

Experimental Philosophy

Kwame Anthony Appiah

Presidential Address, Eastern Division APA December 2007

© Kwame Anthony Appiah 2008

I

Some three score years ago, the Norwegian philosopher Arne Naess found himself dissatisfied with “what are called ‘theories of truth’ in philosophical literature.” “The discussion has already lasted some 2500 years,” he wrote. “The number of participants amounts to a thousand, and the number of articles and books devoted to the discussion is much greater.” In this great ocean of words, he went on, the philosophers had often made bold statements about what “the man in the street” or “Das Volk” or “la conscience humaine” made of truth or Wahrheit or vérité. And Naess had a few simple questions about these claims: “How do the philosophers *know* these things? What is the source of their knowledge? What have they done to arrive at it? ... their writings,” he complained, “contain almost nothing of this matter.”¹ And so Naess began the research that resulted in the publication in 1938 of his first book in English: *“Truth” As Conceived By Those Who Are Not Professional Philosophers*.

Naess’s tone is one of irritated astonishment. “Even superficial questioning of non-philosophers would make it almost impossible for anyone to believe that the philosophers writing about the opinion of ordinary people actually ask others than themselves.... Have the philosophers any interest in writing on a subject capable of empirical treatment without knowing anything about it?” He proposed to do better. His study began by recovering what he called the “verbal reactions,”² both oral and written, of non-philosophers of varying “degrees of philosophical virginity,”³ to a series of questionnaires designed to elicit their views as to the “common characteristics” of the things that were true. The subjects he called *ps*, the common characteristics *c.c.s*, and he referred to himself throughout as I (for Leader). Thousands of hours and 250 *ps* later, I was able to report on the “fundamental formulations” in his subjects’ answers to the questionnaires about the *c.c.s*.

There are scores of distinct formulations ranging from (1) “agreement with reality” to (98) “that one is ready to defend the statement, to direct one’s behaviour according to it.” I patiently examines whether there are statistical correlations among these choices and the

¹ Arne Naess *“Truth” As Conceived By Those Who Are Not Professional Philosophers* (Oslo: I Kommissjon Hos Jacob Dybward, 1938): 12–15. The French passage is on 164. (The original book has the author’s name as Ness; but he is best known under the spelling Naess.)

² Naess *op. cit.* 17–18.

³ Naess *op. cit.* 45.

age, sex, class or educational level of the ps.⁴ I also reproduces brief excerpts of some of his oral examinations, of which the first begins:

l: What is the c.c. of that which is true?

p: — silence —

l: Have those things anything in common?

p: That is not certain.

l: Is it quite incidental, when you in some situations use the word “true”?

p: It is probably founded on something or other.⁵

It would be a serious project to establish what, among the portentous behaviorism of the talk of “verbal reactions,” the solemn abbreviations of every technical term, the mock indignation of l’s comments and the splendid silliness of the interviews, would get the average Norwegian to crack a smile. I myself, as a sample of one, would report that the monograph is hilarious. It is worth re-examination—I am very grateful to Larry Hardin for drawing it to my attention—for this reason alone.

But Naess was making a serious point. It really is an interesting question how we should decide what ordinary people think “truth” means. And it turns out that just asking them produces a vast and indigestible mess. What you have to do, as Naess showed, is to sort, interpret and analyze what people say, and then reflect. Towards the end of the book Naess proposes a name for this enterprise:

The diversity and consistency of amateur theories of truth, point to the possibility of an “experimental philosophy”.⁶

Naess went on to a distinguished philosophical career, writing about skepticism, applied semantics, and the history of philosophy; but his most influential contribution, as philosophers of the environment will be aware, was the invention of the notion of deep ecology, and he was a leader among Norway’s environmental activists at a time when their ranks were thinner. (Aestheticians and others with an interest in music will surely know him as the uncle of Arne Naess Jr., who was for 15 years the husband of Diana Ross.) But I’d like to talk today about the prospects of his suggestion—which was largely ignored outside Oslo—that it was time to take up experimental philosophy.

II

⁴ Naess *op. cit.* 42–45.

⁵ Naess *op. cit.* 32.

⁶ Naess *op. cit.* 161.

Or should I say: take it up again? Because it's a fair question to ask whether experimental philosophy is a matter of innovation or of restoration. In philosophy, perhaps more than any other discipline, we say what we are by telling stories about what we were: hence Naess's pointed reference to a 2500 year-old lineage. But most philosophers look back in ways that are quite unhistorical. We pick and choose among those who carried the label in the past without taking much notice of the people around them who also bore it, people whose work would not belong in our current stories of what we do. And we even pick and chose among the works of those we acknowledge as ancestors, as well. Newton and Locke were both called "philosophers" in the English of their day, but we don't claim Newton; even though, in saying "hypotheses non fingo," he was apparently announcing himself to be an anti-realist about gravitational theory. And we don't take much notice of many of Locke's contributions to the work of the Royal Society, either; even though he would no doubt have been simply puzzled by someone who said this was not philosophy.

