

## 12 Toward an Empirically Responsible Ethics: Cognitive Science, Virtue Ethics, and Effortless Attention in Early Chinese Thought

Edward Slingerland

The two models of ethics that have been dominant in the West since the Enlightenment—and continue to serve as the default in both academic philosophy and public ethical debate—are deontology and utilitarianism. The former is best captured in Kant’s classic rule-based approach, where ethical behavior is to be guided by a hierarchy of maxims, which can be applied to properly classified situations in a fairly straightforward manner. For instance, consider the maxim, “It is wrong to lie.” When presented with a situation, we can consult our definition of a *lie* to determine whether act X in this given situation was or was not an instance of lying, and having made this determination, we can then decide whether it was right or wrong depending on where this particular maxim is located in a hierarchy of maxims—for example, perhaps it is trumped by the maxim that we should strive to preserve life. If we are utilitarians, in any situation we should be able to unproblematically tally up the costs and benefits of proposed courses of action, do the math, and thereby figure out which course of action maximizes whatever good our brand of utilitarianism deems important (happiness, justice, gross national product, etc.).

Despite their differences, both models of ethics might be characterized as *cognitive control* or *high reason* (Damasio 1994) models: In order to be effectively implemented, they require the agent to be consciously aware of all of the relevant factors, to suppress emotional reactions and social biases, and to arrive at and carry out an objective, dispassionately rational decision. The reasoning process drawn upon is amodal, involving the manipulation of abstract maxims or mathematical values. In both cases, the entire process of moral reasoning is transparent and under our conscious control and has nothing to do with the details of our embodiment, or with emotions, implicit skills, or unconscious habits. In other words, effortless attention and action—automatic, nonconscious, embodied engagement with the world—play no role in either model of ethics.

### Problems with Cognitive-Control Ethical Models

The extent to which disembodied, purely rational convictions could realistically be expected to guide ethical behavior has been questioned at least since the early Chinese

philosopher Mencius (4th c. b.c.e.) and his criticism of the Mohists, who defended a form of rational utilitarianism.<sup>1</sup> More to the point of this volume, recent developments in the cognitive sciences have begun to call into question the model of the self upon which both deontology and utilitarianism are based. This model, which we might term *objectivist*,<sup>2</sup> involves conceiving of the self as a disembodied, unitary consciousness, housed within a body but distinct from it and all of the phenomena—emotions, habits, somatic skills—that come along with a body. This disembodied intelligence finds itself confronting a world with a fixed, pregiven structure. This world contains discrete objects, which make themselves known to the disembodied self through (somehow) translated sensory impressions; on the basis of these sensory impressions, the objects can be classified into clearly demarcated categories, with sufficient and necessary conditions for category membership. Categories are then labeled with arbitrary symbols (words), and by our combining these words into sentences, the logical relationships between categories of things in the world can be directly represented and communicated. Reasoning, on this model, consists of “a rule-governed manipulation of connections among symbols ... [whereby] connections among symbols and rule-governed combinations of symbols are established and traced out according to various logical canons or principles” (Johnson 1987, xxiv).

Of course, there exist in the Western tradition older, more pragmatic models of rationality that can be traced back to Aristotle (trans. 1999) and that involve a mind always already in direct, constant contact with a messy world of tangible things. The strongly dualistic model of a disembodied mind really takes rigid hold on Western thought in the Enlightenment. Since at least the 18th century, however, it has become entrenched as the sole proper model of ontology and epistemology in most areas of academic philosophy, including ethics. The purpose of this chapter is, first of all, to review how the picture of human reasoning and decision making that is emerging from the cognitive sciences calls into question some of the basic assumptions of objectivism–rationalism and, therefore, models of ethics based upon “high reason.” I will then argue that the importance in everyday human cognition of effortless attention and action suggests that virtue ethics, a model of ethics characteristic of many world traditions (including pre-Enlightenment Europe), might be preferable to deontology and utilitarianism both descriptively and normatively. Finally, I will illustrate how effortless attention and action might be cultivated and manifested in ethical behavior by looking at a text from early China, the *Analects* of Confucius, which provides one of the earliest accounts of a virtue ethic in world literature.

### Problems with the Objectivist Self and Effortful Attention

The consensus coming out of recent work in cognitive science suggests that pure, bloodless, fully conscious rationality plays a limited role in everyday decision making,

and indeed an absence of emotion—a hallmark of the ideal moral agent for Plato or Kant—apparently transforms us into ethical incompetents. We are rarely fully conscious or in control of what “we” are doing, and indeed the very idea of a unitary, conscious “I” in control of the dumb, animal-like nonself (the body, the emotions) appears to be an illusion. Even such quotidian achievements as ordinary language comprehension and basic perception of our surroundings rely heavily upon tacit know-how and fast and frugal heuristics, guided by embodied and mostly unconscious emotional reactions to our environment. Perception is not concerned primarily with representation but rather with action, and the concepts we acquire from interacting with the world seem to be based primarily upon imagery and sensorimotor schemas. Concepts are therefore not amodal, abstract, and propositional, but perception and body based. Even when dealing with “abstract” concepts or complicated, novel situations, somatic knowledge appears to play a fundamental role.<sup>3</sup> Below I will touch upon each of these themes, noting their relevance for ethics.

### The Importance of Tacit Know-How

Tacit *know-how* has been an increasingly important theme in philosophy, especially among philosophers with some knowledge of cognitive science. John Searle, for instance, asks his readers to consider the sentences “Sally cut the cake,” “Bill cut the grass,” or “The tailor cut the cloth.” None of these sentences are characterized by lexical ambiguity or obvious metaphorical usage,

but in each case the same verb will determine different truth conditions or conditions of satisfaction generally, because what counts as cutting ... will vary with the context.... If somebody tells me to cut the cake and I run it over with a lawn mower or they tell me to cut the grass and I rush and stab it with the knife, there is a very ordinary sense in which I did not do as I was told to do. Yet nothing in the literal meaning of those sentences blocks those wrong interpretations. (Searle 1995, 130–131)

What *does* block the wrong interpretation of these sentences is our recourse to what Searle calls *the Background*: a reservoir of tacit social and ontological assumptions and skills for coping with the world (Searle 1995, 129–137). Involving a type of inarticulable know-how, this Background cannot, according to Searle, be translated into a finite set of explicit sentences, which means that the comprehension of human sentences cannot be reduced to simply the algorithmic transformation of strings of symbols.

A similar point is made by Hilary Putnam, who notes that the ease with which we access our background knowledge disguises the potential ambiguity of most of the sentences that we nonetheless effortlessly and correctly process; it is clear, upon reflection, that we need recourse to “good judgment” in order to figure out what almost any given string of words means in any given context. Putnam points out that this need for contextual judgment undermines the algorithmic model of sentence

processing, for “as Kant long ago said (if not in those terms), there isn’t a recursive rule for ‘good judgment’” (Putnam 1999, 89). He summarizes a point made by Stanley Cavell (1979), in *The Claim of Reason*, that “our ‘attunement’ to another, our shared sense of what is and what is not a natural projection of our previous uses of a word into a new context, is pervasive and fundamental to the very possibility of language—without being something that can be captured by a system of ‘rules’” (89). It is thus apparent that the understanding of even quite pedestrian human utterances involves reliance on a huge reservoir of tacit, nonalgorithmic knowledge and a pragmatic “feel” for the conversational environment. The objectivist paradigm does not seem to capture either how knowledge of the world is stored nor how it is processed online, and this fact is not surprising considering what knowledge is supposed to *do* for creatures such as ourselves—that is, help us to survive long enough to pass on our genes most effectively.

Philosophers such as Gilbert Ryle and Michael Polanyi have developed theoretical accounts of the function of know-how and the distinction between explicit and tacit knowledge, and the importance of implicit, bodily skills for human flourishing can be traced in the West as far back as Aristotle (trans. 1999).<sup>4</sup> Here I would like to focus on a growing body of empirical work coming out of social psychology and behavioral neuroscience that bolster these theoretical accounts by highlighting the crucial role that tacit, nonpropositional forms of knowledge play in everyday human cognition. In their review of the social psychology literature on “automaticity,” for instance, John Bargh and Tanya Chartrand discuss studies revealing the power of priming to affect modes of behavior, the effect of stereotype priming on social judgments, the unconscious acquisition of goals from external stimulation, and the unconscious mimicry of behavior and its effect on social judgments. For instance, subjects whose movements—crossing their legs, playing with their hair—were subtly mimicked by an interviewer subsequently rated the interviewer as more likeable, and the interview process itself as having gone more smoothly, than if the interviewer maintained a relaxed, neutral physical posture. Bargh and Chartrand conclude that, in many areas, people “classify their experience as either good or bad and do so immediately, unintentionally, and without awareness that they are doing it” (Bargh and Chartrand 1999, 474). They describe this process as follows:

Automatic evaluation of the environment is a pervasive and continuous activity that individuals do not intend to engage in and of which they are largely unaware. It appears to have real and functional consequences, creating behavioral readiness within fractions of a second to approach positive and avoid negative objects, and, through its effect on mood, serving a signaling system for the overall safety versus danger of one’s current environment. All of these effects tend to keep us in touch with the realities of our world in a way that bypasses the limitations of conscious self-regulation capabilities. (475–476)

In other words, the social psychology literature documents the pervasive importance of automatic, tacit, and unformulizable heuristics (“good judgment”) on human behavior and attitude formation.

It is also apparent that there are separate human cognitive systems that work on the implicit and explicit levels, with know-how functioning primarily at the former level. Robert Zajonc and colleagues have demonstrated that people can have affective responses to stimuli without being able to consciously recognize them,<sup>5</sup> and Antonio Damasio has shown that skin conductance reactions to emotionally charged stimuli *precede* conscious awareness of emotion: Emotional states happen first, and conscious feelings follow (Damasio 2003, 101). Joseph LeDoux has postulated the existence of two systems of memory, an unconscious, implicit “emotional memory” and an explicit “declarative” memory, and reviews studies indicating that priming, manual skills, and cognitive skills (such as the ability to solve a particular type of puzzle) are preserved in amnesiac patients, suggesting that implicit “know-how” is developed and stored in brain systems separate from those that subserve conscious memory (LeDoux 1996, 195–198).

Of course, it is obvious that the brain systems associated with abstract reasoning and cognitive control can, at least sometimes, bring implicit biases and other sorts of emotions into consciousness in order to modify or override them. Indeed, there is evidence that cortical control is necessary for the normal conscious experience and expression of emotion. Animals that have had their cortex removed, for instance, are still capable of having emotional reactions, but they are not entirely normal—such creatures are easily provoked and seem entirely incapable of regulating their emotional reactions, which suggests that cortical areas normally rein in and control emotional reactions (LeDoux 1996, 80). However, it is equally clear that conscious self-control is something of a limited resource. The work of Roy Baumeister and his colleagues (Baumeister, Bratslavsky, Muraven, and Tice 1998; Muraven, Tice, and Baumeister 1998) has shown that when conscious control is exerted in one domain, this depletes the individual’s ability to exert it in another unrelated domain. This suggests that conscious self-control must be a relatively rare occurrence, since it seems to require a lion’s share of cognitive resources. There is also considerable evidence that conscious intervention in automatic processes can be counterproductive. Baumeister’s work has shown that automatic behaviors are disrupted when people analyze and decompose them (Baumeister 1984). Similarly, Timothy Wilson and Jonathan Schooler have shown in a series of studies that, in many domains, people form automatic and apparently quite adaptive evaluations that can then be disrupted when these people are asked to reflect on their reasons for their evaluative feelings. Untrained subjects who were asked to spontaneously rate the taste of a variety of jams, for instance, assigned ratings that best matched their demonstrated future satisfaction, as well as the ratings

of food industry expert tasters; when asked to rationalize their rankings by analyzing their reasons as they went along, however, the optimality of their ratings decreased significantly.<sup>6</sup> In summary, evolution seems to have off-loaded the vast bulk of our everyday decision making and judgment formation onto automatic, unconscious systems, because such systems are fast, computationally frugal, and reliable.

