11</sup> [Nietzsche quote from *Dawn* here]

Given that the correlation between attitudes and behavior is not overwhelmingly high and that a substantial portion of this correlation can be explained in terms of the impact of behavior on attitudes (rather than the other way around), a number of researchers have concluded that attitudes actually have only a very minimal influence on behavior. So, for example, Haidt (2001) has argued that, although people often have conscious attitudes regarding very general moral questions, these attitudes actually have little impact on people's feelings about the rightness or wrongness of specific acts. Perhaps people's feelings about specific acts are derived not from their conscious moral attitudes but rather from a set of non-conscious mental states (Wilson 2002). Thus, it might be thought that the degree to which an individual discriminates against black people is affected, not so much by that individual's conscious attitudes regarding black people in general, as by certain purely non-conscious prejudices over which the person's conscious attitudes have little causal influence.

But these results do raise another question: if conscious attitudes actually have very little impact on behavior, why do people so frequently assume otherwise? People quite often attribute their own behaviors to particular conscious attitudes, and many people take it to be simply *obvious* that their behaviors are caused by these attitudes. If the attitudes aren't actually causing the behaviors, we need some sort of hypothesis to explain this sense of obviousness.

#### ***E. The Conscious Experience of Will***

One plausible hypothesis would be that people actually have a conscious experience of taking certain considerations into account as they are trying to decide what to do. That is, people might be conscious of weighing various reasons and, on that basis,

choosing to perform an action. When they are asked to explain why they behaved as they did, they immediately assume that their behavior can be understood in terms of the very considerations they consciously considered while deliberating.

Here we find yet another area in which Nietzsche's interests overlap with those of contemporary experimental psychologists. It seems clear that people sometimes engage in a conscious process of deliberation followed by a conscious act of decision, but it is far from obvious that all of these conscious processes really have much impact on people's behavior. As Nietzsche often suggests, it might be that our behavior is actually determined by a struggle among unconscious drives and that the process of conscious deliberation has surprisingly little impact on what we ultimately end up doing (see D 109 and the discussion in Leiter [2002: 95-101]). This possibility has been investigated by a long tradition of psychologists, including Nisbett and Wilson (1977) and Gazzaniga and LeDoux (1978). Most recently, Wegner (2002) has reviewed much of the relevant evidence in his popular book *The Illusion of Conscious Will*.

Wegner, like Nietzsche, starts from the *experience* of willing, and, like Nietzsche, wants to undermine our confidence that the experience accurately tracks the causal reality. To do so, Wegner calls our attention to cases where the *phenomenology* and the *causation* admittedly come apart: one set of cases involve "illusions of control," that is, "instances in which people have the feeling they are doing something when they actually are not doing anything" (2002: 9) (think of a video game, in which you feel your manipulation of the joy stick explains the action on the screen, when in fact, the machine is just running a pre-set program). Another set of well-documented cases involve the "automatisms," that is cases where there is action but no "experience of will" (2002: 8-9)

(examples would include ouija board manipulation and behaviors under hypnotism).

Wegner remarks:

[T]he automatism and illusions of control...remind us that action and the feeling of doing are not locked together inevitably. They come apart often enough to make one wonder whether they may be produced by separate systems in the mind. The processes of mind that produce the experience of will may be quite distinct from the processes of mind that produce the action itself. (2002: 11)

If the cases in question do, indeed, show that phenomenology of willing is not always an accurate guide to causation, they certainly do not show that this is *generally* true. But Wegner wants to establish Nietzsche's more general claim (see Leiter 2005), namely, that the phenomenology of willing systematically misleads us as to the causation of our actions. And in the place of the "illusion of free will" as Wegner calls it, he proposes a different model according to which "both conscious willing and action are the effects of a common unconscious cause" (Holton 2004: 219), but the chain of causation does not run between the experience of willing and the action. Rather, in Nietzschean terms, some type-fact about persons explains both the experience *and* the action (Wegner 2002: 68).

As Wegner sums up his alternative picture of the causal genesis of action:

[U]nconscious and inscrutable mechanisms create both conscious thought about action and the action, and also produce the sense of will we experience by perceiving the thought as cause of the action. So, while our thoughts may have deep, important, and unconscious causal connections to our actions, the

experience of conscious will arises from a process that interprets these connections, not from the connections themselves. (Wegner 2002: 98)

Before we turn to the empirical evidence for this strong claim, it is important to recall an ambiguity in the causal story suggested by Nietzsche's example of Cornaro. On one reading, slow metabolism (the relevant type-fact about Cornaro) explains why Cornaro ate a slender diet, and the fact that he ate a slender diet explains his longevity. If we take this version as an analogue of willing, then the will is, indeed, causal, but it is not the *ultimate* cause of an action: something *causes* the experience of willing and then the will causes the action. Let us call this the view of *Will as Secondary Cause*.

