Document generated on 10/11/2024 11:48 p.m.

Les ateliers de l'éthique
The Ethics Forum

Letting the World In: Empirical Approaches to Ethics

Joseph Heath

Volume 7, Number 3, Fall 2012

URI: https://id.erudit.org/iderudit/1014387ar
DOI: https://doi.org/10.7202/1014387ar

See table of contents

Publisher(s)

Centre de recherche en éthique de l'Université de Montréal

ISSN

1718-9977 (digital)

Explore this journal

Cite this article

Heath, J. (2012). Letting the World In: Empirical Approaches to Ethics. *Les ateliers de l'éthique / The Ethics Forum*, *7*(3), 93–107. https://doi.org/10.7202/1014387ar

Tous droits réservés © Centre de recherche en éthique de l'Université de Montréal, 2012

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

LETTING THE WORLD IN: EMPIRICAL APPROACHES TO ETHICS

JOSEPH HEATH
UNIVERSITY OF TORONTO

In a landmark summary and review of "fin de siècle" moral philosophy, published in the final decade of the 20th century, three major theorists in the field – Stephen Darwall, Allan Gibbard and Peter Railton – lamented the fact that "too many moral philosophers and commentators on moral philosophy — we do not exempt ourselves — have been content to invent their psychology or anthropology from scratch"[1]. While welcoming the advent of empirically and historically informed approaches to the subject, they noted that "any real revolution in ethics stemming from the infusion of a more empirically informed understanding of psychology, anthropology, or history must hurry if it is to arrive in time to be part of *fin de siècle* ethics"[2]. While the revolution that they anticipated did not arrive quite as quickly as hoped, the first decade of the 21st century did finally bring about a change, in the general direction that they had called for. While some moral philosophers continue to see themselves as legislators in the kingdom of ends, engaged in a purely normative project, this view has increasingly come under pressure from those who find in the growing empirical literature on morality, not only a challenge to many preconceived philosophical notions, but also material for fruitful new approaches to a multitude of time-honoured philosophical problems.

Of course, to talk about "empirically and historically" informed approaches to moral philosophy must not be taken to imply any sort of unified vision or approach. The social sciences are divided into rival schools of thought just as philosophy is. There is a mountain of theory and data, much of it relevant to our understanding of morality, but coming from different domains of inquiry, and in some cases involving contradictory theoretical commitments. There are of course the staples of 20th century social science, areas that have been well-recognized fields of study for several decades: Anthropological ethnography remains an incredibly rich source of material for philosophers interested in cultural pluralism. Empirical research on child development has for several decades provided a wealth of information, particularly about the relationship between social cognition and moral emotions. Personality theory in psychology (including also studies in abnormal psychology) has attempted to operationalize many of our folk-psychological ideas about character. And of course evolutionary theory has provided the overarching set of constraints under which any naturalistic theory of morality must be developed.

Beyond this, the late 20th century saw several developments in the human sciences that significantly increased the volume of philosophically relevant work being done, particularly in the area of moral psychology. Probably the most important developments were the result of the so-called "cognitive revolution" in psychology, where the taboo imposed by behaviourism on the investigation of mental states was lifted, and psychologists began to investigate seriously all aspects of cognition, including moral judgment and reasoning[3]. The result was an extraordinary expansion in our understanding of judgment and decision-making, cognitive bias, intuition, memory, executive control and will-power, to name just a few areas. At the same time, significant developments in evolutionary theory, largely the consequence of the development and application of the tools of evolutionary game theory, led to a sharpening in our understanding of several issues crucial to our understanding of morality (inclusive fitness, group selection, culture-gene co-evolution, etc.)[4]. And finally the development and application of more powerful statistical methods led to significant progress in the area of personality theory, allowing that speciality to leave behind its psychoanalytic origins and to become a theoretically freestanding domain of social-scientific inquiry[5].

My objective here is not to provide a systematic overview of these trends, but to pick out and describe three major challenges that have arisen to traditional moral philosophy as a result of developments in the human sciences. Stated somewhat hyperbolically, these are as follows: first, there is the claim that moral virtues, as traditionally conceived, do not exist; second, that moral intuitions are unreliable and easily manipulated; and finally, that human morality could not have evolved through any straightforward extension of mechanisms that produce altruism in other species. I will deal with each of these claims in turn. Beyond the details of these claims, however, there is a more general point that I would hope to make. For those who have been following it, the results of 20th century social science have been extremely disruptive to many aspects of our everyday self-understanding. Psychological research, in particular, consistently suggests that we are not just radically ignorant of how the mind works, but that much of what we glean through introspection is quite mistaken[6]. Such a finding, painstakingly detailed in study after study, cannot possibly leave philosophy unchanged. The idea that moral philosophers can go about their business, picking up where Henry Sidgwick left off, suggests a failure to grasp the import of much of the progress of human knowledge over the course of the past century. So while the transformation that Darwall, Gibbard and Railton called for is certainly underway, the idea that empirical findings might serve as merely a supplement to traditional philosophical methods, rather than a challenge, is beginning to seem increasingly implausible.