There are reasons why we proceed this way, of course. Our contemporary ideas about what makes a question or an answer philosophically interesting are at work in these processes of editing out of the past the stories we choose to tell. Now, you might think that we could explain what philosophy is without telling these stories: that we could say what it is for a question to be philosophical, independently of these stories, and then explain that, in looking back, we are looking for past answers to the philosophical questions, and claiming as ancestors those whose answers were interesting or otherwise valuable contributions. I myself doubt that this sort of approach will do—I doubt the prospects of defining a trans-historical essence for philosophy and looking for the history of that essential subject back through the millennia. Disciplinary identities, so it seems to me, are like many other what you might call historical identities: to decide who is entitled to the label today you need a story about who had it yesterday; and whoever has it tomorrow will be continuing—with modifications—the projects of whoever has it today.

Ernest Renan, the great French historian and nationalist, made a similar point about *national* identities in his lecture "*Qu'est-ce qu'une nation?*" some 125 years ago. "Forgetting," he wrote, "and I would even say historical error, is an essential element in the creation of a nation, and that is why the progress of historical studies is often a danger for the nation itself."⁷ What he meant was that the stories of the past that served the modern national identity needed to leave out certain things that had actually happened in order that they could hold the nation together. Perhaps something like this is true of our discipline as well; but we can

⁷ Ernest Renan, *Qu'est-ce qu'une nation?*, 1882, Chapter 1, para 7. See: <http://www.bmlisieux.com/archives/nation02.htm>. Translation mine.

hope that philosophers, unlike the French, can construct stories that do their work without falsifying the past. What we cannot hope for, certainly, is an unedited past, a story that forgets nothing. And in the editing we will be guided, willy–nilly, by our sense not only of where we are but also of where we need to go. As Renan also said:

The Spartan song: “We are what you were; we will be what you are” is in its simplicity the abridged hymn of every country.”⁸

This is, *mutatis mutandis*, a truth about disciplines as well. The crafting of a disciplinary history matters, in part, because disciplinary identities are also contested (as I do not have to remind the members of the Eastern Division). In editing the past—even the very recent past—you foreground some contemporary questions and diminish the claim of others. Kant or Hegel? Frege or Husserl? Russell or Heidegger? Kripke or Derrida? But most pressingly, we seek, in our ancestry, the contour of an overarching identity that will distinguish us from our neighbors.

III

Consider, for example, just one of those neighbors: psychology. It’s often said that psychology has a short history and a long past. Might the opposite be true of philosophy, taken as an endeavor that sits apart from the realms of empirical research? In a 1668 treatise called *Observations upon Experimental Philosophy*, Margaret Cavendish urged that “the experimental part of philosophy” was not to be “preferred before the speculative,” for “most experiments have their rise from the speculative.”⁹

Still, it’s significant that, in those days, few of our canonical forebears confined themselves to the realm of unsullied abstraction. Descartes spent plenty of time with his hands inside freshly slaughtered cows, and his physiological findings were scarcely marginal to his thought; indeed, without the pineal—as the Danish anatomist Nicholas Steno pointed out in 1669—Descartes has no story of how mind and body are functionally integrated.¹⁰ By the next century, the growing prestige of experimentation was apparent everywhere. The encyclopedist D’Alembert praised Locke for reducing metaphysics to what it should be: *la physique expérimentale de l’âme*—the experimental science of the spirit.¹¹ And

⁸ Renan *op. cit.* Chapter 3, para 1 See:

http://ourworld.computer.com/homepages/bib_lisieux/nation04.htm

⁹ Margaret Cavendish, *Observations Upon Experimental Philosophy* Eileen O’Neil ed. (Cambridge: Cambridge University Press, 2001), 44, 49

¹⁰ René Descartes, *The Passions of the Soul*, in *Selected Philosophical Writings*, J. Cottingham, R. Stoothoff, and D. Murdoch trans. (Cambridge: Cambridge University Press, 1988), 331.

¹¹ The fifth edition of the *Dictionnaire de l’Académie Française* (1798) defines “âme” as “Ce qui est le principe de la vie dans tous les êtres vivans.”

Hume subtitled his great *Treatise of Human Nature*, as we do not sufficiently often recall, *Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*.

Some have seen the epistemological preoccupations of modern philosophy presaged in the work of Thomas Reid, whose “common-sense” school looks like a forerunner to the past century’s ordinary-language philosophy. But Reid himself was emphatic in his suspicion of mere conjecture. Every real discovery, he says, is arrived at by “patient observation, by accurate experiments, or by conclusions drawn by strict reasoning from observation and experiments, and such discoveries have always tended to refute, but not to confirm, the theories and hypotheses which ingenious men had invented.”¹² And Kant, to whom we owe the analytic-synthetic distinction, worked avidly on both sides of the putative divide; Herder revered him mainly for his lectures in geography.