On another reading, the slow metabolism explains *both* the slender diet and the longevity, but there is no causal link between the latter two. Let us call this the view of *Will as Epiphenomenal*. The Cornaro example itself most plausibly suggests the Will as Secondary Cause (surely the slender diet makes a *causal* contribution to the long life), yet other passages in Nietzsche (e.g., D 124) suggest the Will as Epiphenomenal doctrine instead: if the "I will" is really analogous to the person "who steps out of his room at the moment when the sun steps out of its room, and then says: '*I will* that the sun shall rise'" (D 124), then there is no causal link between the experience of willing and the resulting action, just as there is no causal link between the person who wills the sun to rise and the rising of the sun.

Nietzsche's texts on this subject are, we believe, generally ambiguous as to which view of the will he decisively embraces. Wegner's empirical evidence, by contrast, is offered in support of the Will as Epiphenomenal.<sup>12</sup> If Wegner is right, that is good

---

<sup>12</sup> [some ambiguity in Wegner on this score—note Holton's skepticism etc.]

reason, as a matter of interpretive charity, to read Nietzsche as committed to the Will as Epiphenomenal,<sup>13</sup> that is, to read him as holding the view that is (a) supported by his texts, and (b) most likely to be correct as a matter of empirical science.

Psychologists have adduced a variety of kinds of evidence to support the view of the Will as Epiphenomenal, and it will not be possible to review most of them here. Let us concentrate on a few illustrative bits. We begin with a phenomenon that is already well-known to students of the free will literature, but probably less familiar to those interested in Nietzsche. These are the studies by Benjamin Libet and colleagues (e.g., Libet 1985) examining the brain electrical activity (the “readiness potential” or “RP”) that precedes an action (such as moving a finger) and the experience of willing. The researchers found that subjects reported an experience of conscious will that took place at a significant interval *after* the onset of the RP. These findings seem to suggest that an unconscious process takes place first and that the person only then becomes conscious of wanting to perform the action. As Libet (1992: 269) puts it:

[T]he initiation of the voluntary act appears to be an unconscious cerebral process. Clearly, free will or free choice of whether *to act now* could not be the initiating agent, contrary to one widely held view. This is of course also contrary to each individual’s own introspective feeling that he/she consciously initiates such voluntary acts; this provides an important empirical example of the possibility that the subjective experience of a mental causality need not necessarily reflect the actual causative relationship between mental and brain events.

---

<sup>13</sup> In the terms used in Leiter (2002: \_\_-\_\_), this would be to read Nietzsche as embracing Token

This claim has been extremely controversial, and a number of authors have argued that it is possible to explain all of Libet's results without supposing that our behaviors are actually controlled by non-conscious processes (see, e.g., Mele forthcoming). We cannot hope to resolve this controversy here. Instead, we will turn to additional sources of evidence. These other sources of evidence do not involve RPs in any way, but they do provide strong reason to suppose that the conscious experiences accompanying decision-making often mislead us about the factors that truly impact our behavior.

Let us turn, then, to studies of 'split-brain' patients. These are patients who suffered from severe epilepsy and therefore underwent an operation that severed the corpus callosum that normally connects the left and right hemispheres of the brain. After the operation is complete, there is no communication between the two hemispheres; each is isolated from the other. For present purposes, the significance of this work comes from the fact that it is possible to construct a situation in which one hemisphere initiates a behavior and the other hemisphere is then given the task of providing an explanation for the behavior that has been performed. For example, if one places a stimulus in a certain portion of the visual field, it will be detected by the right hemisphere but not by the left. One can then use this technique to send the right hemisphere an order to get up and walk around. The right hemisphere will initiate the requested behavior, but the left hemisphere will have no idea why this behavior is being performed. Now comes the tricky part. The left hemisphere is home to most of the brain regions subserving verbal communication. So if one asks a subject why she is leaving the room, the response will most likely be generated by the left hemisphere. When questions like these are actually given to

---

Epiphenomenalism, contrary to the interpretation emphasized in that earlier work.

subjects, one finds a curious type of response. Subjects do not typically express puzzlement and say, 'I don't really know; I just suddenly found myself getting up.' Instead, they immediately offer a plausible explanation like, 'I thought I'd get up to get a drink.' (For discussion of these experiments, see Gazzaniga & LeDoux 1978.) The fact that split-brain patients are able to come up with these explanations so effortlessly provides some evidence for the view that the mechanism at work in split-brain patients is the same mechanism we see in normal people's attempts to explain their own behavior. That is, it might be that we never have direct conscious access to the factors that really influence our decisions. Perhaps we are always simply engaged in an effort to construct plausible explanations for the things we find ourselves doing.