1. VIRTUE THEORY

The suggestion that there is something wrong with the folk theory of character traits underlying traditional Aristotelian virtue theory has been around for a long time. As early as the 1920s, psychologists found that traits like "honesty" or "compassion" seemed to have weak predictive value, and could easily be over-

whelmed by situational factors[7]. (Since psychologists have such easy access to large numbers of students, some of the earliest studies were on cheating and plagiarism. This resulted in "honesty" being the most exhaustively studied trait.) Some philosophers — Gilbert Harman and Owen Flanagan in particular — noted the significance of this work early on, but the majority of moral philosophers ignored it[8]. The turning point came with the publication of John Doris's book, *Lack of Character*, which presented the data in a systematic and accessible way, but more importantly, framed it in a way that made the underlying challenge to virtue theory impossible to dismiss through mere hand-waving[9].

There are two major aspects of the psychological research on character that are significant. First is the finding that it is difficult to isolate character traits at the medium level of generality posited by traditional virtue theory (such as courage, generosity, or honesty). There is no question that individuals have habits, or specific scripts that they follow in particular situations. For example, at a store when the cashier making change mistakenly hands over too much money, many people, when they notice the error, will automatically give it back. This is typically a settled disposition, often developed at an early age. The problem is that it doesn't correlate in any significant way with behaviour in other sorts of situations, even ones that are only slightly different[10]. Many people, for instance, when they notice that an item has scanned wrong, and that they have been charged less than the marked price, will let the error stand. Again, this may be a very settled disposition. The important point is that knowing how a person behaves in one sort of situation tells you *nothing at all* about how the person will behave in the other sort of situation. And so if you ask the question, "Is this person honest?" the answer will be "It depends." Even the microcategory of "willingness to take advantage of employee error in order to pay less than full price for goods" is still too general to capture the morally significant dimension of the individual's behavioural dispositions.

There is, however, some evidence of the existence of character traits, as traditionally conceived, at a much *higher* level of generality than that posited by virtue theory. Psychologists often talk about the "big five" personality traits: openness, conscientiousness, extraversion, agreeableness and neuroticism[11]. There are two things that are striking about these traits: first, they have no particular moral valence, and are therefore not "virtues" in anything like the classical sense. Second, they *do* have cross-situational predictive validity (weak, but not non-existent)[12]. So, for example, a test used to determine how introverted or extraverted a person is might ask a question such as, "When the telephone rings, does it usually make you feel excited or anxious?" Knowing how a person answers this and other related questions would actually be useful in helping to predict how that person is likely to behave in other situations, such as entering a room full of strangers. This is *precisely* the sort of cross-situational predictive validity that was always assumed for the classic Arisotelian (and later Christian) moral virtues, but which empirical research has subsequently failed to demonstrate.

The second major blow to virtue theory came from a raft of studies showing that, while individuals often focus on personality traits that have no predictive value, they often ignore factors that are of major importance in determining whether people act morally. Perhaps the most celebrated discovery is the power of conformity, imitation, and the associated set of social expectations, in determining conduct[13]. Whatever particular habits or scripts people may have, these can quite easily be overridden not just by what the group does, but even by perceptions or hints as to what the group may do, or what the expectations are[14]. Yet the underlying motive is not one that sits easily with traditional virtue theory, partly because it contains no substantive moral content or orientation toward the good. It also suggests that people are much more strongly influenced by their *current* social environment than the environment in which they were raised, and so the emphasis that virtue theory puts upon socialization winds up being, if not mistaken, then at least deeply misleading[15].

Finally, an integral component of virtue theory is the account that it offers of *vice*. In the same way that people are thought to do good things because of some excellence of character, they are also thought to do bad things because of some defect of character. The superficial plausibility of this hypothesis led to it being the subject of investigation by criminologists, who spent decades studying convicted criminals, trying to determine what sort of personality traits might distinguish them from the general population. Again, the results came back negative for most of the traditional categories. There are some differences — for instance, criminals show a slightly higher level of impulsiveness than the general population — but nothing with the type of substantive content that might qualify it as a vice[16]. Investigators also found that criminals had, for the most part, conventional moral views about what was good and what was bad[17]. There were, however, interesting differences, such as the fact that criminals are far more likely to make self-serving use of conventional excuses[18]. (Thus they typically endorse the general norms under which their conduct would be classified as immoral, but then neutralize the force of these norms, exempting themselves from the judgment through some sort of rationalization.)