The point is not just that the canonical philosophers belong as much to the history of what we now call psychology as to the genealogy of philosophy. It’s that you would have had a hard time explaining to them that *this* part of their work was *echt* philosophy and *that* part of their work was not. Trying to separate their “metaphysical” from their psychological claims is, I fear, rather like trying to peel a raspberry.

And though we typically suppose that psychology calved off from philosophy, you can make a case that it was the other way around. The psychology labs at Harvard are in William James Hall because its inhabitants rightly think of James (who migrated from Harvard’s physiology department to its philosophy department in 1881) as one of their ancestors, just as we contemporary philosophers claim him for ourselves. His colleague Josiah Royce was elected president of the American Psychological Association in 1902, and president of the American Philosophical Association in 1903. The common germline is visible in the history of our professional journals as well. When the philosophical quarterly *Mind* was founded, in 1878, and for a couple of decades afterward, much of it was devoted to articles we would now consider to be psychology. Even in the years following the rise of experimental psychology, a habit of intimacy was presumed. *The Journal of Philosophy* was founded, in 1904, as *The Journal of Philosophy, Psychology, and Scientific Methods*.

“Physique” is defined as “Science qui a pour objet les choses naturelles.” (This dictionary is one of many older French dictionaries available online from the Project for American and French Research on the Treasury of the French Language (ARTFL) *Dictionnaires d’autrefois* site: <http://portail.atilf.fr/dictionnaires/onelook.htm>.)

¹² Thomas Reid, *Essays on the Intellectual Powers of Man*, Essay I., iii, “Of Hypothesis,” in *The Works of Thomas Reid*, Vol. I (New York: Published by N. Bangs and T. Mason, for the Methodist Episcopal Church, J. and J. Harper, Printers, 1822), 367–8.

Modern philosophy has its origins in a sort of self-exile: the powerful swerve away from psychologism that began in the latter half of the nineteenth century—here the touchstones are Bolzano, Frege, Husserl—and that culminated in the golden age of conceptual analysis ... *and* also, I should add, of phenomenology. We know that experimentalists and pure speculators coexisted in the Harvard philosophy department until the mid-nineteen-thirties, and, give or take a few years; you find a similar pattern at other universities. I was educated at Cambridge University, where the degree in philosophy had not long ago before been called the Moral Sciences Tripos, and experimental psychology was still one of the fields in which you could choose to be examined for that degree. To see when what we would recognize as philosophy departments came into existence, look to see when psychology departments came into existence.

Indeed, we should be hard-pressed to establish to everyone's satisfaction that we noble philosophers, and not those knavish psychologists, are the legitimate heirs to what mainly went under the name philosophy in previous eras. These slippery movements of group designations are familiar to all historians, not to mention any sports fan who has watched the Jets, who used to be the Titans, play the Titans, who used to be the Oilers.

IV

So, as Renan might have asked: *Qu'est-ce qu'un philosophe?* In the heyday of analytic philosophy—the decades before and after the Second World War—the answer went like this. Philosophy is now what the best philosophy has always been: conceptual analysis. The claim about what philosophy had always been at its best was underwritten by some reference to Plato's *Theaetetus*, say, or Descartes *Meditations*, each construed as attempts at analysis of the concept of "knowledge." From there you would proceed by identifying conceptual analysis with exploring meanings. Since meanings are what speakers know in understanding their language, any speaker of a language knows already, without looking beyond her own linguistic competence, what she needs to know to do the analysis. Philosophical claims—knowledge is justified true belief, say—are true (*if* true) in virtue of the meanings of the words they contain. This way of doing philosophy presupposed, obviously, theories about concept-possession and knowledge of language.

"Conceptual analysis" was the examination not of just any old concepts but of the important ones; most of these were familiar from the earlier history of philosophy. They were to be explored in an essentially a priori way. The older conceptual analysts would have agreed with Timothy Williamson, when he said in his presidential address to the Aristotelian Society a few years ago, "If anything can be pursued in an

armchair, philosophy can.”¹³ You considered, then, not how things *are* but how we think about them, more or less *however* they are; and the only access you really had to how we *think* was to notice some of the patterns in what we do and don’t *say*.

Though the “we” here was meant to be all those who had a native-speaker’s grasp of the language, the actual conversations were naturally discussions among philosophers. W. H. Auden once wrote,

Oxbridge philosophers, to be cursory,
Are products of a middle-class nursery:
Their arguments are anent
What Nanny really meant.¹⁴

Now, Auden, like many poets, had a wonderful ear for other people’s language; so what he had in mind, I suspect, was the *sound* of philosophers at Oxford, when he was Professor of Poetry there from 1956 to 1961. And, though comic verse does not aspire to be either accurate or fair, he had a point. “Suppose I did say ‘the cat is on the mat’ when it is not the case that I believe that the cat is on the mat, what should we say?” asked J. L. Austin, one of Auden’s leading philosophical contemporaries in the Oxford of those years, in his 1955 William James lectures at Harvard.¹⁵ “What is to be said of the statement,” he went on “that ‘John’s children are all bald’ if made when John has no children?” (I don’t know about you, but that gets my nanny to a T.) For Austin, “What is to be said?” was not an invitation to collect data about how a given population of persons might make sense of these statements. The answer was supposed to be obvious. What a person knows in knowing her language is what everyone competent speaker *should* say in a certain situation; and so, being competent myself, I know what every competent speaker *would* say. That is why it wouldn’t matter if we found someone who *didn’t* say it: it would just show she wasn’t competent.