### **No Unitary Subject: The Objectivist Knower Is Not Master of His Own House**

The objectivist model of reasoning and conscious decision making assumes the presence of a unitary, conscious self—the locus of rationality and will—whose job is to evaluate incoming sense data, classify it, and enforce appropriate conclusions and behavioral decisions on the dumb, recalcitrant emotions or body. While it is acknowledged that this rational self is not always successful in exerting control over other portions of the self, it is assumed that the self is at least *aware* of what “it” is doing and why.

The phenomenon of automaticity discussed above calls this assumption into question, and the outline of the human neural architecture emerging from neuroscientific research indeed calls into question the very idea of a unitary ego as the locus of consciousness. One of the main Cartesian “errors” at which Antonio Damasio takes aim in his famous 1994 book, *Descartes’ Error*, is the concept of a Cartesian theater: a central area of consciousness that experiences the world and the self in a unified fashion and serves as a kind of headquarters of knowledge and decision making. As Damasio notes, there is no single region in the human brain equipped to act as such a central theater; although there are various intermediate-level “convergence zones” that coordinate information coming in from more specialized sensorimotor regions, there is no “master” convergence zone that has an overall view of the entire process (Damasio 1994, 94–96; cf. Damasio 1989). Of course, in our everyday experience we certainly *feel* a strong sense of mental integration—the intuition of a unified self in charge of and informed about everything is very powerful and universal. This is, however, “a trick of timing,” Damasio argues, an illusion “created from the concerted action of large-scale systems by synchronizing sets of neural activity in separate brain regions” (1994, 95; cf. Damasio and Damasio 1994). How this sort of “binding” occurs is still not precisely understood, but what is clear is that there is no little homunculus collecting data and running a central command post in the brain. It is also likely that each of the various interconnected subsystems that together make up the mind encode and process information in their own task-specific manner, merely transmitting the results of their processing to other appropriate subsystems, which means that there is probably not even the kind of central, universal representational format that such a homunculus would need to function (Clark 1997, 136–141).

One of the more dramatic illustrations of the decentered nature of the self emerges from a series of experiments with split-brain patients performed by Michael Gazzaniga

and his colleagues. In these patients the corpus callosum, which normally connects the left and right hemispheres, has been severed (this has been found to be an effective, if last resort, treatment for certain severe forms of epilepsy). The left brain is the seat of verbal ability and interpretative synthesis—in other words, the locus of our sense of unified self—and Gazzaniga and his colleagues found that the illusion of an in-control, unified self that the left hemisphere weaves persists even when it is most certainly *not* in control. For instance, in one experiment, images were selectively presented to each hemisphere: The left hemisphere was shown a chicken claw, the right a snow scene. Subjects were then presented with an array of objects and asked to choose an object “associated” with the image they were shown. A representative response was that of a patient who chose a snow shovel with his left hand (controlled by the right hemisphere and prompted by the snow scene) and a chicken with the right (controlled by the left hemisphere and prompted by the chicken claw). Asked why he chose these items, “he” (that is, his left hemisphere “spin doctor”) replied, “Oh, that’s simple. The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed” (Gazzaniga 1998, 25). Gazzaniga and LeDoux found a similar effect with normative judgments: In one particular patient referred to as P.S., the left hemisphere could correctly identify the emotional valence of a stimulus presented to right hemisphere (“good” or “bad”) without any conscious awareness of the nature of the stimulus (reported in LeDoux 1996, 14–15). In other words, the left hemisphere “was making emotional judgments without knowing what was being judged” (LeDoux 1996, 15).

As Gazzaniga concludes, “The left brain weaves its story in order to convince itself and you that it is in full control” (1998, 25). He argues that, in place of the all-powerful legislator or canny calculator, a more appropriate metaphor for the conscious, verbal self might be a “harried playground monitor, a hapless entity charged with the responsibility of keeping track of multitudinous brain impulses running in all directions at once” (1998, 23), and also responsible for concocting an *ex post facto* story of unified control for the consumption of both itself and others. One is reminded of Nietzsche’s claim that the idea of free will is “the expression for the complex state of delight of the person exercising volition, who commands and at the same time identifies himself with the executor of the order,” taking pleasure in the illusion that “*L’effet c’est moi*” (Nietzsche, 1886/1966, 26).

Lest one think this sort of illusion of self-control is confined to people with extreme trauma, such as a severed corpus callosum, a large body of psychological experimental evidence has demonstrated the existence of a rather deluded “spin doctor” in neurologically normal individuals. The pioneering experiments in this field were performed by Richard Nisbett and Timothy Wilson (Nisbett and Wilson 1977; T. Wilson and Nisbett 1978), who demonstrated in a series of experiments that people often report having thoughts and desires that they could not, in fact, possibly have, and that the

verbal reports given by subjects concerning the effects of stimuli on their judgments and behavior in experiments are often highly inaccurate. In a now classic experiment, Nisbett and Wilson presented shoppers at a mall with a display of identical nylon stockings, laid out from left to right. They observed the well-attested phenomenon that, given such a horizontally oriented presentation of otherwise identical items, people display a preference for the items on their right-hand side: in this experiment, the rightmost stocking was preferred almost four to one over the leftmost. What they found most interesting, however, was the confabulated rationales the subjects concocted to justify their choices—swearing, for instance, that their preferred stocking was clearly of better quality than the identical stocking to its left. None of the subjects mentioned the position of the article, and virtually all of the subjects absolutely denied the possible effect of the article's position on their judgment when directly questioned about it by the researchers, “usually with a worried glance at the interviewer suggesting that they had misunderstood the question or were dealing with a madman” (1977, 244).

Studies of subjects given posthypnotic suggestions show a similar effect. For instance, Philip Zimbardo (Zimbardo, Laberge, and Butler 1993) found that subjects in which both hypnotic arousal and amnesia were induced generated a range of plausible explanations for their mental state that had nothing to do with the actual context of the experiment, and Paul Rozin and Carol Nemeroff (Rozin and Nemeroff 1990) found that subjects justified disgust-based attitudes with rationalizations that proved upon examination to be poor predictors of their actual behavior. Jonathan Haidt and his colleagues have found a similar effect with regard to moral judgments: Judgments resulting from emotional reactions or posthypnotic suggestions are invariably given *ex post facto*—and utterly specious—rational justifications by experimental subjects.<sup>7</sup> Together with the vast literature on the unconscious effects of stereotype, mood, and emotional priming,<sup>8</sup> these results suggest that the deontological or utilitarian self is not, in fact, master of its own house, or even of “itself.”

### Thought Is Image-Based

One of the most fundamental challenges to the disembodied, amodal model of human reason is the increasing consensus in the fields of neuroscience and cognitive science that human thought is primarily image-based and modal in character—that is, deriving its structure from sensorimotor patterns. As opposed to a picture of thought as the manipulation of arbitrary, abstract symbols, cognitive scientists such as Lawrence Barsalou have been arguing for a “perceptual symbol” account of human cognition.<sup>9</sup> According to this model, the symbols manipulated in human thought are understood not as pictures but as “records of neural activation that arises during perception” (Barsalou 1999, 583). These records can be abstracted from and combined in various ways in areas of the brain “upstream” from the sensorimotor cortices (what Damasio



1989 refers to as “convergence zones”), but they always remain to some extent grounded in sensorimotor systems.

There is a huge and constantly growing body of evidence in favor of at least some version of the perceptual symbol account.<sup>10</sup> To begin with, it is clear that offline reasoning and language comprehension is imagery based. In a series of classic experiments on the mental rotation of three-dimensional objects, Roger Shepard, Lynn Cooper, and Jacqueline Metzler showed that reaction times for subjects asked to match objects varied consistently as a function of the angular difference of the objects, suggesting that subjects were mentally simulating physical rotation of the objects in real time.<sup>11</sup> Observation of eye saccades during sentence comprehension reveals that subjects’ eyes react to a described situation in an attenuated but similar manner, as they would if the situation were actually in front of them, suggesting that descriptive sentences are serving as cues for imagistic reconstruction of scenarios (Spivey and Geng 2001). It is also apparent that imagination involves the activation of the appropriate sensorimotor regions. Damasio and his colleagues, for instance, have found that achromatopsia (the loss of color perception) also precludes imaging color in recall (Damasio 1985), and Wexler, Kosslyn, and Berthoz (1998) found that the premotor cortices utilized in actual physical rotation of objects are also activated in mental rotation.<sup>12</sup> Further, this sort of sensorimotor simulation is necessary in processing even less obviously perceptual concepts. Damage to sensorimotor systems, for instance, results in category-specific deficits in cognition: Damage to visual areas selectively disrupts the conceptual processing of categories specified by visual features (e.g., birds), while damage to the motor regions selectively disrupts use of categories specified by motor programs (e.g., tools; Warrington and Shallice 1984). Work on imitation has similarly found that both the perception and conceptualization of action and action-related words requires the activation of the appropriate sensorimotor regions of the brain (e.g., Rizzolatti, Fogassi, and Gallese 2001), and Ronald Langacker’s and Leonard Talmy’s work on cognitive grammar and semantics has demonstrated the superior explanatory power of image-schematic over formal analyses of natural language use (Langacker 1987, 1991; Talmy 2000).

Perhaps the strongest argument in favor of something like the perceptual symbol account is that it avoids two fundamental problems that plague amodal symbolic accounts, the transduction problem (how perceptual signals could get “translated” into amodal symbols) and the grounding problem (how arbitrary, abstract symbols could ever come to refer to something in the world.) As Spivey and his colleagues note:

True digital symbol manipulation would require a kind of neural architecture that is very different from the analog two-dimensional maps that might implement image-schematic representations (cf. Regier, 1996) and that we know populate much of the cortex (e.g., Churchland & Sejnowski, 1992; Swindale, 2001). (Spivey, Richardson, and Gozalez-Marquez 2005, 272)

They argue that Occam's razor therefore favors a view whereby much of perception and cognition "is implemented in two-dimensional spatial formats of representation that we know exist in the brain, without the use of discrete symbolic representations that we have yet to witness." Barsalou sums up this argument against classical amodal theories of meaning by concluding that such theories "are unfalsifiable, they are not parsimonious, they lack direct support, they suffer conceptual problems such as transduction and symbol grounding, and it is not clear how to integrate them with theory in neighboring fields, such as perception and neuroscience" (1999, 580).

### Prototype Categories

Objectivist philosophy relies upon classic Aristotelian categories, which have sharp boundaries and clearly defined sufficient and necessary conditions for category membership. If something like the perceptual symbol account of concepts is correct, this would entail the need for a different model of categorization, and evidence from cognitive psychology and linguistics has long suggested that the mode of categorization generally relied upon by human beings differs significantly from the classical account. Much of the early work in this field was done by Eleanor Rosch and her colleagues,<sup>13</sup> who developed a theory of "radial" categorization based upon a "prototype effect." Categories as they are usually active in human minds are based upon certain exemplars or prototypes; membership in the category is then based upon family resemblance and can be a matter of degree (there can be "better" or "worse" members of a given category). For example, most North Americans have an understanding of the category "bird" that is based upon an image of a sparrow, robin, or jay. Most people can switch into a "logical category" mode and acknowledge that chickens, penguins, and ostriches are "birds," but will continue to insist that these are not particularly "good" examples of birds. The same effect can be seen with social categories such as "bachelor": the Pope, for instance, is not a particularly good instance of a "bachelor" (Lakoff 1987).