None of this is strictly speaking inconsistent with virtue theory. Nevertheless, much of it is highly unexpected from that perspective. Things that virtue theorists thought should make a big difference turn out to make no difference, while things that should have made no difference at all turn out to make a very large difference, when it comes to determining whether individuals will act morally or immorally. It perhaps worth observing that despite the large number of factions within the field of personality theory and social psychology, there is no "Aristotelian" (or even "neo-Aristotelian") school of thought. This may explain why, in the increasingly voluminous philosophical literature on the subject, the dominant tone among virtue theorists has been highly defensive[19]. Rather than mining the empirical research to find work that supports their view, most of the defences that have been offered simply take the theory in the direction of unfalsifiability, often by loosening up the connection between the supposed virtues and actual behavior.[20]

2. MORAL INTUITION

One of the most striking features of 20th-century Anglo-american moral philosophy has been the heavy reliance upon appeal to "moral intuition" as a basis of philosophical argumentation (combined with widespread failure to state explicitly what the status of these intuitions is, or where they are supposed to come from). There are, of course, many other areas of human decision-making where individuals rely upon intuitions to guide judgment. Much of the upshot of 20th century psychological investigation, however, concerns the *unreliability* of these intuitions. Intuitions are typically the result of heuristic problem-solving processes, rigidly adapted to deal with particular challenges that arose in the environment of evolutionary adaptation. Not only do they contain bugs, but they also tend to misfire when deployed in non-standard environments[21]. For example, we have a rather developed system of intuitive physics, which we use to calculate the trajectory of objects in motion through the air, in earth-standard gravity[22]. It contains a notorious bug, however, in that it disregards forward momentum in calculating the trajectory of dropped objects, and therefore anticipates that they will travel straight down. As a result, the untutored will consistently miscalculate the landing point of, say, bombs dropped from airplanes.

The investigation of these heuristics formed the basis for perhaps the best-known project in empirical psychology, centered on the work of Daniel Kahneman and Amos Tversky, who showed that our "intuitions" in a wide range of problem-solving domains are fraught with error. For moral philosophers, this generates an enormous sceptical problem, which most have been loathe to confront. Cass Sunstein was, if not the first, then at least the most explicit in throwing down the gauntlet:

> I believe that some philosophical analysis, based on exotic moral dilemmas, is inadvertently and even comically replicating the early work of Kahneman and Tversky by uncovering situations in which intuitions, normally quite sensible, turn out to misfire. The irony is that where Kahneman and Tversky meant to devise cases that would demonstrate the misfiring, some philosophers develop exotic cases with the thought that the intuitions are likely to be reliable and should form the building blocks for sound moral judgments[23].

Derek Parfit's work is perhaps a paradigmatic example of the tendency that Sunstein condemns, since it relies so heavily on a direct appeal to intuitions, often in highly decontextualized cases. Consider, for instance, Parfit's argument in *Reasons and Persons* against the temporal discounting of harms in a consequentialist calculus[24]. He develops an example involving a person who leaves broken glass lying in the woods. One hundred years later, a child walking in woods cuts herself. Parfit suggests that this act is wrong, regardless of how long it takes for the harm to occur — that the temporal distance between the agent and the victim makes no difference. (He compares this to a person who shoots an

arrow into the woods, and injures someone far away, unseen. Here the spatial distance would seem to make no difference, in assessing the wrongness of the act. He then claims that "remoteness in time has, in itself, no more significance than remoteness in space"[25].)

The example is very compelling. And yet the issue of temporal discounting is one that has been extensively studied by economists, who have a professional interest in determining what sort of attitudes people have toward tradeoffs between present and future. The empirical research suggests that the way that the problem is *framed* is what is doing most of the work in Parfit's argument. Using different scenarios, economists have been able to elicit preferences that imply positive, negative, and zero social discount rates[26]. In one influential study, Maureen Cropper, Sema Aydede and Paul Portney found that people were willing to sacrifice the lives of 45 people in a hundred years, in order to save one life in the present[27]. On the other hand, people can also be induced to assign greater value to future lives[28]. Cass Sunstein and Richard Thaler have argued, on this basis, that for anyone who has surveyed the literature, "the most sensible conclusion is that people do not have robust, well-ordered intergenerational time preferences"[29]. Because we almost never have to think about future people, we simply don't have a moral intuition with respect to temporal discounting, which isn't to say that we can't be induced to have one.

At very least, this example suggests that when a particular frame ("a child cuts herself on broken glass") elicits a particular intuition, but another frame elicits some other, some defence of the frame needs to be offered, some account of why this is the right way of posing the problem[30]. (Why is it a child, and not a middle-aged hunter?) And yet, as Walter Sinnott-Armstrong has argued, this shows that one cannot be justified in trusting a moral intuition non-inferentially[31]. The intuition that *p* is wrong cannot itself justify that claim; one must show that the frame that elicits the intuition is also somehow the correct one. This is, as Sinnott-Armstrong observes, not impossible. It simply means that intuitions must play a less central role in philosophical reflection than they have to date.

There are several other objections to the reliance on intuitions that have been thematized of late. First of all, there was the widely discussed intervention by Joshua Greene and his collaborators, who did a series of fMRI studies on individuals, looking to see which regions of the brain were activated as they contemplated various moral dilemmas[32]. Greene made the suggestion that the deontological "intuitions" individuals deployed were essentially emotional responses, whereas the consequentialist ones relied on parts of the brain associated with reasoning and calculation. This served as the basis for his provocative suggestion that deontological judgments, far from being driven by moral reasoning, are actually just rationalizations of emotional responses.