Philosophy gained an institutional aerie all to itself just about the time the theory of meaning on which it was based was being subjected to sustained assault by leading exponents of the analytic tradition itself; most influentially in the work of W. V. O. Quine, who persuaded many people that the idea of an analytic truth, the idea that a sentence could be true in virtue solely of the meanings of the words it contained, was mistaken. Belief in analyticity was, Quine famously argued, one of the

¹³ Timothy Williamson “Armchair philosophy, metaphysical modality and counterfactual thinking” (Presidential Address), *Proceedings of the Aristotelian Society* 105, 1 (2005): 1–23.

¹⁴ W. H. Auden, *Academic Graffiti* (New York: Random House, 1972), 25.

¹⁵ Published as J. L. Austin *How to do Things with Words* (Oxford: Clarendon Press, 1962), 50.

“dogmas of empiricism”; epistemology, in his view, “simply falls into place as a chapter of psychology and hence of natural science.”¹⁶

Anti-psychologism, which had enabled philosophers to hivel themselves off from experimental science, is now just one position in philosophy among others. The separation of philosophy from the empirical succeeded as an institutional project, but faltered as an intellectual one. Michael Dummett has written that certain errors of Frege and Husserl “have left philosophy open to a renewed incursion from psychology, under the banner of ‘cognitive science’. The strategies of defense employed by Husserl and Frege will no longer serve: the invaders can be repelled only by correcting the failings of the positive theories of those two pioneers...”¹⁷ Dummett’s bellicose rhetoric, even if tongue-in-cheek, suggests a genuine measure of unease, but no such strategy for repelling the marauders has gained widespread acceptance. Indeed, *anti-anti-psychologism* is now perfectly conventional. Philosophy, after Quine, was in the peculiar position of having surrounded itself with a moat—only to have drained it of water.

V

Now, in bringing the empirical entanglements of canonical philosophers back into view I am doing just what Renan suggested we do with national identities. I am crafting a genealogy that supports a conception of the subject to which I am sympathetic. In deciding what story to tell of philosophy’s past, those who were convinced of the importance of the distance between, say, philosophy and psychology, picked their way through the past accordingly. The recent return to these shores of the epithet “experimental philosophy” is—as one tendency in our profession might put it—a return of the repressed.

There are all kinds of ways in which experimentation has been brought to bear in our discipline. For decades, of course, philosophers of mind have been working closely with their peers in psychology and psycholinguistics and computer science; there has been an effort to ground the philosophy of language, too, in more naturalistic theories of the mind (an effort to which my first two books belong). Philosophers who work on consciousness can tell you all about Capgras Syndrome and research in various forms of neurologically induced agnosia.

The relevance of empirical research tends to be more hotly contested in the obviously normative reaches of moral theory. But here, too, the “renewed incursions” have been hard to miss. Over the past decade, for instance, there’s been a debate between virtue ethicists and critics armed with findings from social psychology—in particular,

¹⁶ W. V. O. Quine, “Epistemology Naturalized” in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), 82–83.

¹⁷ Michael Dummett, *Frege and Other Philosophers* (Oxford: Oxford University Press, 1991), 287.

empirical evidence for what's called "situationism." These critics draw on decades of research suggesting that much of what people do is best explained not by traits of character but by systematic human tendencies to respond to features of their situations that nobody previously thought to be crucial at all.¹⁸

Situationists think that someone who is, say, reliably honest in one kind of situation will often be reliably dishonest in another. Back in 1972, experimental psychologists had found that, if you dropped your papers outside a phone booth in a shopping mall, you were far more likely to be helped by someone who had just had the good fortune of finding a dime waiting for them in the return slot. A year later, John Darley and Daniel Batson discovered—this is probably the most famous of these experiments—that Princeton seminary students, even those who had just been reflecting on the Gospel account of the Good Samaritan, were much less likely to stop to help someone "slumped in a doorway, apparently in some sort of distress," if they'd been told that they were late for an appointment. More recently, experiments showed that you were more likely to get change for a dollar outside a fragrant bakery shop than standing near a "neutral-smelling dry-goods store."¹⁹

Many of these effects are extremely powerful: huge differences in behavior flow from differences in circumstances that seem of little or no normative consequence. Putting the dime in the slot in that shopping mall raised the proportion of those who helped pick up the papers from 1 out of 25 to 6 out of 7; i.e. from almost no one to almost everyone. Seminarians in a hurry are six times less likely to stop like a Good Samaritan.²⁰ Knowing what I've just told you, you should surely be a little less confident that "she's helpful" is a good explanation next time

¹⁸ I rely on the excellent account in John Doris *Lack of Character* (Cambridge: Cambridge University Press, 2002). For situationism in psychology see: Lee Ross and Richard E. Nisbett, *The Person and the Situation* (Philadelphia: Temple University Press, 1991).