Finally, there are studies in which researchers directly compare verbal reports based on conscious experience with experimental data about the factors that truly do explain the variance in behavior. In studies of this final type, researchers actually manipulate certain aspects of a situation and check to see whether these manipulations have any impact on the decisions people end up making. Thus, some subjects are placed in a situation that includes a particular factor  $x$ , while others are placed in a situation that is exactly like the first except for the absence of factor  $x$ . After deciding what to do in the situation, all subjects are asked the simple question 'Did factor  $x$  make any difference in your decision?' This task may appear to be a relatively easy one. After all, subjects have just had a conscious experience of taking certain considerations into account, and it seems that they should therefore be in a good position to know which factors played a key role in their decisions. Yet study after study shows that people find it extraordinarily difficult to answer questions like this one.

One of the first experiments in this vein was performed by Maier (1931). Subjects were asked to solve a puzzle that involved moving certain pieces of string. The one factor that most facilitated performance on this task was a subtle movement in the string subjects were supposed to be moving, but the subjects themselves resolutely insisted that this factor had no influence at all on their decisions. Instead, they claimed that they had been greatly helped by a ‘hint’ that involved a weight spinning on another string (despite the fact that systematic experimental study showed that this hint was entirely useless).

Maier’s study is not just one isolated case. A great many experiments have been run using this basic design, and almost all of them yield the same pattern of results. Nisbett and Wilson (1977) review dozens of studies in this vein and conclude that conscious experience actually gives people almost no insight into the real causes of their own behavior. To the extent that people do know why they perform the behaviors they do, it is mostly because people are making use of information that would also be available to outside observers.

Putting all of these sources of evidence together, we arrive at a strong case for a single basic thesis. It seems that people do sometimes have the conscious experience of making decisions on the basis of certain considerations, but that this sort of experience can sometimes mislead people about the true causes of their behavior. Thus, the mere fact that people have a conscious experience of choosing to perform a behavior on the basis of certain moral principles does not show that these moral principles actually have any causal impact on the behaviors people perform (as the Kantians require). The only way to know whether a given factor actually has any causal impact is to perform

systematic experiments, and as we have seen, these experiments tend to suggest that consciously held moral principles have very little influence on behavior.

## **V. Conclusion**

We have been concerned with three rival views in moral psychology — one that emphasizes habits acquired through childhood upbringing, one that emphasizes conscious moral principles, and one that emphasizes heritable psychological traits. Philosophers have devoted considerable time and thought to the first two of these views, and partisans of each view have shown great theoretical sophistication in clarifying the relevant concepts and working out the key ethical implications. But this philosophical ingenuity is never accompanied by any empirical evidence showing that the factors under discussion actually play any important role in people's lives, and when one looks to the empirical literature, one finds shockingly little evidence that either childhood upbringing or conscious moral principles have a substantial impact on people's moral behavior. It seems likely, then, that much of the recent work on these issues has been taken up with an attempt to work out the implications of a moral psychology that is not actually instantiated in real human beings.

By contrast, the third view — the one that we find in Nietzsche — turns out to be the one that garners the most support from a growing body of empirical evidence. This evidence suggests that heritable psychological traits influence many aspects of people's lives, including their moral behavior. Investigation into the philosophical implications of this psychological account has only begun in recent years (see, e.g., Leiter 2002). Such an investigation would, in many ways, resemble the projects we find in the Aristotelian and Kantian traditions, but it would differ in that it would be concerned with the

implications of a hypothesis that seems quite likely to turn out to have a sound basis in empirical reality.