There were some obvious problems with the argument[33]. In particular, Greene offered no reason to think that the theory of value *underlying* the consequentialist

calculus was not based on the same sort of emotional reactions. In this respect, what he was really doing was presenting an essentially sceptical challenge to moral reasoning in general, yet optimistically assuming that it undermined only the position of his opponents. At the same time, the sceptical challenge is a significant one, particularly for moral philosophers who place great weight on the authority of their intuitions. What Greene's challenge relied upon, in part, is the wealth of psychological evidence revealing the limits of subjective experience when it comes to understanding our own reasoning processes[34]. The central accusation — that "deontology... is a kind of moral confabulation"[35] — can't be settled by introspection, or by any other armchair method. Starting with the early studies on confabulation in the 1960s, psychologists have established, quite firmly, that we cannot tell through introspection whether reasoning is actually guiding our conduct or not[36]. This is indeed one of the most disconcerting findings of 20th century psychology: sometimes we actually control our conduct through decision, and sometimes we just make up stories to explain our behaviour after the fact, and yet we can't tell through introspection when we're doing one or the other[37].

Finally, it is worth mentioning the most long-standing difficulty with moral intuitions, which is that many philosophers continue to treat them as though they offered insight into the nature of reality, or some sort of universal moral truth, despite ample evidence that these intuitions are not shared by those outside the circles of what Pierre Bourdieu referred to as *homo academicus*. This has of course long been obvious to anyone interested in either history or non-Western culture[38]. The rise of behavioural economics, however, has produced a body of very well-controlled studies, examining how people around the world respond to a set of social interactions involving cooperation and principles of fairness. Psychologists Joseph Henrich, Steven Heine and Ara Norenzayan have coined the acronym WEIRD societies (Western, Educated, Industrialized, Rich and Democratic) to describe the background of individuals who make up the sample group for most psychology studies. They go on to show, as the titles suggests, that in multiple dimensions, WEIRD peoples are distinct outliers with respect to a wide range of judgments (and that Americans are distinct outliers within the class of WEIRD peoples). They conclude that "members of WEIRD societies, including young children, are among the least representative populations one could find for generalizing about humans" on a wide range of topics, including "visual perception, fairness, cooperation, spatial reasoning, categorization and inferential induction, moral reasoning, reasoning styles, self-concepts and related motivations..."[39] Two factors that they point to, in particular, include the unusual nature of Western, especially American, child-rearing practices, along with the high levels of urbanization. ("Since such urban environments are highly 'unnatural' from the perspective of human evolutionary history, many conclusions drawn from subjects reared in such informationally impoverished environments must remain rather tentative"[40].)

Of course, psychological studies have also shown that when people are told that people just like them are subject to a particular bias, they automatically interpret this to mean "everyone just like me, except me, is subject to this bias"[41]. Thus moral philosophers will no doubt go on treating themselves as something more than just native informants, reporting on the finer details of bourgeois liberal morality. Should they at some point start to treat the problem more seriously, however, it seems an inevitable consequence that "intuitions" will come to play a less authoritative, or at least less central, role in moral philosophy. For instance, there may come a time when the "method of reflective equilibrium" and its variants come to be seen as a discredited approach to the development of moral theory.

3. EVOLUTIONARY THEORY

While "evolutionary ethics" is not a particularly fruitful or even influential research program, most moral philosophers nevertheless would like to think of themselves as developing theories that are consistent with a scientific worldview. Thus there is a general desire to produce theories that are, if not inspired by an evolutionary perspective, then at least compatible with one. And indeed, many moral philosophers, even those whose central preoccupations lie elsewhere, nevertheless make some effort to explain how the moral psychology that they posit could have arisen through natural selection. This is certainly the case with, among others, Peter Singer[42], John Mackie[43], Peter Railton[44], Allan Gibbard[45] and even Derek Parfit[46].

Evolutionary theory, however, is not a fixed point of reference, but rather a highly dynamic field of inquiry, which has undergone significant changes over the course of the 20th century. Two developments in particular were of obvious significance for moral philosophy. First of all, there was the powerful attack launched against group selection theory by George Williams in the 1960s, which showed that traits could not be explained on the grounds that they were "good for the species," without some further specification of how this translated into a reproductive advantage at the level of the individual[47]. This line of reasoning was further amplified with the development of evolutionary game theory, most influentially in the work of John Maynard Smith[48]. Given the role that the prisoner's dilemma played in driving interest in rationality-based game theory, evolutionary theorists acquired a much more vivid awareness of the importance of free-riders problems, and of the unlikelihood that any evolutionary system would achieve an optimal outcome. Both of these trends made it clear that altruism required some sort of special explanation.