The dominance in everyday thought of prototype-based categorization is to be expected from the perspective of the perceptual symbol account of cognition. If concepts are a form of sensorimotor simulation, categorization will be based upon imagined exemplars and family resemblances. As Barsalou (1999, 587) notes, categorization understood from this perspective will also not be designed as a rigid net for exhaustively cataloguing and organizing sets of clearly defined objects in the world but rather as a dynamic, contextual, and embodied means of gaining access to categorical inferences—that is, suggestions as to how to interact successfully with encountered objects and situations and to reason about absent (future) entities. Experimental work by W. Kyle Simmons and colleagues<sup>14</sup> found that functional magnetic resonance imaging (fMRI) images of brain activity in sensory regions of subjects asked to confirm the property of categories corresponded significantly with the predicted sensory profile of the category, and studies by Pecher, Zeelenberg, and Barsalou (2003) on property veri-

fiction tasks found significant temporal costs in shifting from one sensory modality to another—suggesting that subjects are activating sensorimotor prototype images in processing categories. As Raymond Gibbs (2006, 83) concludes, “prototypes are not summary abstractions based on a few defining attributes, but are rich, imagistic, sensory, full-bodied mental events.”

### Metaphor and the Importance of Imaginative Extension

If categories as they generally function in human minds are *not* classic Aristotelian categories, this means that reasoning—the classification of events or objects into categories, and the relating of these categories to one another—must involve something other than propositional transformation. The most likely suspect for this “something other” is sensorimotor-based imagination.

That imagination is crucial for moral reasoning has been the central argument of Mark Johnson for two decades,<sup>15</sup> and it has also been argued by philosophers such as Martha Nussbaum. In a discussion of the importance of imagination and literature for morality, Nussbaum notes that moral knowledge

is not simply intellectual grasp of propositions; it is not even simply intellectual grasp of particular facts; it is perception. It is seeing a complex, concrete reality in a highly lucid and richly responsive way; it is taking in what is there, with imagination and feeling. (Nussbaum 1990, 152)

Part of what this sort of moral perception involves is the categorization of novel situations in terms of learned prototypes, which in turn involves a kind of intuitive pattern matching rather than conscious rule following. Johnson discusses the work of Linda Coleman and Paul Kay (Coleman and Kay 1981; refined by Sweetser 1987) on the prototype semantics of the English word *lie*, which seems to demonstrate radial category structure: There are better and worse instances of what constitutes a lie, and subjects’ judgment of whether or not a given act constitutes a lie depends upon a set of implicit criterion that are contextually weighted, as well as upon what Sweetser refers to as “idealized cognitive models” of knowledge and communication.<sup>16</sup> Applying these models to novel situations involves the reactivation of previous sensorimotor experiences, the identification of relevant features in the novel situation, and the recruitment of both implicit and explicit social knowledge. This process cannot be captured in a propositional maxim-following or cost–benefit analysis.

What this means is that moral education will involve training individuals—explicitly or implicitly—to develop more and more sophisticated imagistic models, as well as the ability to extend them in a consistent manner. As Johnson explains, in any kind of reasonably complex situation, “moral reasoning cannot consist merely in the rational unpacking of a determinate concept. Instead, it requires imaginative extensions to nonprototypical cases” (Johnson 1993, 100). Such extension often involves

the use of metaphors or analogies, and thus both internal moral reasoning and public moral debate will often take the form of battling metaphors—which metaphor or analogy best captures the current situation? Is the current U.S. position in Iraq a “quagmire” like Vietnam, or is it like the difficulties encountered in the early period of implementing the Marshall Plan? When a senator vetoes an aid bill to help Sudanese famine victims, is he snatching food out of the mouths of hungry children, or is he helping the Sudanese to learn to stand on their own two feet? How we choose to metaphorically frame a situation is probably the single most crucial element in how we will morally reason and morally *feel* about it, which leads us to our next point.

### Embodied Emotions in Human Cognition

As Gerd Gigerenzer and Reinhard Selten (2001) explain in an introductory essay on the concept of “bounded rationality,” economists and psychologists have, since the 1950s, been moving away from models of behavior that assumed humans are optimal calculators toward models that assume that, in most situations, human beings rely upon domain-specific, “fast and frugal” heuristics.<sup>17</sup> These heuristics generally do not result in rationally optimal results but often outperform general-purpose, time-consuming, and “information-greedy” optimizing strategies, especially in the specific situations of partial knowledge and computational limitations for which they have evolved. A representative example is the “recognition heuristic” (Goldstein and Gigerenzer 2002), whereby an organism presented with a choice between two options—say, two potential food items—simply chooses the one that has been previously encountered over the one that is unknown. It is not hard to imagine how this strategy might be adaptive: Items that have been consumed before by you or by a conspecific are more likely to be edible than an item chosen randomly from the environment, and this might very well outweigh any potential advantage derived from discovering a superior new foodstuff. What is perhaps less intuitively obvious is how such crude heuristics can outperform more “rational” strategies even in quite complex and evolutionarily novel situations, such as stock market investment (Borges, Goldstein, Ortman, and Gigerenzer 1999).

These heuristics and biases often take the form of tacit skills, unformulizable hunches, or—the focus of this section—emotional reactions. In the last decade there has been an explosion of literature on the role of emotions in human reasoning in such fields as behavioral neuroscience, cognitive science, economics, social psychology, and philosophy.<sup>18</sup> Because of space restrictions, I will focus here on the work of perhaps the best known pioneer in this field, Antonio Damasio, and in particular his theory of “somatic marking.” In his discussion of the “body-minded brain,” Damasio points out that the mind evolved in order to ensure the survival of the entire mind-body unit, and he argues that the best way to do this is by “*representing the outside world in terms of the modifications it causes in the body proper*,” that is, representing the

environment by modifying the primordial representations of the body proper whenever an interaction between organism and environment takes place" (Damasio 1994, 230). The result is a set of "somato-motor maps" that provide a "dynamic map of the overall organism anchored in body schema and body boundary" (Damasio 1994, 231). Thus, when we are presented with a situation—or called upon to imagine a situation (neurophysiologically not that different a process)—we rely upon the "dispositional representations" (Damasio 1994, 104) that constitute our full repository of knowledge in order to comprehend it, and these representations inevitably include emotional information. As Damasio (1999, 161) observes, "When we recall an object ... we retrieve not just sensory data but also accompanying motor and emotional data.... We recall not just sensory characteristics of an actual object but the past reactions of the organism to the object." In other words, the images that form the basis of our concepts are somatically "marked" with visceral and often unconscious feelings of "goodness" or "badness," urgency or lack of urgency, and so on, and these feelings play a crucial role in everyday, "rational" decision making.<sup>19</sup>

In *Descartes' Error* (1994), Damasio describes his work with patients suffering from damage to the ventromedial prefrontal cortex, a center of decision making in the brain. The accidents or strokes that had caused this damage had spared these patients' "higher" cognitive faculties—their short- and long-term memories, abstract reasoning skills, mathematical aptitude, and performance on standard IQ tests were completely unimpaired. They were also perfectly physically healthy, with no apparent motor or sensory disabilities. Nonetheless, these patients had been brought to Damasio's attention as a physician because, despite their apparent lack of physical or cognitive impairment, they were no longer functional members of society. In real-life decision-making contexts they were appallingly inept, apparently incapable of efficiently choosing between alternate courses of action, taking into account the future consequences of their actions, or accurately prioritizing the relative importance of potential courses of action.

One representative example is the patient Damasio refers to as "Elliot." Formerly a successful businessman and respected husband and father, Elliot's life began to unravel after he was operated on for a brain tumor, a procedure that involved removing parts of his ventromedial prefrontal cortex. As Damasio (1994, 36) describes it, "Elliot's smarts and his ability to move about and use language were unscathed. In many ways, however, Elliot was no longer Elliot." Elliot needed to be prompted to get up and prepare to go to work in the morning and, once there, seemed incapable of managing his time properly, focusing his attention effectively, or completing even the most routine of tasks:

Imagine a task involving reading and classifying documents of a given client. Elliot would read and fully understand the significance of the material, and he certainly knew how to sort out the documents according to the similarity or disparity of their content. The problem was that he

was likely, all of a sudden, to turn from the sorting task he had initiated to reading one of those papers, carefully and intelligently, and to spend an entire day doing so. Or he might spend a whole afternoon deliberating on which principle of categorization should be applied: Should it be date, size of document, pertinence to the case, or another? The flow of work was stopped. (Damasio 1994, 36)

Understandably, Elliot was soon fired. He proved no more successful in negotiating his way through unemployed life. He developed bizarre collecting habits, took up a bewilderingly diverse array of projects (often dropping them almost as quickly as he had picked them up), entered into questionable financial ventures with disreputable individuals, lost his life's savings, divorced several times, and finally was reduced to living off of social security disability payments.

In the view of Damasio and his colleagues, the problem with ventromedial prefrontal cortex patients such as Elliot is that they lack “somatic markers”—the unconscious, visceral normative weights that ordinarily accompany our representations of the world. This prevents them from unconsciously assigning different values to different options, thereby rendering their “decision-making landscape hopelessly flat” (1994, 51). In any given situation, the number of theoretically possible courses of action is effectively infinite, and the human mind is obviously not capable of running simultaneous analyses of all of them at once. Therefore, the body contributes by biasing the reasoning process with somatic markers—often unconsciously—before it even begins. Patients such as Elliot perform well on abstract moral reasoning and utilitarian calculation tasks because such abstract analyses are artificially simplified. Thrown into a real-life situation, but deprived of the biasing function of somatic markers, they seem to attempt to dispassionately consider *all* of the options theoretically open to them, with the result that they become paralyzed by indecision or simply commit themselves to what appear to outside observers as poorly considered and capriciously selected courses of action.<sup>20</sup> Some researchers have compared their situation to that of alcoholics or compulsive gamblers, in that mere conceptual knowledge that something is harmful is not necessarily adequate—in the absence of the appropriate somatic markers—to motivate a person to actually avoid those harmful things. The problem with impulsive behavior may thus not be too much emotion but rather not *enough* emotion.

Of course, it is important to note that, despite the crucial importance of somatic markers for normal decision making, navigating through the world by means of hunches and know-how does not necessarily lead to advantageous results. Damasio remarks upon the ability of one of his prefrontal patients to calmly steer his way through a skid on icy roads: one sort of scenario where a person's immediate emotional response to a perceived danger (slamming on the brakes) typically leads to unhelpful behavior. More generally, it is clear that human beings are sometimes quite *bad*—that is, not rationally ideal—decision makers, especially when operating in

modern industrial societies, far outside of their ancestral environment. In the field of economics, Daniel Kahneman and the late Amos Tversky have been the best known proponents of a move away from rational choice theory toward more psychologically realistic models that take into account the role of the nonrational heuristics and biases that guide everyday decision making.<sup>21</sup> Dispassionate calculation makes it clear that we are likely to achieve a much better payoff investing \$20 weekly in some conservative mutual fund rather than using that money to buy lottery tickets, but the reasoning processes of many are (incorrectly, in this case) biased by the powerfully positive somatic marker attached to the image of the multimillion-dollar payoff. Similarly, the powerfully negative image of a jetliner falling in flames from the sky prevents many from making the “rational” decision to fly rather than drive, even though commercial airline travel is statistically much safer than automobile travel. George Loewenstein and his colleagues have formulated a “risks as feelings” hypothesis very similar to Damasio’s somatic marker theory, finding that human risk assessment of an imagined scenario is driven largely by vividness, not probability of that scenario actually occurring. In one study (Loewenstein, Weber, Hsee, and Welch 2001), they found that people are willing to pay more for airline travel insurance covering death from “terrorist acts” than for insurance covering death from “all possible causes”! At the other extreme, people tend to be underinsured against emotionally “pallid” risks like floods. Other studies have found that people are also much more responsive to warnings that are linked to individuals and anecdotes than those put in statistical terms.

Thus, while navigating by means of powerful, reasoning-biasing somatic markers must have been adaptive in our dispersed, hunter-gatherer environment of evolutionary adaptation, it sometimes leads us into errors of judgment in the more complex world of settled agricultural societies, especially when modern technology is thrown into the mix. More generally, recognizing the importance of somatic markers in no way requires us to neglect the crucial importance of good old-fashioned “offline,” bloodless rational calculation and algorithmic reasoning—indeed, the fact that humans are even *capable* of such forms of reasoning indicates that they have proven their worth over evolutionary time. The best way to view the work of Damasio and his colleagues is as a corrective to the fetishization of reason in the post-Enlightenment period and as an indication that the objectivist-rationalist model as it has been traditionally formulated has some serious and fundamental limitations.