¹⁹ A. M. Isen and P. F. Levin (1972). "The effect of feeling good on helping: Cookies and kindness," *Journal of Personality and Social Psychology*, 21 (1972), 384–388; J. M. Darley and C. D. Batson, "From Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology* 27 (1973), 100–108; K. E. Matthews and L. K. Cannon, "Environmental noise level as a determinant of helping behavior," *Journal of Personality and Social Psychology* 32 (1975), 571–577; R. A. Baron and J. Thomley, "A whiff of reality: Positive affect as a potential mediator of the effects of pleasant fragrances on task performance and helping," *Environment and Behavior*, 26 (1994), 766–784. All are cited in Doris, *op. cit.*, 30–34, 181.

²⁰ And people are about one tenth as likely to help someone behind a curtain who has had what sounds like an accident, if there's someone else standing by who does nothing. B. Latane and J. Rodin, "A lady in distress: Inhibiting effects of friends and strangers on bystander intervention," *Journal of Experimental Social Psychology*, 5 (1969), 189–202. As cited in Doris.

someone stops to assist you in picking up your papers, especially if you're outside a bakery!

Philosophers inspired by situationists have argued that this reality is at odds with the conception of human character that underlies virtue ethics. For when virtue ethicists ask us to be virtuous, they typically mean that we should have, or cultivate, persistent, multitrack dispositions to, say, act compassionately, or honestly. Their situationist critics object that we're simply not built that way—that character, as the virtue ethicists conceive it, is about as real as phlogiston. Crudely put: If there's no such thing as character, then the project of a character ethics—a morality centered on virtues—is a waste of time.

VI

But can mere facts about how we *are* disqualify an account of how we *ought* to be? Perhaps nobody is fully virtuous; still, virtue ethics is hardly alone in assigning a role to elusive ideals. Our models of rationality are also shot through with such norms. They tell us not how we do reason but how we ought to reason; and you don't need to be a Kahneman or a Tversky to know that we don't do it how we ought to.²¹ If you have been following debates about the role of ideals in cognitive psychology, you might think that the answer is to treat claims about virtues as moral heuristics. One eloquent advocate of modern virtue ethics, Rosalind Hursthouse, encourages this approach—even though it is decisively not her own—when she insists that a virtue ethics is just as helpful as, say, utilitarianism, in offering guidance as to what we should do in particular cases: quite simply, we should do what a virtuous person would do.²²

But there are many difficulties, I think, for a heuristics of virtue.²³ Here is one: virtues are not merely instrumental. Virtue ethics wants us to aim at *becoming* a good person, not just at maximizing the chance that we will do what a good person would do. (As the 19th century logician, Archbishop Richard Whately, once observed, honesty may be the best policy, but this is not a maxim that guides an honest person.)

By contrast, cognitive heuristics are, so to speak, twice dipped in means–end rationality. First, the right outcome is defined by what someone possessed of infinite cognitive resources would do. Second, means–end rationality is used to determine how people with limited cognitive resources can maximize their chances of doing what's right

²¹ D. Kahneman and A. Tversky. "On the Psychology of Prediction" *Psychological Review*, 80, (1973), 237–251.

²² Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999); see also her "Virtue Ethics," *The Stanford Encyclopedia of Philosophy* (Fall 2003 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/fall2003/entries/ethics-virtue/>.

²³ For more of them, see my *Experiments in Ethics* (Cambridge: Harvard University Press, 2008), Chapter 4.

according to the first test. When we try to concoct a heuristic of virtue, we must start, analogously, by defining the right outcome as what someone ideally virtuous would do. Since we're not ideally virtuous, the heuristics model should next introduce means–end rationality to maximize your chance of doing what's right by the first test.

The trouble is, of course, that virtue ethics requires that we aim at the good for reasons that aren't reducible to means–end rationality. With the cognitive heuristic, what matters is the outcome: but if virtue ethics tells you that outcomes aren't the only thing that matters, then you cannot assess heuristics by means–end rationality—by looking at the probability that they will produce certain outcomes.

That doesn't leave virtue ethics without argumentative recourse—far from it. But the confrontation with social psychology has forced virtue ethicists to clarify the contours of their account—to make claims and concessions about which psychological claims are and are not necessary for their view; and, quite significantly, many of them have made arguments about what the psychological evidence actually shows about human nature. From our metaphilosophical perspective, what matters is not so much whether the situationists' claims are right as whether they are relevant. In that sense, there may be victory even in defeat.