The second major development was of course the emergence and popularization of "inclusive fitness" and "selfish gene" theory. Despite some of Richard Dawkins's rhetoric, what his argument suggests is that the "selfishness" promoted by natural selection is going to occur at the level of the *gene*, not that of the individual organism[49]. This puts to rest the old philosophical chestnut about whether true altruism is possible, or whether all action is at some level self-interested. If we are just robots, constructed by our genes in order to advance *their*

interests, then it follows almost immediately that the gene should be willing to sacrifice the robot entirely, if by doing so it is able to benefit copies of itself found elsewhere in the environment. This is, of course, the logic of kin selection. It is an elementary implication of selfish gene theory that natural selection can produce organisms capable of what is, from the perspective of the *individual*, true altruism. Thus Michael Ghislain's famous slogan, "scratch an altruist, watch a hypocrite bleed," is shown to be inconsistent with an evolutionary perspective[50].

The flip side of the selfish gene/inclusive fitness perspective, however, is the expectation that altruism should be limited in scope. Careful study of social insects — bees, wasps, ants — showed that they have a non-standard reproductive biology, which increases the coefficient of relatedness between female members of a colony or hive, and therefore helps explain the wider range of prosocial behavior (and, *inter alia*, the complex division of labour) that they exhibit. The effect of these discoveries was to heighten the mystery surrounding human cooperation, precisely because we seem to lack any such mechanism. As Robert Boyd and Peter Richerson put it:

> Humans are, arguably, a new page in the natural history of animal cooperation. Our reproductive biology is similar to other social mammals. Among our close relatives, the apes and monkeys, genetic relatedness and reciprocal altruism support a diverse array of small-scale societies, but no other spectacular ones. Humans have built complex societies by some mechanism or mechanisms different from any other known highly social species. At the same time, there are remarkable parallels between human and ape social behavior and material culture, not to mention many convergences between humans and other social and tool-using species. Consistent with classical comparative anatomy and modern molecular studies, human behavior is clearly recently derived from ape behavior[51].

The first hypothesis to attract widespread attention was that the mechanism of reciprocity, which had been posited to explain certain symbioses in nature, could be adapted to explain human ultrasociality. Much of this early discussion, however, has come to seem overly simplistic in retrospect. Robert Trivers, for example, in his classic article on reciprocal altruism, picks out three examples to discuss[52]. The first is the famous example of "cleaning symbioses" among fish, the second is warning calls among birds, but the third is the "psychological system underlying human reciprocal altruism"[53]. This hypothesis — that human ultrasociality is just an extension of the same mechanism that one sees on display in other species — has not fared well. First, there is the fact that the mechanism is not very robust. It seems to be more suited to promoting dyadic cooperation than cooperation in large groups[54]. Second, growing scepticism has developed about the extent of reciprocal altruism in nature (as opposed to "byproduct mutualism," or "pseudoreciprocity")[55]. In particular, the problem of how these systems get started has come to seem rather serious. (There would have to be a first

mover. But if reciprocity is required, in order to make the behaviour beneficial, how could anyone ever make the first move?) Thus the current literature is characterized by widespread scepticism about the ability of reciprocity to sustain much of the limited altruism that one finds in other animal species, much less the ultrasociality one finds among humans.

The upshot of these debates is significant because, among the philosophers who try to provide some account of the evolutionary plausibility of their views, the most common strategy has been to appeal to some form of reciprocity as the source of the benefits that could render moral behaviour adaptive. This is true from John Mackie in the 1970s to Richard Joyce today[56]. The inclination here is understandable enough. When one contemplates the enormous benefits that come from the systems of cooperation at the heart of human society, it seems natural to suppose that these benefits must provide an explanation for the behavioural dispositions that make them possible. And yet contemporary evolutionary models show that this is far from obvious.

In part for these reasons, the most state-of-the-art discussions of morality in an evolutionary context take place within the framework of so-called gene-culture co-evolutionary theory[57]. The range of hypotheses that have been presented under this rubric is formidable, and the literature is developing quickly. In part this is due to the realization that human culture-dependence is also a somewhat mysterious phenomenon, from an evolutionary perspective. The benefits of cultural learning are enormous, and so again there is a temptation to think that these *must* provide some sort of an adaptive explanation of the behavioural dispositions that facilitate their emergence. But since the benefits are largely those of *cumulative* cultural transmission, they cannot explain why the first generation to exhibit the required dispositions (e.g. the social learning algorithms) derived any benefit from them. Thus culture, far from providing any sort of a solution to the problem about the origins of morality, has become part of a complex of human traits that stand in need of explanation.

All of this has bolstered the impression that the mechanism that sustains human ultrasociality is going to be extremely non-obvious. The mere fact that it doesn't occur elsewhere in nature already suggests this. For philosophers, this means that evolutionary theory is going to constrain hypothesis-formation with respect to our understanding of morality much more than was earlier assumed, since merely pointing to the benefits of cooperation as an "adaptive" explanation for the phenomenon of moral constraint can no longer suffice.