Relating this work to ethics, a growing number of cognitive scientists and philosophers have come to agree with David Hume and the Stoics that normative judgments are ultimately derived from human emotional reactions. Damasio has argued that our sense of “goodness” corresponds to our sense of bodily wellness, which is not surprising considering that “achieving survival coincides with the ultimate reduction of unpleasant bodily states and the attaining of homeostatic ones, i.e., functionally balanced biological states” (1994, 179). Martha Nussbaum has similarly argued for a

“neo-Stoic” “cognitive-evaluative” view of emotions, which views them as “intelligent responses to the perception of value” (Nussbaum 2001, 1).<sup>22</sup> Although he explicitly wishes to distance himself from “naturalists” who would apply scientific reasoning to the realm of the human, the work of Charles Taylor (1989) is nonetheless very helpful when it comes to the relationship of emotions to ethical values. One of Taylor’s most important points is that human beings, by their very nature, can only operate within the context of a normative space defined by a framework of empirically unverifiable beliefs. The Enlightenment conceit that one can dispense with belief or faith entirely, and make one’s way through life guided solely by the dictates of objective reason, is nothing more than that—a conceit, itself a type of faith in the power of a mysterious faculty, “reason,” to reveal incorrigible truth. In addition to the panoply of “weak evaluations”—such as a preference for chocolate over vanilla ice cream—that we are familiar with, humans are also inevitably moved to assert “strong,” or normative, evaluations. This latter type of evaluation is based on a set of explicit or implicit ontological claims and therefore is perceived as having objective force rather than being a merely subjective whim. For instance, I don’t particularly like chocolate ice cream and believe that the flavor of vanilla ice cream is superior. I don’t, however, expect everyone to share my preference, and am certainly not moved to condemn my wife for preferring chocolate. I am also not inclined to sexually abuse small children, but this feels like a different sort of preference to me: Abusing small children seems *wrong*, and I would condemn and be moved to punish anyone who acted in a manner that violated this feeling. If I were pressed on the matter, this condemnation would be framed, moreover, in terms of beliefs about the value of undamaged human personhood and the need to prevent suffering and safeguard innocence.

In the fields of cognitive science and social psychology, there is a growing set of empirical literature supporting Taylor’s position, showing that people distinguish between “response-dependent” evaluations (Taylor’s weak evaluations), merely conventional evaluations, and moral evaluations (Taylor’s strong evaluations), the last of which are seen even by young children as having an objective quality and being universally applicable.<sup>23</sup> According to Taylor, the distinguishing feature of a strong evaluation is that it is based upon an ontological claim—a claim about how the world really *is*, which gives the evaluation objective force. Evidence from cognitive science and psychology has, however, called into question the directionality of the causal link between ontological belief and moral emotion, suggesting that in many cases the causality may derive from the emotion, with the ontological belief tacked on as an *ex post facto* justification. Jonathan Haidt et al. (1993) found that, when people are presented with verbal scenarios, their affective reactions to them were better predictors of their moral judgments than their claims about harmful consequences, and that people who have a strong negative affective reaction to a scenario often have to struggle to provide a rational justification, with sometimes rather silly results. Simi-



larly, Shaun Nichols (2002) has shown that affectively charged but conventionally neutral scenarios are judged along the same dimensions as moral violations, while Kari Edwards and William von Hippel (1995) found that social attitudes are best changed by altering people's affectively charged intuitions rather than by rational argument and that affectively based opinions were held with much more confidence than rationally based ones. In a recent study, Thalia Wheatley and Jonathan Haidt (2005) demonstrated that judgments of both how disgusting and how morally wrong a given behavior is were made more severe by a flash of hypnotically induced queasiness—directly implicating the importance of “gut reactions” for moral judgments.

In defense of his “social intuitionist” approach to moral judgment, Haidt sums up a vast body of literature suggesting that conscious moral reasoning “is usually a post hoc construction, generated after a judgment has been reached” (2001, 814).<sup>24</sup> These results strongly support the idea that Taylorian strong evaluations arise from affectively laden sensorimotor intuitions, such as disgust, which are then justified through the invocation of ontological claims or rational justification. As Haidt (2001, 814) remarks, “Faced with a social demand for a verbal justification, one becomes a lawyer trying to build a case rather than a judge searching for the truth.” This has led the neuroscientist Joshua Greene, who has studied the various brain regions involved in moral decision making, to conclude that deontological moral principles are ultimately “a kind of moral confabulation” (Greene 2007, 63).

### Gut Reactions and Cognitive Control

As important as work on moral “gut reactions” has been as a corrective to objectivist models, one should not be tempted into dismissing the importance of cognitive control for ethical reasoning and behavior. First of all, although emotional reactions can inspire the creation of ex post facto conscious beliefs, it is also clear that causality often flows in the opposite direction: Explicit beliefs, for instance, frequently engender affective responses in otherwise affectively neutral situations. To take a case of religious belief, being touched by a low-caste Indian on the street is not likely to inspire much of an emotional–moral response in me, while it is likely to inspire physical revulsion and moral outrage in a conservative high-caste Hindu, for whom the caste system is part of the normative structure of the cosmos. Even here, though, we should note that this type of culturally specific ontological belief can elicit strong evaluations from an individual only because it is metaphorically tied into basic physiological responses to uncleanness, contamination, and disgust: Physical “mixing” of castes is revolting to a conservative Hindu because of a worldview in which lower castes are understood as metaphorically “unclean,” which then inspires innate affective responses to pollution and contamination.

Another problem with taking too strongly the affective determination of conscious moral reasoning is that the process can be blocked: It is clear that top–down control,

based on rational beliefs, can override affective reactions. In a cross-cultural study of disgust and moral reactions to various scenarios performed by Haidt et al. (1993), interesting differences were found in the reactions of individuals from cultures varying in their degree of “Westernization” and from various socioeconomic classes. The results suggest that conscious reference to an “ethics of autonomy”—the framework of beliefs concerning the importance of diversity, individual freedom, rights, and so forth that forms the basis of the Western liberal conception of the self—allowed individuals from high socioeconomic classes and living in Westernized cultures to resist converting their affectively negative responses to moralized ones. When presented with certain scenarios—such as a person cutting up the national flag to use as a toilet-cleaning rag or eating a family dog that had been killed in an accident—these individuals felt at least mild disgust or disapproval but often overrode this feeling to declare that these actions were not morally wrong, no matter how much they might personally feel uncomfortable with them.<sup>25</sup>

Joshua Greene et al. (Greene, Sommerville, Nystrom, Darley, and Cohen 2001) document a similar phenomenon in an fMRI study of moral reasoning, where subjects have quite divergent reactions to two versions of a classic thought experiment. In the trolley version, the subject is asked to imagine a runaway trolley rolling down a track toward a Y-shaped rail junction, with one person tied to one set of tracks and five people tied to another set. The subject controls a switch that determines which set of tracks the trolley will be diverted to, and most subjects fairly quickly conclude that they would switch the trolley away from the five people and toward the lone person. In the footbridge version, the subject is asked to imagine being on a footbridge over a single track to which five people have been tied, standing next to a single but rather large individual. The only way to stop the trolley is to push this large person off the bridge to certain death (the subject being too small to stop the trolley), but this will have the effect of saving the other five. The utilitarian calculus is the same, but in this scenario most subjects say they would not be willing to push the person off the bridge. Greene et al. found that emotional regions of the brain were strongly implicated in subjects’ reactions to the footbridge scenario but not to the trolley version, which suggests that the differing responses are the result of a visceral, negative response to the idea of actively pushing a person off a bridge. Interestingly, a subset of subjects in this experiment did endorse the “proper” utilitarian response to the footbridge scenario, and the timing of the activation of various brain regions as documented in the fMRI results suggests that this response involved activating brain regions associated with abstract reasoning and cognitive control in order to override the affective responses triggered by the scenario.<sup>26</sup> These sorts of results provide empirical support for the “high-reason” conviction, endorsed by philosophers since the time of Plato and Xunzi, that the rational faculties *can* and *should* supervise and—when appropriate—override the reactions of our more emotional faculties.

An important question for someone interested in an empirically grounded and practically plausible model of morality is, however, the proper degree of *salience* to be given to this type of conscious, rational control. It is clearly possible. What is less clear is how much of an effect it has, and can be expected to have, on the quotidian functioning of a real moral agent making his or her way through the world. It is important to note that the sort of conscious overriding of automatic emotional responses documented in the Greene et al. 2001 study significantly interfered with the reaction time of participants, which suggests that cognitive control is a fairly costly and time-consuming process. In a recent follow-up study, Greene and his colleagues (Greene, Morelli, Lowenberg, Nystrom, and Cohen 2008) found increased reaction times for utilitarian judgments under cognitive load. This accords with the work of Baumeister et al. (1998), cited above, documenting the so-called “ego-depletion” phenomenon: Conscious supervision is a limited cognitive resource.

It is even more important to note that the sort of moral dilemmas that are the staple of deontological and utilitarian theorizing are radically simplified decision-making frames—ethical decision making in the real world takes place in an environment characterized by time pressure, limited and often inaccurate information, indistinct physical and temporal boundaries, and often only limited or entirely non-existent conscious involvement.<sup>27</sup> Part of the problem with the scenarios such as the footbridge case examined by Greene et al. is that they are artificially simplified (in real life, would a person *really* know for sure that the large man’s body will stop the train or that he or she would be able to get him over the railing in time?), are unmotivated (some subjects *say* they would push the man off, but that does not mean they would actually do it in real life, faced by the real human being), and focus attention on dramatic moments of conscious choice, when in fact most of what counts as moral judgment in real life is probably automatic, or at best semiconscious. Am I performing a conscious utilitarian calculation (or deontic reasoning) every time I pass a homeless person on my way to the bus? I am a professor and catch a student plagiarizing. The rules say he should get an “F” for the class, but he is the first generation in his family to go to college; my “impression” of him is that he is a basically honest, extremely hard-working kid driven to cheat as an act of desperation under extreme pressure; he seems to be experiencing genuine, excruciating remorse; throwing the book at him will result in his scholarship’s being taken away and expulsion from school, and so forth. When I decide what to do about his case, can I really be said to be performing a utilitarian calculation? What are the numerical parameters I am supposed to be working with? (How many points for remorse? Is it a sliding scale, depending on intensity?) It seems that these sorts of situations, rather than the trolley or footbridge situations, are what people face most often as the noticeable tip of the iceberg of moral decision making. Then, of course, there is a huge hidden iceberg of moral “choice” that is operating unconsciously all the time when we simply move through the world:

Are we nice to the check-out girl at the supermarket? Do we notice the old person getting on the bus and give him or her our seat? Do we pay attention when our spouse is talking to us about something important to her, although we have not yet had our coffee and are thinking intensely about utilitarianism and really just want to get back to our writing?

As William Casebeer has noted, the problem with “experimental regimens that isolate ‘dry’ thinking-about-moral-things from ‘wet’ here-I-am-doing-moral-things” is that they “can unnecessarily restrict the scope of the neural mechanisms that are activated” (Casebeer 2003) and produce a distorted picture of what real-life moral reasoning is like.<sup>28</sup> It is clear that the kind of dispassionate reflection and cognitive control required to override somatic–emotional reactions is a very limited resource, which suggests that active attention—though an important human ability—is a rather fragile foundation on which to build an ethical theory. We now turn to a model of ethics that does *not* rely primarily on active cognitive control and algorithmic reasoning but instead aims to cultivate self-activating, automatic, effortless dispositions to act in virtuous manner.