VII

In recent years, however, philosophers have done more than draw upon research by experimentalists in other disciplines. The recent currency of the phrase “experimental philosophy” often refers to research that, in the mold of Arne Naess, has actually been conducted by philosophers ... often, as with Naess, on nonphilosophers. Much of this work is in a continuation of the project of conceptual analysis.²⁴ If conceptual analysis is the analysis of “our” concepts, then shouldn't one see how “we”—or representative samples of us—actually mobilize concepts in our talk? So one strain of this work seeks to elicit and tabulate intuitions people have about various scenarios.

The use of such scenarios, or thought experiments, is a hallmark of philosophy. Yet the newer philosophical experimentalists seem to have noticed that many thought experiments in philosophy were, so to speak, at a double remove from reality. Not only were the scenarios unrealized, the claims about how we would respond to those scenarios were also simply asserted, rather than demonstrated.

²⁴ One impetus to this experimental turn is the fact that, since the 1970s, it has been common among philosophers to refer to our concepts of intention, belief, desire, and so on under the rubric “folk psychology.” And, once we did so, it raised the possibility of error—or, anyway, the possibility of the possibility of error, since many denied, on conceptual grounds, that there was such a possibility.

Recall Hume's Missing Shade of Blue argument. If a man had never encountered a particular shade of blue, and is now presented with a sequence of deepening shades, absent that one, will he notice the gap and be able to imagine the unseen shade? Hume, the great empiricist, writes, "I believe there are few but will be of opinion that he can." That has been the usual protocol. We conjure a scenario, and then announce that, in such case, "it would be natural to say" X, Y, or Z. (In the empirical spirit, I should report that, when I typed the phrase "it would be natural to say" into Google's Book Search, it happily returned, as its top search results, passages by Gilbert Ryle, Peter Strawson, Max Black, and Bertrand Russell.)

Most thought experiments are unrealized for good reasons. With the stroke of a pen, Frank Jackson can summon up in imagination Mary, the scientist raised in a world without color. *Actually* raising such a scientist, however, would be arduous, time consuming, and costly ... and likely to get bogged down in human subjects review committees. Any attempt to reproduce Judith Jarvis Thompson's thought experiment about the comatose violinist would run into protests from the musician's union. Yet the other part—finding out what people would think when contemplating such scenarios—can be expeditiously and inexpensively done. So why not remove at least some of the thought from our thought experiments?

VIII

This approach is well exemplified by the work of Joshua Knobe, who, in his best-known study, asked subjects to consider two scenarios. In the first, the chairman of a company is asked to approve a new program that will increase profits and also help the environment. "I don't care at all about helping the environment," the chairman replies. "I just want to make as much profit as I can. Let's start the new program." So the program is launched and the environment is helped. The second story is identical—except that the program will *hurt* the environment. Once again, the chairman is indifferent to the environment, and the program is launched in order to increase profits, with the expected results.

Rather than surmising what "it would be natural to say," Knobe totted up the responses of actual subjects, and found that when the program helped the environment, only 23% percent agreed that the chairman had "helped the environment intentionally." When the program harmed the environment, though, 82% agreed that the chairman had "harmed the environment intentionally." And a similar pattern recurred when various other scenarios were tested.

As Cavendish told us, most experiments have their rise from the speculative, and Knobe's research is part of a larger exploration of how judgments about responsibility, intentional action, and causation can be affected by moral or, anyway, evaluative considerations. In his original

view, then, our intuitions about intention aren't incoherent; rather, they track with our ascriptions of praise and blame. Similar judgments shape our intuitions about causation, he finds. If a mishap is the combined result of more than one factor, one of which is a misdeed, most subjects say that the misdeed, not another necessary factor, was the "cause." In fact, that very dynamic is visible in the way the question about the chairman was framed: it supposed that the chairman "harmed" the environment, as opposed to, say, allowed the environment to be harmed. (We don't automatically suppose that people have performed, or even caused, the foreseeable consequences of their actions.)

In this line of studies, a seeming anomaly in our folk concept of intentional action is identified and defended. In other studies, Knobe and collaborators are not so generous toward the anomalies they find. One has to do with intuitions about determinism and moral responsibility, and aims to shed light on the dispute between compatibilists and incompatibilists. In that study, subjects were asked to imagine a fully deterministic universe, in which everything that happens is completely caused by whatever happened before. They're told about Mark, who arranges, in this universe, to cheat on his taxes, as he has done many times before. Is Mark fully morally responsible for cheating on his taxes? Most people said no. But the responses changed when the scenarios involved were, as they put it, "high affect" cases: for example, when people were told about Bill, who, as he often has done in the past, stalks and rapes a stranger. In that case, most subjects thought that he *was* fully morally responsible for his misconduct. Now, the investigators here—Josh Knobe and Shaun Nichols—are persuaded by psychological research that patients who are incapable of affective responses don't act like people who engage in cool moral assessment; rather, they act like people who don't see the point of moral assessment. Our philosophers hold that affect is part of our competence in making judgments about responsibility. But they also believe that it can lead to performance errors. When strong emotion converts us to compatibilism, we have been led into error.