CONCLUSION

Although the development of more empirically-informed approaches to the study of morality has caused some anxiety in philosophical circles, particularly among those who feel that it threatens their intellectual property values, it is nevertheless difficult to regard it as anything less than a welcome development. Empirical research can no doubt serve as an important stimulus to philosophical

reflection, but it can also provide a measure of intellectual *discipline* that philosophers sometimes lack, with their reliance upon thought-experiments, introspective psychology, and just-so stories. At the same time, it is not just philosophers who stand to benefit from closer contact and exchange with the human sciences. The empirical literature is itself often vague in its use of terminology, and glosses over important distinctions. Even something as elementary as the difference between deontology or consequentialism, or the right and the good, is not always sharply drawn. Thus there is significant room for mutually beneficial collaboration between philosophers and those who are trained in empirical research methods.

From what I have said so far, it should be obvious that I regard the overall trend toward increased contact with the human sciences to be salutary. I would like to close this discussion, however, on a cautionary note. Although it is all well and good for philosophers to go out and read the latest social science, most philosophers have very weak training in scientific methodology, and (in my experience) tend not to have a lot of respect for it either. (For example, the educational background of most philosophers is typically "humanistic," and so includes, for example, no training in the use of statistics.) Thus when dealing with social science, it is important for philosophers to recognize that serious investigators take methodological concerns very seriously. For example, as a general consequence of the commitment to scientific method there is an emphasis upon falsification as that is almost entirely absent among philosophers. (Most philosophers are able to explain at length why their own view is correct, but are hard-pressed to specify what would show it to be *incorrect.*)

This produces two dangers. On the one hand, it can lead philosophers into flippant dismissal of inconvenient results. Often this is based on a failure to appreciate how much work social scientists put into control and testing. I myself have witnessed the following interaction, which I think illustrates the danger. After a listening to a research presentation by a psychologist, a philosopher asks (in a tone of triumphant refutation) "How do you know that people didn't interpret the question this way, rather than that?" The psychologist replies, "Because we ran a series of focus groups, in multiple regions, and tried out several formulations before settling on the wording of the question." He hadn't mentioned this in the presentation, because he simply took it for granted that one would test a survey instrument before deploying it, and that the audience would know this as well. Philosophers, however, sometimes think that social-scientific results can be defeated just by imagining other possible interpretations, failing to realize that a lot of time and energy may have been spent ruling out those interpretations.

On the other hand, a lack of methodological sophistication can generate the opposite problem, where philosophers are far too credulous in accepting experimental results. For those who are enthusiastic about empirical approaches, there is a temptation to treat the results of all published studies as simply "facts," regardless of such things as sample size, strength of the correlations, and suc-

cessful replication. Again, speaking to psychologists, it still surprises me when I mention a particular experimental result, only to have someone say, "We tried for years to replicate that and couldn't." This is, of course, the type of information you get through the oral culture of the discipline, since negative results are seldom published. Experts in the field have a nuanced sense of what is believable and what isn't, because they evaluate particular studies against a vast background knowledge of what else has been demonstrated, and how particular research programs have played out over time. Thus there are real pitfalls for the philosophical explorer, reading about in an unfamiliar literature, since it can be difficult to determine how believable a given study is when taken in relative isolation.

The point of these reflections is to suggest that, while there is an enormous amount to be gained from the development of more empirically informed approaches to moral philosophy, there is also a lot to be said for the development of more collaborative approaches to the investigation of these questions. It is important for philosophers to be cautious when assessing experimental results, without being glib in dismissing them. Finding this balance can be difficult, which is precisely why it can be useful to engage more seriously with those whose disciplinary training lies precisely in the assessment and interpretation of these results

NOTES

1 Stephen Darwall; Allan Gibbard; Peter Railton, "Toward Fin de Siecle Ethics: Some Trends," *The Philosophical Review*, 101 (1992): 115-189, p.188.

2 Darwall, Gibbard and Railton, "Toward Fin de Siecle Ethics," p. 189.

3 For short summary, see George A. Miller, 'The Cognitive Revolution: A Historical Perspective," *Trends in Cognitive Science*, 7 (2003): 141-144.

4 For short summary, see Martin A. Nowak, "Five Rules for the Evolution of Cooperation," *Science*, 314 (2006): 1560-1563.

5 See Oliver P. John and Sanjay Srivastava, "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives," in Lawrence A. Pervin and Oliver P. John, *Handbook of Personality*, 2nd edn. (New York: Guilford, 1999).

6 Daniel Kahneman has recently provided a wonderful summary of much of this research, in *Thinking Fast and Slow* (Toronto: Doubleday, 2011).

7 Hugh Hartshorne and Mark A. May, *Studies in the Nature of Character, Vol. 1 Studies in Deceit* (New York: MacMillan, 1928). The classic presentation of the situationist view is Lee Ross and Richard E. Nisbett, *The Person and the Situation* (Philadelphia: Temple University Press, 1991).