## References

- Aronson, E. 1969. "The Theory of Cognitive Dissonance: A Current Perspective," in *Advances in Experimental Social Psychology*, vol. 4, ed. L. Berkowitz, (New York: Academic Press), pp. 2–34.
- Baumrind, D. 1967. "Child Care Practices Antecedent Three Patterns of Preschool Behavior," *Genetic Psychology Monographs* 75: 43-88.
- Baumrind, D. 1991. "The Influence of Parenting Style on Adolescent Competence and Substance Use," *Journal of Early Adolescence* 11: 56-95.
- Bem, D. J. 1972. "Self-Perception Theory," in *Advances in Experimental Social Psychology*, vol. 6, ed. L. Berkowitz, (New York: Academic Press), 1-62.
- Bem, D. J., & Allen, A. 1974. "On Predicting Some of The People Some of the Time: The Search for Cross-Situational Consistencies in Behavior," *Psychological Review* 81: 506-520.
- Bouchard T. J. 1994. "Genes, Environment and Personality," *Science* 264:1700-1701.
- Carver, C. S. 1975. "Physical Aggression as a Function of Objective Self-Awareness and Attitudes toward Punishment," *Journal of Experimental Social Psychology* 11: 510-519.
- Cloninger, C. R., Sigvardsson, S., Bohman, M., & Knorrning, A. 1982. "Predisposition to Petty Criminality in Swedish Adoptees," *Archives of General Psychiatry* 39: 1242-1247.

- Cooper, John M. 1975. *Reason and Human Good in Aristotle* (Cambridge, Mass.: Harvard University Press). Page references are to the reprint edition (Indianapolis: Hackett, 1986).
- Crowe, R. 1974. "An Adoption Study of Antisocial Personality," *Archives of General Psychiatry* 31: 785-791.
- Darley, J. M. & Latané, B. 1973. "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 27, 100-108.
- Deater-Deckard, K., & Plomin, R. 1999. "An Adoption Study of the Etiology of Teacher Reports of Externalizing Problems in Middle Childhood," *Child Development* 70: 144-154.
- Deigh, John. 1996. *The Sources of Moral Agency: Essays in Moral Psychology and Freudian Theory* (Cambridge: Cambridge University Press).
- Doris, John M. 2002. *Lack of Character: Personality and Moral Behavior* (Cambridge: Cambridge University Press).
- Edelbrock, C., Rende, R. D., Plomin, R., & Thompson, L. A. 1995. "A Twin Study of Competence and Problem Behavior in Childhood and Early Adolescence," *Journal of Child Psychology and Psychiatry* 36: 775-785.
- Eley, Thalia, Lichtenstein, Paul & Stevenson, Jim. 1999. "Sex Differences in the Etiology of Aggressive and Nonaggressive Antisocial Behavior: Results from Two Twin Studies," *Child Development* 70: 155-168.
- Epstein, Seymour. 1979. "The Stability of Behavior: On Predicting Most of the People Much of the Time," *Journal of Personality and Social Psychology* 37: 1097-1126.

- Feingold, Alan. 1992. "Good-Looking People Are Not What We Think," *Psychological Bulletin* 111: 304-341.
- Fendrich, James. 1967. "A Study of the Association among Verbal Attitudes, Commitment and Overt Behavior in Different Experimental Situations," *Social Forces*, 45: 347-355.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA : Stanford University Press.
- Festinger, L & Carlsmith, J. M. 1959. "Cognitive Consequences of Forced Compliance." *Journal of Abnormal and Social Psychology* 58, 203-210.
- Funder, D. C. & Ozer, D. J. 1983. "Behavior as a Function of the Situation." *Journal of Personality and Social Psychology* 44:107-112.
- Gazzaniga, M & LeDoux, J. (1978). *The Integrated Mind*. New York: Plenum Press.
- Haidt, J. 2001. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108: 814-834.
- Harris, Judith Rich. 1995. "Where is the Child's Environment? A Group Socialization Theory of Development," *Psychological Review* 102: 458-489.
- Harris, Judith Rich. 1998. *The Nurture Assumption: Why Children Turn Out the Way They Do*. (New York: Free Press).
- Harman, Gilbert. 1999. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 99: 315-331.
- Hartshorne, H. & May, M.A. 1928. *Studies in Deceit* (New York: McMillan).

- Hoffman, Lois. 1991. "The Influence of the Family Environment on Personality: Accounting for Sibling Differences," *Psychological Bulletin* 110: 187-203.
- Holton, Richard. 2004. Review of *The Illusion of Conscious Will* by Daniel Wegner, *Mind* 113: 218-21.
- Hutchings, B. & Mednick, S. 1975. "Registered Criminality in the Adoptive and Biological Parents of Registered Male Criminal Adoptees," in *Genetic Research and Psychiatry*, eds. Fieve, R. R., Rosenthal, D., & Brill, H. (Baltimore: Johns Hopkins University Press), 105-116.
- Jones, E. E. & Nisbett, R. E. 1972. "The Actor and the Observer: Divergent Perceptions of the Causes of Behavior," in *Attribution: Perceiving the Causes of Behavior*, eds. E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins and B. Weiner (Morristown, NJ: General Learning Press), 79-94.
- Kamtekar, R. 2004. 'Situationism and Virtue Ethics on the Content of Our Character,' *Ethics*, 114.
- Korsgaard, Christine. 1996. *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press).
- Kraus, Stephen. 1995. "Attitudes and the Prediction of Behavior: A Meta-Analysis of the Empirical Literature," *Personality and Social Psychology Bulletin* 21: 58-75.
- Kraut, Richard. 2001. "Aristotle's Ethics," in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL: <http://plato.stanford.edu/entries/aristotle-ethics/>
- LaPiere, R. 1934. "Attitudes vs. Actions," *Social Forces* 13: 230-237.
- Leiter, Brian. 2002. *Nietzsche on Morality* (London: Routledge).
- . 2005. "Nietzsche's Theory of the Will," unpublished manuscript.