IX

Are the compatibilists really driven by emotion? You might use a brain scan to make sure. Consider Josh Greene's work in moral psychology, where he and his colleagues have studied the fMRI images of people thinking through those celebrated trolley-car experiments. Why do people think it's OK to reroute a runaway trolley car from a track where five pedestrians will be killed to a track where just one will be killed—but not OK to stop the trolley by pushing a large man hovering innocently nearby onto the track? Greene thinks it's related to the fact that when subjects contemplate pushing the large man, the parts of the brain that "light up"—the medial frontal gyrus, the posterior cingulate gyrus, and

the angular gyrus—are regions associated with emotion.²⁵ If, like Greene, your sympathies are broadly consequentialist, and, so to speak, revisionary, you'd conclude that our reluctance to support tipping the large man arises from morally irrelevant considerations—how “up close and personal” the action is, as Greene puts it.²⁶ Emotions have overridden cool judgment, and here are the pictures to prove it.

Especially given the glamour of neurological imaging, though, we might want to bear in mind that slides and surveys are not arguments. For what inferences would you make if you started with the premise that pushing the large man onto the tracks below was wrong? Then those brain scans, assuming they show what they're supposed to show, would be marvelous evidence for the moral authority of affect: our emotions are more exquisitely attuned to the moral features of a situation than are our crude powers of calculation, you would say. In just this way, it's open to compatibilists to laud the superior analysis elicited by the more emotionally fraught case—the rapist—and to find performance error in people's permissive views about the tax cheat. Neither questionnaires nor brains scan are likely to settle debates between deontologists and consequentialists, or compatibilists and incompatibilists. And, of course, they are not offered to that end.

For here, as with the debate with the situationists, the confrontation with evidence can help theorists clarify what claims they are and are not committed to. Indeed, the argument over the usefulness of experimental philosophy itself can be illuminating—that is to say, useful. In the old days of conceptual analysis, it could sometimes seem as if the philosopher was a kind of fancy lexicographer, sorting out the definitions of various grand words. But that paradigm suggests lurking trouble, for within lexicography itself there has always been a tug between normative and descriptive—or, as you might say, empirical—impulses.

Suppose—if you'll tolerate a thought experiment about thought experiments—the A.P.A. decided to hire Zogby International to poll-test those thought experiments we love so well: scenarios about the Experience Machine; Twin Earth; Mary the Color Scientist; Thompson's comatose violinist; Williams's body-swaps. What could the survey results tell us that we don't already know?

Well, we might learn that, in some instances, what philosophers supposed “would be natural to say” wasn't what non-philosophers found it natural to say. Some of our intuitions might be less popular than we

²⁵ Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, Jonathan D. Cohen, “An fMRI Investigation of Emotional Engagement in Moral Judgment,” *Science*, Vol. 293, no. 5537, (14 September 2001), 2105–2108.

²⁶ Greene maintains that people should “develop a healthy distrust of moral common sense,” and that “our social instincts were not designed for the modern world.” <http://www.wjh.harvard.edu/~jgreene/> Accessed December 25 2007.

assumed. Some of our intuitions might turn out to be culturally specific. When Edouard Machery and colleagues posed a famous thought experiment of Kripke's to students, they found that those from Hong Kong had quite a different pattern of response than those from New Jersey. But my guess is that in most cases, the results would shore up the basic intuition it was meant to pump; and that, where it did not, philosophers, too, have already been left divided. What we're not going to end up with is some sort of metaphysics by plebiscite; we wouldn't want to. For most of us don't believe the truth is simply what most of us believe.

X

In a friendly review in the *Journal of Philosophy* of Arne Naess's monograph with which I began, Ernest Nagel remarked:

Since most philosophers will not be prepared to undertake the sort of "dirty work" to which Dr. Naess invites them, he will no doubt remain an outcast from the philosophic community and will have to find what solace he can in being a "mere" scientist.²⁷

Well, it took half a century, Professor Nagel: but I think your hypothesis has been disconfirmed. Many of today's experimental philosophers, as we've seen, like Naess, think it is important to elicit the intuitions of non-philosophers: in general, we want our theories to be about intention, not some guild-specific variant, shmintention, as it were; we want to be describing Truth, not something else we guild members have come to honor with that name.

It's common to analogize folk psychology with folk physics. But, of course, professional physicists can happily leave folk physics far behind as they tinker with their Calabi-Yau Manifolds and Gromov-Witten invariants. By contrast, moral psychology, however reflective, can't be dissociated from our moral sentiments, because it's basic to how we make sense of one another and ourselves. In a deliberately awkward formulation of Bernard Williams's, moral thought and experience "must primarily involve grasping the world in such a way that one can, as a particular human being, live in it."²⁸

That's where philosophy and natural philosophy part company. Chemists don't read Humphrey Davy, and physicists don't read Aristotle. So why do philosophers still read, quarrel with, or defend, figures from centuries ago? A couple of thousand years later, we're still kicking around

²⁷ E.N. Rev: "Truth" as Conceived by Those Who Are Not Professional Philosophers by Arne Ness, *The Journal of Philosophy*, Vol. 36, No. 3. (Feb. 2, 1939): 79.