8 Gilbert Harman, "Moral Philosophy meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society*, 99, (1999):315-331; Owen Flanagan, *Varieties of Moral Personality: Ethics and Psychological Realism* (Cambridge, MA: Harvard University Press, 1991). See also Peter Railton, "Made in the Shade: Moral Compatibilism and the Aims of Moral Theory," in Jocelyn Couture and Kai Nielsen, eds. *On the Relevance of Metaethics: Canadian Journal of Philosophy, Supplementary Volume 21* (Calgary: University of Calgary Press, 1995), pp. 79-106.

9 John Doris, *Lack of Character* (Cambridge: Cambridge University Press, 2002).

10 Or more specifically, as David C. Funder observes, "The relevance of personality for behavior is revealed not by a lack of change across situations but by the maintenance of individual differences. Such maintenance is indexed by the correlation coefficient, which reflects the degree to which subjects who perform a given behavior more often than others do in one situation also do so relatively more in a second situation," *Personality Judgment* (San Diego: Academic Press, 1999), p. 42.

11 McCrae RR, Costa PT., Jr. "Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model." In: Wiggins JS, ed. *The Five-Factor Model of Personality: Theoretical Perspectives*. (New York: Guilford; 1996). pp. 51–87

12 See, e.g., Debbie S. Moskowitz, "Cross-Situational Generality and the Interpersonal Circumplex," *Journal of Personality and Social Psychology*, 66 (1994):921–933. For overview, see Bernardo J. Carducci, *The Psychology of Personality,* 2nd edn, (Oxford: Wiley-Blackwell, 2009), pp. 298-301.

13 The most influential set of experiments on conformity were by Solomon Asch,. For summary, see Solomon E. Asch, "Opinions and Social Pressure," *Scientific American*, 193 (1955): 31-35. See also Elliot Turiel, *The Culture of Morality* (Cambridge: Cambridge University Press, 2002).

14 For a demonstration of how expectations affect behaviour discussed by both Harman and Doris, see John M. Darley and C. Daniel Batson, "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior." *Journal of Personality & Social Psychology*. 27 (1973): 100-108.

15 This issue was, of course, put on the agenda most forcefully by Stanley Milgram's experiments on obedience to authority. The literature on this is voluminous, but for a balanced survey, see Thomas 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.

16 For discussion with respect to white collar crime, see James Coleman, *The Criminal Elite* (New York: St. Martin's, 1989), pp. 200-204.

17 John Braithwaite, *Crime, Shame and Reintegration* (Cambridge: Cambridge University Press, 1989), p. 23 and passim, for a useful survey of relevant literature).

18 Gresham K. Sykes and David Matza. "Techniques of Neutralization: A Theory of Delinquency." *American Sociological Review* 22 (1957): 667-70.

19 For two prominent replies to the situationist challenge, see Rachana Kamtekar, "Situationism and Virtue Ethics on the Content of our Character." *Ethics,* 114 (3): 458-491 and Gopal Sreenivasan, "Errors about Errors: Virtue Theory and Trait Attribution." *Mind* (441): 47-68,

20 For an example of this, see Julia Annas, "Comments on John Doris's 'Lack of Character'," *Philosophy and Phenomenological Research*, 71 (2005): 636-642, at 637-38.

21 For useful review of the "misfiring" literature, see Keith Stanovich, *The Robot's Rebellion* (Chicago: University of Chicago Press, 2004), pp. 31-79

22 Mark McCloskey, "Intuitive Physics," *Scientific American*, 248 (1983): 122-130.

23 Cass Sunstein, "Moral Heuristics," Behavioral and Brain Sciences, 28 (2005): 531-573 at 541.

24 Derek Parfit, *Reasons and Persons*, (Oxford: Clarendon, 1986), pp. 356-57.

25 Parfit, *Reasons and Persons*, p. 357.

26 For some example of dependence upon the frame, see Shane Frederick, "*Measuring Intergenerational Time Preference:* Are *Future Lives Valued Less?" Journal of Risk and Uncertainty*, 26 (2003): 39-53.

27 Maureen Cropper, Sema Aydede and Paul Portney, "Preferences for Life Saving Programs: How the Public Discounts Time and Age," Journal of Risk and Uncertainty, 8 (1994): 243-265. See also Maureen L. Cropper, Sema K. Aydede, and Paul R. Portney, *Rates of Time Preference for Saving Lives*, 82 Am Econ Rev 469, 472 (1992).