### Virtue Ethics and Effortless Attention

In light of the work on human cognition reviewed above, a model of ethics generally referred to as *virtue ethics*—the dominant model of ethics in many world cultures for thousands of years and one that relies much less on cognitive control than modern Western ethical theories—might begin to look rather appealing. There is some disagreement about what constitutes a virtue ethic, and there are stronger and weaker ways of understanding how virtue ethics differs from deontology or consequentialism.<sup>29</sup> Here I would like to focus on a set of features most relevant to the phenomenon of effortless attention and which might be seen as central, and distinguishing, features of a virtue ethic: the idea that ethical behavior results not from rational, conscious rule following or calculation but rather through the activation of stable, spontaneous, and at least partially emotional dispositions, which cause one to respond ethically to specific situations and to reliably perceive the world in certain normatively desirable ways. For instance, someone possessing the virtue of courage can be counted on to correctly recognize situations that call for courageous action, effortlessly grasp what the appropriate “courageous” response would be, and put that response into action in a spontaneous and unselfconscious manner.

Many, if not most, virtue ethics hold that such virtues do not develop naturally but require various forms of cultural training—a process that begins with some conception of raw human nature and aims to transform it in accordance with some conception of what a “flourishing” or proper human life looks like. This training will inevitably involve some general rules and principles, but because the goal is to produce

a self-activating disposition and a particular mode of perception, the primary tools tend to be role modeling, mentor-guided imaginative extension, and cultural practices that engage the body and the emotions, such as ritual and music. Properly developed virtuous dispositions have been compared to a kind of “know-how,” or skill, and display a degree of flexibility, context sensitivity, and effortlessness that is absent in behavior strictly guided by rules or abstract calculations. As such, virtues would seem to rely on the sort of implicit knowledge and automaticity that, as the research reviewed above suggests, play such a large role in everyday human activity. One might say that the very essence of virtue ethics is a suspicion of the power of cognitive control and a consequent desire to get at and reshape the vast iceberg of human cognition that lies beneath the surface of active consciousness.

As one would expect, the virtue ethical model is not without its problems. In particular, the circularity of the so-called “good person criterion”—the definition of *courage* in any given situation is what the courageous person would do—gives deontologists and utilitarians fits. The reliance upon tradition and social norms to provide models of virtue can easily lead—and historically has led—to authoritarianism and hidebound conservatism, precisely the ills that deontology and utilitarianism were formulated to counteract. Nevertheless, given the features of human cognition outlined above, it may very well be the case that we do not have much of a choice when it comes to ethical models. As attractive as deontology and utilitarianism may be in certain respects, if human beings are simply incapable of consistently employing deontological or utilitarian principles in everyday life, we may be stuck with virtue ethics whether we like it or not.

One of many services that a knowledge of cognitive science can offer scholars in the humanities is a chance to finally settle certain perennial, and previously interminable, theoretical debates (Slingerland 2008). Since the Mencius–Mohist debates in 4th c. b.c.e. China, virtue ethicists have been arguing for the theoretical and practical limitations of rationalist ethics—the latest and most local iteration of this argument kicking off with G. E. M. Anscombe’s “Modern Moral Philosophy” (Anscombe 1958)—and there seems no a priori way to settle the debate. If even some of the discoveries about human cognition outlined above hold up to further scrutiny, this might decisively tilt the balance toward virtue ethics, at least with regard to particular subissues. If emotions play a crucial role in everyday human decision making, a model of ethical training aimed at shaping and properly directing the emotions begins to look quite attractive.<sup>30</sup> If much of our everyday decision making and judgment formation is automatic, nonconscious, and not easily amenable to cognitive control, we might be justifiably concerned about the efficacy of ethical models that rely exclusively on conscious reasoning. If bodily based metaphors and metaphoric blends are central and ineliminable features of human thought, ethical models based on amodal, propositional cognitive processes begin to appear radically inadequate—not necessarily a

knockout blow to deontology and utilitarianism (whatever their cognitive plausibility, one might continue to argue for them as unattainable but crucially regulative ideals) but certainly progress of sorts. Virtue ethics may be characterized by its own theoretical problems and prone to various types of social abuses, but it may be, realistically speaking, the only ethic we've got.

The modern revival of virtue ethics has traditionally looked back to Aristotle for inspiration, and many of the features of human psychology discussed above can be easily accommodated by the Aristotelian conception of the self. As some scholars of Chinese thought have argued, however,<sup>31</sup> there are other virtue ethicists besides Aristotle, and looking beyond Aristotle to traditions such as early Confucianism can give us a broader, richer picture of what a cognitively plausible ethic in action might look like. I will conclude with a very brief sketch of an early Chinese virtue ethic, paying particular attention to the central role given to effortless attention.

### Effortless Attention in Confucian Virtue Ethics

In a previous work on Warring States Chinese thought (Slingerland 2003b), I argued that the “mainstream”<sup>32</sup> thinkers of this period share as a spiritual goal the ideal of *wu wei*, or *effortless action*. Literally meaning no doing or no effort, *wu wei* refers to an ideal state whereby the individual acts, effortlessly and unselfconsciously, in a manner that harmonizes with the normative order of the cosmos. In the case of the Confucians, *wu wei* also involves effortlessly and spontaneously according with the dictates of traditional ethical and ritual standards. In this work I argued that the question of how to achieve *wu wei* is the central religious problematic in early Chinese thought and that the tension involved in consciously striving to achieve effortlessness—trying not to try—drives much of the secondary theorizing about issues such as the character of human nature, the best metaphoric models for self-cultivation, and the relationship of the individual to inherited cultural traditions.

Here I will focus on the case of Confucius, as portrayed in the *Analects*,<sup>33</sup> arguably the earliest detailed account of a virtue ethic in action. The *Analects* clearly portrays effortless action and effortless attention as a goal to be reached only after a long period of intense training and personal reformation, although it is somewhat inconsistent concerning the issue of how the ethically ideal state relates to one's innate nature. Certain passages suggest that cultural training merely refines and “adorns” an already-present innate potential, while others seem to endorse a model whereby cultural forms fundamentally reshape an inherently crude natural endowment.<sup>34</sup> For the purposes of this discussion, I will leave these differences to the side in order to focus on the *Analects*'s portrayal of the state of effortless action and effortless attention as personified in the figure of Confucius near the end of his life.

The Confucian project of self-cultivation, aimed at producing an individual (the “gentleman” or “sage”) endowed with the virtue of Goodness (*ren* 仁),<sup>35</sup> involves a strict regimen of training in which the individual subordinates himself to traditional standards of practice and judgment. This training advances along several fronts, including practice in religious and social rituals, intensive learning of transmitted classics, and training in music and other traditional cultural forms. The aim is to achieve a state where outer form and inner state are perfectly harmonized. As Confucius observes in *Analects* 6.18, commenting on the relationship between inborn nature (“native substance”) and the refinement created by cultural training:

When native substance overwhelms cultural refinement, the result is a crude rustic. When cultural refinement overwhelms native substance, the result is a foppish pedant. Only when culture and native substance are perfectly mixed and balanced do you have a gentleman.<sup>36</sup>

This passage is perhaps the earliest expression of an ideal that later became very important in Confucian writings: the doctrine of holding fast to the “mean” (*zhong* 中). A perfect balance between native substance and cultural refinement is thus the ideal state, although if one is to err it should be on the side of substance.

### Flexibility and Autonomy

One clear virtue ethical characteristic of Confucius’s thought is that his program of self-cultivation is designed to transform the student into a particular type of person, not to merely win assent to a set of principles or teach a method of calculation. For instance, in 9.24 the Master remarks:

When a man is rebuked with exemplary words after having made a mistake, he cannot help but agree with them. However, what is important is that he change himself in order to *accord* with them. When a man is praised with words of respect, he cannot help but be pleased with them. However, what is important is that he actually *live up* to them. A person who finds respectful words pleasing but does not live up to them, or agrees with others’ reproaches and yet does not change—there is nothing I can do with one such as this.

Nominal assent to the Confucian Way is insufficient—one must love the Way and strive to embody it in one’s person. This is why Confucian moral training involves long periods of time engaged in concrete, embodied practices such as ritual and music—the latter of which included singing, the playing of musical instruments, and dancing.<sup>37</sup>

The “offline,” theoretical aspect of Confucian training emphasizes the absorption of proper language and cognitive models acquired through study of the classics, memorized and rehearsed until they become fully internalized and unconscious patterns of thought. It also includes the consideration and group discussion of case examples from the past. Here is a typical exchange:

5.19 Zizhang said, “Prime Minister Ziwen<sup>38</sup> was given three times the post of prime minister, and yet he never showed a sign of pleasure; he was removed from this office three times, and yet never showed a sign of resentment. When the incoming prime minister took over, he invariably provided him with a complete account of the official state of affairs. What do you make of Prime Minister Ziwen?”

The Master said, “He certainly was dutiful.”

“Was he not Good?”

“I do not know about that—what makes you think he deserves to be called Good?”

We have to imagine that this conversation, though initiated by the disciple Zizhang, is taking place in the company of assembled disciples, and Confucius’s judgment of this historical figure is thus intended as a general lesson for them all. One of the most interesting—and to some analysts, most infuriating—aspects of the *Analects* is Confucius’s failure to precisely define his supreme virtue of Goodness, as well as his reluctance to pronounce any particular person to be Good. When asked if a particular individual might be considered Good, Confucius typically goes no farther than conceding to the person one of the lesser virtues as in the exchange above. The reason for this is not only that true Goodness is extremely difficult to obtain, but also that—as the master virtue of being a perfected person—it involves a kind of flexibility and grace that is hard to perceive in second-hand accounts of contemporaries or historical figures.

The difficulty of conveying genuine Goodness in words is also the reason that much of Confucius’s pedagogy is focused on context-specific injunctions or bits of advice tailored to the needs of individual disciples. Indeed, some commentators believe that the reservations Confucius expresses about Prime Minister Ziwen in 5.19 may have less to do with his actual opinions of this figure than with his concerns about the moral qualities of the disciple Zizhang himself. His flexibility in tailoring his pedagogical message to the needs of the student is famously illustrated in 11.22:

Zilu asked, “Upon learning of something that needs to be done, should one immediately take care of it?”

The Master replied, “As long as one’s father and older brothers are still alive, how could one possibly take care of it immediately?”<sup>39</sup>

[On a later occasion] Ran Qiu asked, “Upon learning of something that needs to be done, should one immediately take care of it?”

The Master replied, “Upon learning of it, you should immediately take care of it.”

Zihua [having observed both exchanges], inquired, “When Zilu asked you whether or not one should immediately take care of something upon learning of it, you told him one should not, as long as one’s father and elder brothers were still alive. When Ran Qiu asked the same question,



however, you told him that one should immediately take care of it. I am confused, and humbly ask to have this explained to me.”

The Master said, “Ran Qiu is overly cautious, and so I wished to urge him on. Zilu, on the other hand, is too impetuous, and so I sought to hold him back.”

This is a paradigmatic example of how the Master’s teachings were variously formulated depending upon the individual needs of his students—a Confucian version of the Buddhist practice of *upaya*, or “skillful means.”<sup>40</sup>

This is not to deny that Confucius taught his disciples certain general principles, such as the famous “negative Golden Rule”—“Do not do to others what you would not want done to you”—described in 15.24 as “a single teaching that can be a guide to conduct throughout one’s life.” Modern Western students of the text, especially those with backgrounds in analytic philosophy, often make such general principles the central focus of their interpretations of the *Analects*,<sup>41</sup> but I believe this leads to a distortion of how the text was traditionally intended and used in premodern China. Rather than focusing primarily on such principles, Confucius’s pedagogical technique is designed to create in his students an ability to apply Confucian ideals in a context-sensitive and flexible manner. This *feel*—a type of rational emotion or embodied know-how—allows the perfected Confucian sage to act in accordance with the ethical ideals and rules while at the same time displaying a level of autonomy and flexibility impossible for one who is merely going by the book. Indeed, one cannot be said to have properly mastered a set of principles until one knows how to apply them skillfully and in a context-sensitive manner. As Confucius notes in 13.5:

Imagine a person who can recite the several hundred Odes by heart but, when delegated a governmental task, is unable to carry it out, or when sent abroad as an envoy, is unable to engage in repartee. No matter how many Odes he might have memorized, what good are they to him?