- Libet, B. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *The Behavioral and Brain Science* 8: 529-566.
- Libet, B. 1992. "The Neural Time-Factor in Perception Volition, and Free Will." *Revue de Métaphysique et de Morale* 97: 255-272.
- Loehlin, J. C. 1992. *Genes and Environment in Personality and Development* (Newberry Park, CA: Sage).
- Maccoby, E. E., & Martin, J. A. 1983. "Socialization in the Context of the Family: Parent-Child Interaction," in *Handbook of Child Psychology*, vol. 4, *Socialization, Personality, and Social Development*. ed. E. M. Hetherington (New York: Wiley).
- Maier, N. (1931). "Reasoning in Humans II: The Solution of a Problem and its Appearance in Consciousness." *Journal of Comparative Psychology*, 12, 181-194.
- McGuire, W. J. 1985. "Attitudes and Attitude Change," in *The Handbook of Social Psychology*, eds. G. Lindzey & E. Aronson (New York: Random House), 238-241.
- Mele, A. Forthcoming. "Decisions, Intentions, Urges, and Free Will: Why Libet Has Not Shown What He Says He Has," in J. Campbell, M. O'Rourke, and D. Shier, eds. *Explanation and Causation: Topics in Contemporary Philosophy* (Cambridge, MA: MIT Press).
- Merritt, M. 2000. "Virtue Ethics and Situationist Personality Psychology," *Ethical Theory and Moral Practice* 3:365-383.
- Milgram, S. *Obedience to Authority*. New York: Harper Colophon, 1975.
- Mischel, Walter. 1968. *Personality and Assessment* (New York: Wiley).

- Nichols, Shaun. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment* (New York: Oxford University Press).
- Nisbett, R., & Wilson, T. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Plomin, Robert & Daniels, Denise. 1987. "Why are Children in the Same Family So Different from One Another?" *Behavioral & Brain Sciences* 10:1-16.
- Prinz, Jesse. 2006. *The Emotional Construction of Morals* (Oxford: Oxford University Press).
- Ross, Lee, & Nisbett, Richard. 1991. *The Person and the Situation: Perspectives of Social Psychology* (New York: McGraw Hill).
- Sabini, John & Maury Silver. 2005. "Lack of Character? Situationism Critiqued," *Ethics* 115: 535-562.
- Schmitz, S., Fulker, D. W., & Mrazek, D. A. 1995. "Problem Behavior in Early and Middle Childhood: An Initial Behavior Genetic Analysis," *Journal of Child Psychology and Psychiatry* 36: 1443-1458.
- Schulsinger, F. 1972. "Psychopathy: Heredity and Environment," *International Journal of Mental Health* 1: 190-206.
- Schneewind, J.B. 1992. "Autonomy, Obligation, and Virtue: An Overview of Kant's Moral Philosophy," in *The Cambridge Companion to Kant*, ed. Paul Guyer (Cambridge: Cambridge University Press).
- Steele, C. M. 1988. "The Psychology of Self-Affirmation: Sustaining the Integrity of the Self," *Advances in Experimental Social Psychology*, vol. 21, ed. L. Berkowitz (New York: Academic Press), 261-302.

- Wegner, D. 2002. *The Illusion of Conscious Will* (Cambridge, Mass: MIT Press).
- Westen, Drew. 1998. "The Scientific Legacy of Sigmund Freud: Toward a Psychodynamically Informed Psychological Science," *Psychological Bulletin* 124: 333-371.
- Wicker, A. W. 1969. "Attitudes vs. Actions: The Relationship of Verbal and Overt Behavioural Responses to Attitude Objects," *Journal of Social Issues* 22: 41-78.
- Wilson, T. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious* (Cambridge, Mass.: Harvard University Press).