28 Frederick, "*Measuring Intergenerational Time Preference," p. 45. For discussion, see Cass R. Sunstein and Richard H. Thaler, "*Libertarian Paternalism Is Not an Oxymoron," *University of Chicago Law Review*, 70 (2003): 1159-1202 at *1180, note 80.*

29 Sunstein and Thaler, "Libertarian Paternalism," p. 1180.

30 One can begin to see some push-back against the more extreme instances of reliance upon moral intuition. One reviewer, for instance, took issue with Frances Kamm's insistence that her readers take "as fixed data points in moral reflection" extremely fine-grained intuitions. "Her intuitions are exquisitely sensitive to the specific physical properties of trolleys, bombs, bridges, pills, knives, doctors, transplanted organs, sore throats, headaches, bathtubs, Lazy Susans, life-support machines, snow shovels, islands, yachts, et cetera." Jeffrey Brand-Ballard, review of "F.M. Kamm, Intricate Ethics: Rights, Responsibilities, and Permissible Harm," in *Notre Dame Philosophical Reviews* (http://ndpr.nd.edu/news/25299-intricate-ethics-rights-responsibilities-and-permissible-harm/).

31 Walter Sinnott-Armstrong, "Framing Moral Intuitions," in Walter Sinnott-Armstrong, ed. *Moral Psychology, Vol. 2* (Cambridge, MA: MIT Press, 2008), p. 68.

32 See Joshua D. Greene, "The Secret Joke of Kant's Soul," in Walter Sinnott-Armstrong, ed. *Moral Psychology, Vol. 3* (Cambridge, MA: MIT Press, 2008), pp. 35-79.

33 Most critics of Greene seem quite satisfied with the response in Selim Berker, "The Normative Insignificance of Neuroscience," *Philosophy and Public Affairs*, 37 (2009): 293-329.

34 See Greene, "Secret Joke," pp. 61-63. Also see Kahneman, *Thinking Fast and Slow*, p. 124.

35 Greene, "Secret Joke," p. 63.

36 See Timothy Willson, *Strangers to Ourselves* (Cambridge, Mass: Harvard Univesity Press, Belknap, 2002)

37 William Hirstein, *Brain Fiction* (Cambridge, MA: MIT Press, 2005). For an argument that we are at least sometimes doing the latter: see Roy F. Baumeister, E.J. Masicampo and Kathleen D. Vohs, "Do Conscious Thoughts Cause Behavior?" *Annual Review of Psychology*, 62 (2011): 331-361.

38 It is also one of the reasons that French and German philosophers tend not to appeal to moral intuitions – the influence of historicist doctrines, post-modernism, modesty about the universality of Western rationalism.

39 Joseph Henrich, Steven J. Heine, and Ara Norenzayan. "The Weirdest People in the World?" *Behavioral and Brain Sciences*,, 33 (2010): 61–135 at.61.

40 Ibid, p. 67.

41 Kahneman, *Thinking Fast and Slow,* p. 173.

42 Peter Singer, *The Expanding Circle: Ethics, Evolution, and Moral Progress*. (Princeton, NJ: Princeton University Press, 2011)

43 J.L. Mackie, *Ethics: Inventing Right and Wrong*, p. 113.

44 Railton, "Moral Camouflage or Moral Monkeys," The Stone (July 18, 2010): opinionator.blogs.nytimes.com/2010/07/18/moral-camouflage-or-moral-monkeys/

45 Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990), p. 64-68.

46 Derek Parfit, *On What Matters, Vol. 2* (Oxford: Oxford University Press, 2010), p. 494-97.

47 George C. Williams, *Adaptation and Natural Selection* (Princeton, NJ: Princeton University Press, 1966).

48 John Maynard Smith, *Evolution and the Theory of Games* (Cambridge: Cambridge University Press, 1982).

49 Richard Dawkins, *The Selfish Gene,* 2nd ed. (Oxford: Oxford University Press, 1989), p. 2.

50 Michael T. Ghiselin, *The Economy of Nature and the Evolution of Sex* (Berkeley, CA: University of California Press, 1974). For further discussion, see Franz de Waal, *Primates and Philosophers: How Morality Evolved* (Princeton, NJ: Princeton University Press, 2006), pp. 9-16.

51 Peter Richerson and Robert Boyd, "The Evolution of Human Ultra-Sociality," in Irenas Eibl-Eibesfeldt and Frank Salter, eds., *Ideology, Warfare, and Indoctrinability* (New York:: Berghan Books, 1998), pp. 71–95 at 72.

30 Robert Trivers, "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology*, 46 (1971): 35-57.

30 Trivers, "Evolution of Reciprocal Altruism," p. 47.

54 Robert Boyd and Peter Richerson, "The Evolution of Reciprocity in Sizable Groups," in *The Origin and Evolution of Cultures* (Oxford: Oxford University Press, 2005), pp. 145-165.

55 Olof Leimar and Richard C. Connor, "By-product Benefits, Reciprocity, and Pseudoreciprocity in Mutualism," in Peter Hammerstein, *Genetic and Cultural Evolution of Cooperation* (Cambridge, MA: MIT Press, 2003), pp. 203-222.

56 Mackie, *Ethics*, p. 113. Richard Joyce, *The Evolution of Morality* (Cambridge, MA: MIT Press, 2006), pp. 140-141.

57 For an early statement of this view, see Robert Boyd and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985). For a more recent statement, see Herbert Gintis, *The Bounds of Reason* (Princeton, NJ. Princeton University Press, 2009).