The goal is to develop a *sense* for the practice and not to be overly constrained by its formal structure. This sort of situation-centered reasoning resembles Aristotelian *phronesis* (practical wisdom), and ultimately what is right in the ethical realm corresponds to what the gentleman (i.e., the good person) would do.

Indeed, the entirety of Book Ten of the *Analects*—an extended account of Confucius’s ritual behavior—can be seen as a model of how the true sage flexibly adapts the principles of ritual to concrete situations. While this chapter is often skipped over in embarrassment by Western scholars sympathetic to Confucianism but nonetheless appalled by the seemingly pointless detail and apparent rigidity of behavior (“With a black upper garment he would wear a lambskin robe; with a white upper garment he would wear a fawnskin robe; and with a yellow upper garment he would wear a fox-fur robe.”—10.6), this discomfort is based upon a fundamental misunderstanding.

While the scope and detail of Confucian ritual certainly (and quite rightly) seem alien to a modern Westerner, it is important to understand that what is being emphasized in this chapter is the ease and grace with which the Master embodies the spirit of the rites in every aspect of his life—no matter how trivial—and accords with this spirit in adapting the rites to new and necessarily unforeseeable circumstances.

Consider some snapshots of Confucius in action from Book Ten, with occasional comments interposed for the sake of clarification:

10.2 At court, when speaking with officers of lower rank, he was pleasant and affable; when speaking with officers of upper rank, he was formal and proper. When his lord was present, he combined an attitude of cautious respect with graceful ease.

Confucius effortlessly adapted his countenance and behavior to the demands of the social situation; he was not overly familiar with his colleagues nor obsequious to his superiors.

10.3 When called upon by his lord to receive a guest, his countenance would become alert and serious, and he would hasten his steps. When he saluted those in attendance beside him—extending his clasped hands to the left or right, as their position required—his robes remained perfectly arrayed, both front and back. Hastening forward, he moved smoothly, as though gliding upon wings. Once the guest had left, he would always return to report, “The guest is no longer looking back.”

According to most commentators, Confucius is here serving as the “Master of Reception”; to his left would be the “Master of Introductions” and to his right the “Supreme Master of Ceremonies.” Each had a set speech to deliver, and after each speech they would bow both to their guests and to one another. As for his final report, it was the custom in ancient China for the guest to turn around and bow repeatedly as he left; the host (or the host’s proxy, in this case Confucius) could return to his place only after this process was over. Here we see Confucius fulfilling his ritual duties with both precision and grace.

10.23 When receiving a gift from a friend—even something as valuable as a cart or a horse—he did not bow unless it was a gift of sacrificial meat.

There was apparently no specific clause in the rites that dictates this specific response to this particular situation; rather, Confucius, by virtue of his sensitivity to the ritual value of sacrificial meat relative to a sumptuous—but nonceremonial—gift, simply knows how to respond properly.

That Confucius’s flexibility in applying ritual is the theme of Book Ten is made clear in the last passage, 10.27:

Startled by their arrival, a bird arose and circled several times before alighting upon a branch. [The Master] said, “This pheasant upon the mountain bridge—how timely it is! How timely it is!” Zilu saluted the bird, and it cried out three times before flying away.

This poetic, somewhat cryptic passage seems like a non sequitur at the end of a chapter devoted to short, prosaic descriptions of ritual behavior—unless, that is, it is seen as a thematic summary of the chapter as a whole. Timeliness (*shi*) is Confucius's particular forte, and indeed he is known to posterity (through the efforts of Mencius) as the “timely sage”: the one whose ritual responses were always appropriate to circumstances. As Mencius explains in 5:B:1:

When Confucius decided to leave Qi, he emptied the rice from the pot before it was even done and set out immediately. When he decided to leave Lu he said, “I will take my time, for this is the way to leave the state of one’s parents.” Moving quickly when it was appropriate to hurry, moving slowly when it was appropriate to linger, remaining in a state or taking office when the situation allowed—this is how Confucius was.... Confucius was the sage whose actions were timely.

We have thus seen that, by internalizing the rules and conventions that define a practice such as the rites, one is able at the same time to achieve a certain degree of autonomy in applying them. This autonomy, in turn, can allow one a certain degree of critical distance: Once the meaning embodied in the norms is grasped, the norms themselves can potentially be evaluated, criticized, or even altered. Hence we have the famous passage, *Analects* 9.3, where Confucius accedes to a modification in the rites:

The Master said, “A ceremonial cap made of linen is prescribed by the rites, but these days people use silk. This is frugal, and I follow the majority. To bow before ascending the stairs is what is prescribed by the rites, but these days people bow after ascending. This is arrogant, and—though it goes against the majority—I continue to bow before ascending.”

According to commentators, the linen cap specified by ritual was an elaborate affair—consisting of many layers and involving intricate stitching—and Confucius’s contemporaries had begun replacing it with a simpler silk version. Confucius apparently feels that this does not interfere with its basic function. When approaching a ruler or other superior sitting on a raised dais, ritual dictates bowing before ascending the stairs, but Confucius’s contemporaries had taken to ascending the stairs and only bowing when directly before their ruler. This is a more substantial change, and Confucius rejects it as not ritually proper. This passage thus describes the sort of judgment and flexibility that can be exercised by an accomplished ritual practitioner. Rites are expressive of a certain sense or feeling, and thus an alteration in the actual rite is permissible if it will not—in the opinion of one who has fully mastered the rites and thus internalized it—alter its essential meaning.

### Spontaneity, Harmony, and Ease

After extended training in traditional cultural practices, the emotions are ultimately harnessed by Confucians to produce moral behavior that springs spontaneously from

personal inclination. This means that, for the ethically perfected person, proper action requires no conscious thought or effort and is governed spontaneously by what we might term effortless attention. Music was considered by the early Confucians to be one of the most powerful tools for affecting this sort of emotional transformation, and the metaphor of musical perfection also served for Confucius as a metaphor for the perfected spiritual state:<sup>42</sup>

3.23 The Master was discussing music with the Grand Music Master of Lu. He said, "What can be known about music is this: when it first begins, it resounds with a confusing variety of notes, but as it unfolds, these notes are reconciled by means of harmony, brought into tension by means of counterpoint, and finally woven together into a seamless whole. It is in this way that music reaches its perfection."

Music here serves as a model or metaphor for the process of self-cultivation: starting in confusion, passing through many phases and culminating in a state of *wu wei* perfection. Music as a model of harmony is also the theme of 8.8, which succinctly summarizes the process of Confucian self-cultivation in three phrases: "The Master said, 'Find inspiration in the *Odes*,<sup>43</sup> take your place through ritual, and achieve perfection with music.'" Steps 1 and 2 represent, respectively, cognitive shaping through learning and behavioral shaping through ritual training. In the third stage, the joy inspired by the powerfully moving music of the ancients brings the cognitive and behavioral together into the unselfconscious, effortless perfection that is *wu wei*. *Mencius* 4:A:27, which invokes the metaphor of dance, represents perhaps the best commentary on this passage:

The substance of benevolence (*ren*) is the serving of one's parents; the substance of rightness is obeying one's elders; the substance of wisdom is to understand benevolence and rightness and to not let them go; the substance of ritual propriety is the regulation and adornment of benevolence and rightness; and the substance of music is the joy one takes in benevolence and rightness. Once such joy is born, it cannot be stopped. Once it cannot be stopped, then one begins unconsciously to dance it with one's feet and wave one's arms in time with it.

That the inner state of the actor be harmonized with outer behavior, to the point that it becomes effortless and unselfconscious, is crucial for Confucius. This is a clear point of departure from most classic formulations of deontology. In a famous passage from the *Groundwork of the Metaphysics of Morals* (Kant, 1785/1964, 65–66), for instance, Kant argues that a shopkeeper who refrains from cheating his customers simply because he is honest—that is, he is unselfconsciously honest "out of inclination" (*aus Neigung*)—cannot be considered genuinely moral. True morality requires consciously acting "out of duty" (*aus Pflicht*), not merely "in conformity with duty" (*Pflichtmäßig*): consciously rehearsing the moral law, and then deliberately putting that law into practice through an act of cognitive control. In the *Analects*, there is not only no place for duty-bound action in the Kantian sense, but such behavior would be looked down

upon as forced and inauthentic. The praise for the sage-king Shun by Mencius—that “his actions flowed from benevolence and rightness, he did not merely put them into practice” (4:B:19)—is in this sense quite anti-Kantian: Shun is considered virtuous precisely *because* he acted *aus Neigung* rather than *aus Pflicht*. As the famous Song commentator Zhu Xi explains in his commentary on *Mencius* 4:B:19, “Benevolence and rightness were already rooted in Shun’s heart, and all of his actions sprang from there; it is not the case that he merely valued benevolence and rightness and therefore forced himself to put them into practice.”

Conscious deliberation and cognitive control, though necessary in the early stages of self-cultivation, are thus hallmarks of inauthenticity and ethical clumsiness. Despite the importance of hard work, struggle, and endurance in the early stages of Confucian self-cultivation, it is fully expected that any Confucian gentleman worth the title will have come to fully internalize—behaviorally automatize—such practices as ritual and music. The result will be an individual with an entirely transformed set of dispositions. This theme is captured in Confucius’s spiritual autobiography in *Analects* 2.4:

The Master said, “At fifteen I set my mind upon learning; at thirty I took my place in society; at forty I became free of doubts; at fifty I understood Heaven’s Mandate; at sixty my ear was attuned; and at seventy I could follow my heart’s desires without overstepping the bounds of propriety.”

We can see the reshaping of Confucius’s initial dispositions as taking place in three pairs of stages. In the first pair (stages 1 and 2), the aspiring gentleman commits himself to the Confucian Way, submitting to the rigors of study and ritual practice until these traditional forms have been internalized to the point that he is able to “take his place” among others. In the second pair, the practitioner begins to feel truly at ease with this new manner of being and is able to understand how the Confucian Way fits into the order of things and complies with the will of Heaven. The clarity and sense of ease this brings with it leads to the final two stages, where one’s dispositions have been so thoroughly harmonized with the dictates of normative culture that one accords with them spontaneously. In such a state of *wu wei*, one can “follow one’s heart’s desires”—that is, surrender to automaticity and effortless attention—while still according perfectly with the dictates of traditional etiquette and morality. Some of the work on automaticity and spontaneous judgment formation reviewed above concludes that, in certain situations, automatic judgments may be more accurate than those arrived at through tortuous reasoning, and automatic behaviors may be both more efficacious than those under active cognitive control and easily disrupted by conscious interference. All of the early Confucians believed, realistically or not, that the state of supreme effortlessness enjoyed by the gentlemen would endow him with perfect efficacy in the world, including the ability to influence and transform others. Confucian *wu wei* is thus not merely a personal ideal but also has crucial political and

soteriological effects,<sup>44</sup> all of them flowing from a state of unselfconscious and effortless attention.

### Conclusion

Allow me to quote an apropos call to arms issued by David Hume over 200 years ago:

Men are now cured of their passion for hypotheses and systems in natural philosophy, and will hearken to no arguments but those which are derived from experience. It is full time that they should attempt a like reformation in all moral disquisitions; and reject every system of ethics, however subtle or ingenious, which is not founded on fact and observation. (Hume, 1777/1976, 174–175)

What I would like to suggest is that the body of empirical evidence emerging from cognitive science, cognitive linguistics, neuroscience, social psychology, and primatology that I have reviewed above indicates that the so-called “virtue ethical” model best describes how real human beings actually *engage* in moral reasoning (a descriptive claim) and therefore provides us with the best framework for formulating a psychologically realistic model of moral reasoning and moral education (a normative claim).<sup>45</sup> Since we are invoking Hume, we must, of course, guard against the error of slipping from *is* into *ought*. I think, however, that psychological feasibility is an important desideratum for any ethical theory. If deontology and utilitarianism require us to think or behave in manners that are simply not possible or sustainable in quotidian life, this should temper our enthusiasm for adopting them as moral ideals.