²⁸ Bernard Williams, *Moral Luck* (Cambridge: Cambridge University Press, 1981), 52.

versions of a question from *Euthyphro*: Is an action good because the gods love it or do the gods love it because it's good? You could say—it has been said—that science is the “successful” part of philosophy, “philosophy” the remains. But there's a reason that we've having the same debates: it's that these are stories about *us*. The other methods of inquiry, when they came of age, left the family and lighted out on their own. Philosophy, bound to make sense of a distinctively human realm of meaning, can't sever its ties to ordinary human intelligibility—to the language we use everyday to make sense of ourselves and others.

XI

The new experimental philosophy, I hope I've made clear, poses no threat to philosophical analysis. It offers stimulus, challenge, interest, and not just new sources of funding. Indeed, if anything shadows the prospects of experimental philosophy, it is, I think, that our notion of experiment, of the empirical, may be too conservative, too narrow. In the social sciences—not least in economics—we have lately heard a great deal about “natural experiments,” the offerings of history. (Two of the last three John Bates Clarke medalists in economics have specialized in natural experiments.) If our former colleagues in the other moral sciences find nourishment in natural experiments, perhaps we should not be so neglectful of this resource. I want to suggest, in closing, that one of the many things that philosophers might usefully do, *as* philosophers, is to attend to such natural experiments—to examine the moral history of our species.

Consider how the practice of dueling came to be regarded as an exercise not in honor but in ignominy; how foot-binding, after a millennium, came to be discarded as a barbarism; how slavery went from being considered part of the natural order to being an unpardonable offense. Would moral philosophers go wrong to take an interest in how these arguments were actually waged and won? There is, after all, an ancient adage that history is philosophy taken from examples. It is usually attributed to the Augustan historian of the Roman republic, Dionysius of Halicarnassus—after Herodotus, the most famous historian to have been born in that Ionian city. So it was a long-ago historian who invited us philosophers to the party.

Surely we have good reason to accept the invitation: we can absorb a certain amount of historical richness without relinquishing the rigor we rightly prize. Some of the most exciting work in the philosophy of science, after all, has arisen from theorists who have a detailed interest in the actual work of actual scientists. Moral theorists are inclined to suppose, reasonably enough, that issues involving specific judgments about specific customs involve too great a level of specificity, and that our distinctive contribution, *as* philosophers, is to attend to more fundamental issues.

So it's worth remembering, yet again, that the arguments of those we like to consider our disciplinary ancestors have often depended on stories about the actual doings of actual people. It's worth recalling Hume's famous footnote in his *Enquiry Concerning Human Understanding* in which he denounced the tendency of writers "on *moral, political, or physical* subjects, to distinguish between *reason* and *experience*, and to suppose, that these species of argumentation are entirely different from each other."²⁹ Hume's *History Of England*—five volumes of empirical information, elegantly organized—has rightly been seen as expressing ideas about morality and politics and psychology. For him, it was an extension of the project of a work like the *Enquiry Concerning Human Understanding*. For him, history was too important to relegate entirely to the historians. Ernest Nagel talked, wryly, in that review of Naess about "dirty work"; bringing philosophy to the realm of the natural experiment is still dirtier. But, if I may paraphrase Chesterton and generalize outrageously, this approach has not recently been tried and found wanting; it has been found difficult and left untried.

So I am a pluralist on many levels: the wild eclecticism of our profession delights me. Insofar as method is concerned I have only this modest pluralist suggestion: that we would do well to sustain a variety of traditions of reflection on questions that matter to us. Unless you already know all the answers, I say, you don't even know for sure which questions are worth asking.

Is it worth asking, for instance, how shifts in our moral judgments, changes in our basic values, take place? What would happen if a few more of our contemporary ethicists turned their subtle anatomizing intelligence—as Hume or Montesquieu once did—to the thick, untidy realm of our moral history? John Stuart Mill famously talked about "experiments in living," and our species has engaged in a great many of such experiments, sometimes disastrously. We, like our ancestors, are all subjects in a vast ongoing natural experiment, p.'s without an L; we feel our way through life in a world, a planet, whose fragility Arne Naess was among the first to recognize. There are arguments to be made, truths to be defended or discovered. Our own experiments in living—the changes in moral perceptions we ourselves have lived through—provide a trove of information that will fit on no questionnaire. Experimental philosophy of this sort might indeed be dirty work. But doesn't someone have to do it?

²⁹ David Hume *An Enquiry Concerning Human Understanding*, A. Selby-Bigge 2nd ed. (Oxford: Clarendon Press, 1902), 43.