Traditional world cultures, such as early Confucianism, provide us with a variety of virtue ethical models replete with profound psychological insights, as well as clever and effective methods for transforming and redirecting our nonconscious selves. As we learn more about how the human mind works, such ethical traditions take on a more than merely antiquarian interest, helping us to fill in an enormous blind spot that has hindered modern Western ethical thinking for the past several hundred years. The transformation required in ethics is similar to that currently taking place in certain public health campaigns, which have traditionally tended to focus on conveying specific bits of conscious knowledge (“wearing condoms prevents transmission of the AIDS virus”). As a recent *New York Times* piece (Duhigg 2008) observes, some of these campaigns are now borrowing techniques from corporate marketing departments<sup>46</sup> and focusing more on changing unconscious habits rather than transmitting conscious knowledge. If much of our everyday behavior is driven by effortless attention and spontaneous habits, modern ethical theorists need, in a similar way, to pay more attention to how such tacit modes of perceiving and acting function, how they are formed, and how—when our cognitive control systems deem it necessary—they can be reformed or redirected.

## Notes

1. Van Norden (2007) provides a helpful summary of the Mencius–Mohist debate.
2. See Johnson (1987, ix–xiii) for a characterization of “objectivism” as I am using the term.
3. For a helpful, recent discussion of the objectivist model of concepts and the embodied alternative, with a wealth of experimental evidence, see Gibbs (2006, chapter 4).
4. See especially Ryle’s famous distinction between “knowing how” and “knowing that” (Ryle 1949; Polanyi 1967).
5. See Kunst-Wilson and Zajonc (1980) and the literature review in Zajonc (1980). For a very helpful review article on the topic of emotional processing and automaticity, see Pessoa (2005), which concludes that emotional processing appears to enjoy a degree of autonomy from conscious “top–down” processes, but not complete automaticity.
6. T. Wilson and Schooler (1991); cf. T. Wilson, Kraft, and Dunn (1989) and T. Wilson (2002).
7. See especially Haidt, Koller, and Dias (1993), Haidt (2001), and Wheatley and Haidt (2005).
8. For a recent literature survey and account of the role and power of the “adaptive unconscious,” see T. Wilson (2002).
9. Also see Johnson (1987), Lakoff (1987), and Langacker (1987) for similar arguments that linguistic representations have an analog, spatial component, as well as Damasio’s claim that “images are the main content of our thought” (1994, 107).
10. For recent reviews, see the essays collected in Pecher and Zwaan (2005).
11. See the essays collected in Shepard and Cooper (1982).
12. Cf. the recent study by Winawer, Witthoft, Huk, and Boroditsky (2005) demonstrating that imagined and implied motion appear to recruit the same neural circuits as are involved in viewing actual motion.
13. Rosch (1973) and Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976); also see Lakoff (1987).
14. W. K. Simmons, Pecher, Hamann, Zeelenberg, and Barsalou (2003), reviewed and discussed in K. W. Simmons and Barsalou (2003).
15. See especially Johnson (1987, 1993), as well as Lakoff and Johnson (1999).
16. See Johnson (1993, 91–98) for a discussion of this work.
17. The term “bounded rationality” was coined by Herbert Simon in 1956. For more on the topic, see the essays gathered in Gigerenzer, Todd, and Group (1999) and Gigerenzer and Selten (2001).
18. For just a sampling, see Damasio (1994), LeDoux (1996), Nussbaum (2001), Prinz (2007), Sinnott-Armstrong (2008a, 2008b), and Solomon (2003, 2004).

19. Cf. Gerald Edelman's (1992, 121) concept of "value memory," where "current value-free perceptual categorization interacts with value-dominant memory" through the online construction of scenes.
20. Cf. Gigerenzer and Selten's (2001, 9) characterization of emotions as "effective stopping rules for search and a means for limiting search spaces."
21. See especially the essays gathered in Kahneman, Slovic, and Tversky (1982) and Kahneman and Tversky (2000). Kahneman and Tversky tend to emphasize the suboptimal nature of ordinary decision making. In contrast, the "bounded rationality" movement discussed above contends that, in ecologically realistic situations, the rationally optimal decision (the economist's usual standard for the "right" decision) is not necessarily the best or most adaptive because of the costs of information gathering and processing or lost opportunities.
22. Cf. Lakoff and Johnson's argument that normative judgments arise from background feelings and that bodily ease or "well-being" is the basic source domain for our understanding of "goodness" (Lakoff and Johnson 1999, 290–292). For representative recent works on morality and emotion, see the essays collected in Solomon (2004), Nichols (2004), and Prinz (2007).
23. See the literature reviewed in Nichols (2002, 221–222) and Nichols and Folds-Bennett (2003) for results with 4- to 6-year-old children.
24. See Haidt (2007) for a more recent literature review.
25. See Jones (2007) for a recent review article on disgust and moral judgments.
26. Greene, Nystrom, Engell, Darley, and Cohen (2004) also found increased activity in brain regions associated with cognitive control prior to utilitarian judgments.
27. Daniel Dennett's (1995, 495–502) example of a philosophy fellowship competition is a helpful illustration of some of the features of realistic, "myopic and time-pressured" decision making.
28. He and Patricia Churchland (Casebeer and Churchland 2003, 187–188) have argued for a more "ecologically valid experimental regime" that takes into account that real moral reasoning is "hot" (affective states are a crucial part), social (decisions are not made in a social vacuum but are subject to social cues), distributed (embedded in a large web of stimulation), organic (context sensitive), genuine (personally involved rather than abstract), and directed (about actual things in the world).
29. The essays in Crisp and Slote (1997) provide a helpful introduction to virtue ethics. Also see Van Norden (2007, 33–37) for a discussion of "radical" versus "moderate" versions of virtue ethics.
30. In a piece composed some years ago that included a discussion of the relative advantages of virtue ethics and deontology (Slingerland 2001), an anonymous referee pointed out to me that Kant, to take one example, recognized the need for pragmatic, situation-specific judgment in applying universal principles to specific situations and that, therefore, the gulf between a sophisticated Kantianism and virtue ethics is not as large as it might at first seem. Indeed, defenders



of the Kantian tradition have long argued that what is attacked by virtue ethicists is often a caricature of the true Kantian position, which in fact gives a prominent role to practical reason (e.g., Schneewind 1990). However we understand Kantian practical reason, one point on which Kant was absolutely unequivocal is that emotions and habits—as accidental, empirical, *heteronomous* qualities of human beings—can have absolutely nothing to do with morality. To the extent that more recent takes on Kant attempt to give a positive role to emotions or to analogical imagination (e.g., Johnson 1987), they seem to simply confirm the point that traditional deontology is plagued by deep problems.

31. For example, S. Wilson (1995), Chong (1998), Ivanhoe (2000), Slingerland (2001), and Van Norden (2007).

32. A somewhat anachronistic term for the so-called Daoists and Confucians whose worldview later came to dominate East Asian thought.

33. The *Analects* (*lunyu* 论语; lit. “Classified Sayings”) was probably put together sometime after the death of the historical Confucius (551–479 b.c.e.). Our present version is a somewhat heterogeneous collection of material from different time periods, probably representing different lineages of disciplines, although scholars differ in their identification of the different strata, as well as in the significance they attribute to these differences.

34. Confucius’s two immediate followers in the Warring States, Mencius and Xunzi, each focus on and develop one of these two themes, with Mencius advocating a more thoroughgoing “internalism” and Xunzi formulating a fairly extreme “externalist” model of self-cultivation.

35. In the *Analects*, *ren* refers to the highest of virtues, the overarching virtue of being truly good or truly human (being cognate with the term for person, *ren* 人). In post-*Analects* texts, it has the more specific sense of empathy or kindness between human beings—especially for a ruler toward his subjects—and in such contexts is usually translated as *benevolence*. Although we see hints of this later usage in the *Analects* (12.22, 17.21), it is much more commonly used there in the more general sense of goodness.

36. All citations from the *Analects* are taken from Slingerland (2003a).

37. On the role of the arts in Confucian self-cultivation, see Eno (1990) and Gier (2001).

38. A prime minister who was renowned for his integrity and devotion to the state.

39. That is, one should continue to defer to the judgment of one’s elders and not take the initiative.

40. For more on the contextual nature of Confucius’s teachings, refer to Setton (2000).

41. For a classic example of Confucius interpreted as a deontologist, with a particular focus on the Negative Golden Rule, see Roetz (1993, especially 133–148).

42. See Cook (1995) and Brindley (2006) for excellent discussions of this themes. In this context, it is very telling that the utilitarian Mozi rejected Confucian music, failing to see the moral value

of music because the cultivation of emotions and dispositions played no role in his extremely rationalistic ethical scheme.

43. One of the classic Confucian texts.

44. See Slingerland (2003b) for more on *wu wei* as a spiritual and political ideal.

45. Observations along these lines have been made by neuroscientifically literate philosophers such as Flanagan (1991), Churchland (1998), and Casebeer (2003). John Doris (1998, 2002) and Gilbert Harman (1999) have famously argued that findings in social psychology suggest that stable human character traits do not exist, which, of course, would call into question the very *raison d'être* of virtue ethics (This was first brought to my attention by Eric Hutton, personal communication, September 2005). In fact, large-scale meta-analyses—for example, Roberts, Kuncel, Shiner, Caspi, and Goldberg (2007)—show that, while situational effects can be quite strong, stable personality traits have at least as strong an effect. Also see Kupperman (2001), Kamtekar (2004), and Hutton (2006) for philosophical responses to Doris's and Harman's positions.

46. The marketing departments, of course, were, in turn, inspired by precisely the sort of social psychology research reviewed above. The fact that people flogging toothpaste and air fresheners are more empirically up-to-date when it comes to human cognition than most philosophers says much about the ivory tower quality of contemporary academic ethics.

## References

- Anscombe, G. E. M. 1958. Modern moral philosophy. *Philosophy* 33 (124):1–19.
- Aristotle. 1999. *Nicomachean ethics*. 2nd ed. Indianapolis: Hackett.
- Bargh, J., and T. Chartrand. 1999. The unbearable automaticity of being. *Am. Psychol.* 54:462–479.
- Barsalou, L. 1999. Perceptual symbol systems. *Behav. Brain Sci.* 22:577–609.
- Baumeister, R. 1984. Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *J. Pers. Soc. Psychol.* 46:610–620.
- Baumeister, R., E. Bratslavsky, M. Muraven, and D. Tice. 1998. Ego depletion: Is the active self a limited resource? *J. Pers. Soc. Psychol.* 74:1252–1265.
- Borges, B., D. Goldstein, A. Ortmann, and G. Gigerenzer. 1999. Can ignorance beat the stock market? In *Simple heuristics that make us smart*, ed. G. Gigerenzer, Peter Todd, and the ABC Research Group. New York: Oxford University Press.
- Brindley, E. 2006. Music, cosmos, and the development of psychology in early China. *Toung Pao* 92:1–49.
- Casebeer, W., and P. Churchland. 2003. The neural mechanisms of moral cognition: A multi-aspect approach to moral judgment and decision-making. *Biol. Philos.* 18:169–194.

- Casebeer, W. D. 2003. Moral cognition and its neural constituents. *Nat. Rev. Neurosci.* 4:840–845.
- Cavell, S. 1979. *The claim of reason*. New York: Oxford University Press.
- Chong, K.-C. 1998. Confucius' virtue ethics: Li, Yi, Wen and Chih in the Analects. *J. Chin. Philos.* 25:101–130.
- Churchland, P. 1998. Toward a cognitive neurobiology of the moral virtues. *Topoi* 17:83–96.
- Churchland, P., and T. Sejnowski. 1992. *The computational brain*. Cambridge: MIT Press.
- Clark, A. 1997. *Being there: Putting brain, body and world together again*. Cambridge: MIT Press.
- Coleman, L., and P. Kay. 1981. Prototype semantics: The English word lie. *Language* 57:26–44.
- Cook, S. 1995. *Unity and diversity in the musical thought of warring states China*. Unpublished Ph.D. dissertation, University of Michigan.
- Crisp, R., and M. Slote, eds. 1997. *Virtue ethics*. New York: Oxford University Press.
- Damasio, A. 1985. Disorders of complex visual processing: Agnosias, achromatopsia, Balint's syndrome, and related difficulties of orientation and construction. In *Principles of behavioral neurology*, ed. M. M. Mesulam. Philadelphia: Davis, 259–288.
- Damasio, A. 1989. The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* 1:123–132.
- Damasio, A. 1994. *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Damasio, A. 2003. *In search of Spinoza: Joy, sorrow, and the feeling brain*. New York: Harvest House.
- Damasio, A., and H. Damasio. 1994. Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In *Large-scale neuronal theories of the brain*, ed. C. K. J. Davis. Cambridge: MIT Press, 61–74.
- Dennett, D. 1995. *Darwin's dangerous idea: Evolution and the meaning of life*. New York: Simon & Schuster.
- Doris, J. 1998. Persons, situations, and virtue ethics. *Nous* 32:504–530.
- Doris, J. 2002. *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.
- Duhigg, C. 2008. Warning: Habits may be good for you. *New York Times*. (Business, July 13).
- Edelman, G. 1992. *Bright air, brilliant fire: On the matter of the mind*. New York: Basic Books.
- Edwards, K., and W. von Hippel. 1995. Hearts and minds: The priority of affective versus cognitive factors in person perception. *Pers. Soc. Psychol. Bull.* 21:996–1011.
- Eno, R. 1990. *The Confucian creation of heaven*. Albany: State University of New York Press.

- Flanagan, O. 1991. *Varieties of moral personality: Ethics and psychological realism*. Cambridge: Harvard University Press.
- Gazzaniga, M. 1998. *The mind's past*. Berkeley: University of California Press.
- Gibbs, R. 2006. *Embodiment and cognitive science*. Cambridge: Cambridge University Press.
- Gier, N. 2001. The Dancing Ru: A Confucian aesthetics of virtue. *Philos. East West* 51:280–305.
- Gigerenzer, G., and R. Selten, eds. 2001. *Bounded rationality: The adaptive toolbox*. Cambridge: MIT Press.
- Gigerenzer, G., P. M. Todd, and A. R. Group. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.
- Goldstein, D., and G. Gigerenzer. 2002. Models of ecological rationality: The recognition heuristic. *Psychol. Rev.* 109:75–90.
- Greene, J. 2007. The secret joke of Kant's soul. In *Moral psychology: The neuroscience of morality: Emotion, disease, and development*. vol. 3. ed. W. Sinnott-Armstrong. Cambridge: MIT Press, 35–79.
- Greene, J. D., S. A. Morelli, K. Lowenberg, L. E. Nystrom, and J. D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107:1144–1154.
- Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44:389–400.
- Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293:2105–2108.
- Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* 108:814–834.
- Haidt, J. 2007. The new synthesis in moral psychology. *Science* 316:998–1002.
- Haidt, J., S. Koller, and M. Dias. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *J. Pers. Soc. Psychol.* 65:613–628.
- Harman, G. 1999. Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society* 99:315–331.
- Hume, D. (1777/1976). *Enquiries concerning human understanding and concerning the principles of morals*. 3rd ed. Oxford: Clarendon.
- Hutton, E. 2006. Character, situationalism, and early Confucian thought. *Philos. Stud.* 127:37–58.
- Ivanhoe, P. J. 2000. *Confucian moral self cultivation*. 2nd ed. Indianapolis: Hackett.
- Johnson, M. 1987. *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

- Johnson, M. 1993. *Moral imagination: Implications of cognitive science for ethics*. Chicago: University of Chicago Press.
- Jones, D. 2007. The depths of disgust. *Nature* 447:768–771.
- Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D., and A. Tversky. 2000. *Choices, values, and frames*. Cambridge: Cambridge University Press.
- Kamtekar, R. 2004. Situationalism and virtue ethics on the content of our character. *Ethics* 114:458–491.
- Kant, I. (1964). *Groundwork of the metaphysics of morals*. Translated by H. J. Paton. New York: Harper Torchbooks. (Original work published 1785)
- Kunst-Wilson, W., and R. Zajonc. 1980. Affective discrimination of stimuli that cannot be recognized. *Science* 207:557–558.
- Kupperman, J. 2001. The indispensability of character. *Philosophy* 76:239–250.
- Lakoff, G. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, G., and M. Johnson. 1999. *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.
- Langacker, R. 1987. *Theoretical prerequisites*. vol. 1. *Foundations of cognitive grammar*. Stanford: Stanford University Press.
- Langacker, R. 1991. *Descriptive applications*. vol. 2. *Foundations of cognitive grammar*. Stanford: Stanford University Press.
- LeDoux, J. 1996. *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.
- Loewenstein, G., E. Weber, C. Hsee, and N. Welch. 2001. Risk as feelings. *Psychol. Bull.* 127:267–286.
- Muraven, M., D. Tice, and R. Baumeister. 1998. Self-control as a limited resource: Regulatory depletion patterns. *J. Pers. Soc. Psychol.* 74:774–789.
- Nichols, S. 2002. Norms with feeling: Towards a psychological account of moral judgment. *Cognition* 84:221–236.
- Nichols, S. 2004. *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.
- Nichols, S., and T. Folds-Bennett. 2003. Are children moral objectivists? Children's judgments about moral and response-dependent properties. *Cognition* 90 (23–N):32.

- Nietzsche, F. (1966). *Beyond good and evil*. Translated by W. Kaufmann. New York: Vintage. (Original work published 1886)
- Nisbett, R. E., and T. D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychol. Rev.* 84:231–259.
- Nussbaum, M. 1990. *Love's knowledge: Essays on philosophy and literature*. Oxford: Oxford University Press.
- Nussbaum, M. C. 2001. *Upheavals of thought: The intelligence of the emotions*. Cambridge: Cambridge University Press.
- Pecher, D., R. Zeelenberg, and L. Barsalou. 2003. Verifying different-modality properties for concepts produced switching costs. *Psychol. Sci.* 14:119–124.
- Pecher, D., and R. Zwaan, eds. 2005. *Grounding cognition: The role of perception and action in memory, language and thinking*. Cambridge: Cambridge University Press.
- Pessoa, L. 2005. To what extent are emotional visual stimuli processed without attention and awareness? *Curr. Opin. Neurobiol.* 15:188–196.
- Polanyi, M. 1967. *The tacit dimension*. Garden City, NY: Doubleday.
- Prinz, J. 2007. *The emotional construction of morals*. New York: Oxford University Press.
- Putnam, H. 1999. *The threefold cord: Mind, body, and world*. New York: Columbia University Press.
- Regier, T. 1996. *The human semantic potential: Spatial language and constrained connectionism*. Cambridge: MIT Press.
- Rizzolatti, G., L. Fogassi, and V. Gallese. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2:661–670.
- Roberts, B., N. Kuncel, R. Shiner, A. Caspi, and L. Goldberg. 2007. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspect. Psychol. Sci.* 2:313–345.
- Roetz, H. 1993. *Confucian ethics of the Axial Age*. Albany: State University of New York Press.
- Rosch, E. 1973. Natural categories. *Cognit. Psychol.* 4:328–350.
- Rosch, E., C. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. 1976. Basic objects in natural categories. *Cognit. Psychol.* 8:382–439.
- Rozin, P., and C. Nemeroff. 1990. The laws of sympathetic magic. In *Cultural psychology: The Chicago Symposia on Human Development*, ed. R. S. G. H. James Stigler. Cambridge: Cambridge University Press, 205–232.
- Ryle, G. 1949. *The concept of mind*. London: Hutchinson.
- Schneewind, J. 1990. The misfortunes of virtue. *Ethics* 101:42–63.

- Searle, J. 1995. *The construction of social reality*. New York: Free Press.
- Setton, M. 2000. Ambiguity in the Analects. *J. Chin. Philos.* 27:545–569.
- Shepard, R., and L. Cooper. 1982. *Mental images and their transformations*. Cambridge: MIT Press.
- Simmons, K. W., and L. Barsalou. 2003. The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cogn. Neuropsychol.* 20:451–486.
- Simmons, W. K., D. Pecher, S. Hamann, R. Zeelenberg, and L. Barsalou. 2003. *fMRI evidence for modality-specific processing of conceptual knowledge on six modalities*. Paper presented at the Meeting for the Society of Cognitive Neuroscience, New York.
- Sinnott-Armstrong, W., ed. 2008a. *Moral psychology: The cognitive science of morality: Intuition and diversity*. vol. 2. Cambridge: MIT Press.
- Sinnott-Armstrong, W., ed. 2008b. *Moral psychology: The neuroscience of morality: Emotion, brain disorders, and development*. vol. 3. Cambridge: MIT Press.
- Slingerland, E. 2001. Virtue ethics, the Analects, and the problem of commensurability. *J. Relig. Ethics* 29:97–125.
- Slingerland, E. 2003a. *Confucius: Analects: With Selections from traditional commentaries*. Indianapolis: Hackett.
- Slingerland, E. 2003b. *Effortless action: Wu-wei as conceptual metaphor and spiritual ideal in early China*. New York: Oxford University Press.
- Slingerland, E. 2008. *What science offers the humanities: Integrating body and culture*. New York: Cambridge University Press.
- Solomon, R., ed. 2003. *What is an emotion? Classic and contemporary readings*. New York: Oxford University Press.
- Solomon, R., ed. 2004. *Thinking about feeling: Contemporary philosophers on emotion*. New York: Oxford University Press.
- Spivey, M., and J. Geng. 2001. Oculomotor mechanisms activated by imagery and memory: Eye movements to absent images. *Psychol. Res.* 65:235–241.
- Spivey, M., D. Richardson, and M. Gozalez-Marquez. 2005. On the perceptual-motor and image-schematic infrastructure of language. In *Grounding cognition: The role of perception and action in memory, language and thinking*, ed. D. Pecher and R. Zwaan. Cambridge: Cambridge University Press, 235–241.
- Sweetser, E. 1987. *The definition of lie*. In *Cultural models in language and thought*, ed. D. N. Q. Holland. Cambridge: Cambridge University Press, 43–66.
- Swindale, N. 2001. Cortical cartography: What's in a map? *Curr. Biol.* 11:R764–R767.
- Talmy, L. 2000. *Towards a cognitive semantics*. Cambridge: MIT Press.

- Taylor, C. 1989. *Sources of the self: The makings of modern identity*. Cambridge: Harvard University Press.
- Van Norden, B. 2007. *Virtue ethics and consequentialism in early Chinese philosophy*. New York: Cambridge University Press.
- Warrington, E., and T. Shallice. 1984. Category-specific impairments. *Brain* 107:829–854.
- Wexler, M., S. Kosslyn, and A. Berthoz. 1998. Motor processes in mental rotation. *Cognition* 68:77–94.
- Wheatley, T., and J. Haidt. 2005. Hypnotic disgust makes moral judgments more severe. *Psychol. Sci.* 16:780–784.
- Wilson, S. 1995. Conformity, individuality, and the nature of virtue: A classical Confucian contribution to contemporary ethical reflection. *J. Relig. Ethics* 23:263–289.
- Wilson, T. 2002. *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge: Harvard University Press.
- Wilson, T., D. Kraft, and D. Dunn. 1989. The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *J. Exp. Soc. Psychol.* 25:379–400.
- Wilson, T., and R. Nisbett. 1978. The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Soc. Psychol.* 41:118–131.
- Wilson, T., and J. Schooler. 1991. Thinking too much: Introspection can reduce the quality of preferences and decisions. *J. Pers. Soc. Psychol.* 60:181–192.
- Winawer, J., N. Witthoft, A. Huk, and L. Boroditsky. 2005. Common mechanisms for processing of perceived, inferred, and imagined visual motion. *J. Vis.* 5:491.
- Zajonc, R. 1980. Feeling and thinking: Preferences need no inferences. *Am. Psychol.* 35:151–175.
- Zimbardo, P., S. Laberge, and L. Butler. 1993. Psychophysiological consequences of unexplained arousal: A posthypnotic suggestion paradigm. *J. Abnorm. Psychol.* 102:466–473.