The Affective Dog and Its Rational Tale: Intuition and Attunement*

Peter Railton

Intuition—spontaneous, nondeliberative assessment—has long been indispensable in theoretical and practical philosophy alike. Recent research by psychologists and experimental philosophers has challenged our understanding of the nature and authority of moral intuitions by tracing them to “fast,” “automatic,” “button-pushing” responses of the affective system. This view of the affective system contrasts with a growing body of research in affective neuroscience which suggests that it is instead a flexible learning system that generates and updates a multidimensional evaluative landscape to guide decision and action. With this latter view in mind, I revisit some of the classic hypothetical scenarios used in experimental moral psychology.

Glendower: I can call spirits from the vasty deep.
Hotspur: Why, so can I, or so can any man; But will they come when you do call for them? (Shakespeare, Henry IV, Part 1)

* Although I pattern my title on Jonathan Haidt's seminal paper, “The Emotional Dog and Its Rational Tail,” Psychological Review 108 (2001): 814–34, I do not intend this essay to be a response to that article in particular. I will draw upon a number of features of Haidt’s arguments and evidence along the way, but I should emphasize at the outset that Haidt’s view of the nature and status of intuitive moral responses is complex and has evolved over the last decade. Indeed, it seems fair to say that thinking among psychologists and experimental philosophers on the question of intuition is in continuous evolution; this article attempts to frame a general discussion, which by its nature cannot be true to the full range and complexity of the literature. For this, I apologize in advance.

I have benefited from discussion of the issues touched on in this article with many more individuals and audiences than I can thank here. I owe a special debt to my fellow discussants in the Ethics Georgetown workshop on experimental philosophy, some of whose contributions appear in this symposium. Others to whom I am especially obliged include Robert Audi, Roy Baumeister, Kent Berridge, Sarah Buss, Charles Butter, Fiery Cushman, Justin D’Arms, Stephen Darwall, Julien Deonna, John Doris, Phoebe Ellsworth, Heidi Feldman, Tamar Szabo Gendler, Allan Gibbard, Patricia Greenspan, Gilbert Harman,

Ethics 124 (July 2014): 813–859
© 2014 by The University of Chicago. All rights reserved. 0014-1704/2014/12404-0007$10.00
I. INTRODUCTION

Many a philosopher has called intuitions from the depths of Plato’s cave or the human psyche, but those that have come have been less impressive than one might have hoped. They have been unreliable (consider early modern intuitions of the impossibility of a vacuum or of the Euclidean nature of physical space), quarrelsome (Classical and Intuitionist mathematicians each find their own view of appropriate standards of proof self-evident), and fickle (recent work in psychology and experimental philosophy has shown that “intuitive” moral assessments can vary disquietingly with framing effects, such as seemingly irrelevant details of the wording or order of queries).

Still, it is difficult to see how we can manage to avoid appealing to intuition—in thought generally, and in normative or evaluative thought especially. Nearly everyone will agree that debilitating pain is, other things equal, a bad thing, to be avoided or alleviated. Yet what is this if not an intuition? Certainly it seems more obvious and compelling than any argument we might give for it. It cannot be analytic, since an attribution of badness, unlike painfulness, is thought to entail the existence of a prima facie reason to act. Closer to the mark would be saying that we experience the badness of debilitating pain firsthand whenever we’re having a real bout of it. But spelling this out involves the problematic notion of “evaluative perception.” Suppose we say that the badness of pain is simply self-evident. Yet isn’t this just another way of saying that it is known in an immediate, nondeliberative, “you-know-it-when-you-see-it” way?—that is, intuitively. Generalizing, the same can be said of “self-evident” principles of reasoning such as modus ponens. On pain of circularity, they cannot be known by reasoning. Neither can they be generalizations from experience. Nor can they be mere tautologies, for they are substantive norms for guiding thought. Indeed, the great philosophers of the past bequeathed us the notion of “intuition” from awareness that deliberative, concept-based knowledge presupposes nondeliberative, nonconceptual ways of knowing.

It seems, then, that we won’t easily be rid of the idea of intuition or intuitions. So we face the question: can we, suitably chastened by the misadventures of the past, still make a case for attributing some prima facie authority to intuition or intuitions?

II. EVERYDAY INTUITION AND INTUITIONS

To answer, we’ll need some way of identifying intuition or intuitions, at least as they interest us here. The term ‘intuition’ has no proprietary sense, and while intuition has been conceived by some philosophers as a distinctive, quasi-perceptual rational faculty, a source of synthetic a priori truths, the everyday mental phenomena we typically call ‘intuition’ or ‘intuitions’ are quite diverse in content, often a posteriori, and give little evidence of their origins.

Let’s start as untendentiously as we can, with what is manifest in everyday life, namely, that we often find ourselves with a spontaneous “sense” that (i) some thought, action, state of affairs, individual, or object is right or wrong, good or bad, appropriate or inappropriate, familiar or alien, credible or dubious, promising or hopeless, reasonable or excessive, true or false, and so on, which (ii) does not require explicit, effortful reasoning or judgment, and (iii) can arise nonvoluntarily and unbidden. Moreover, it (iv) can persist in the face of contrary conscious judgment, while (v) still remaining in some degree compelling or motivating and thus such that (vi) we are reluctant to give it up or ignore it, even though (vii) we cannot articulate a satisfactory explanation or justification for it.

We will say that features (i)–(vii)—plus an eighth, to be added shortly—characterize intuition or intuitions in the *observational sense*. In this sense, there need not be a distinctive or unified faculty behind intuitions, nor need they actually possess the authority they are felt to have. While it is customary among philosophers to speak of intuitive judgments, intuitions often—perhaps paradigmatically—appear in the first instance as a “sense” or “feeling.” This “sense” can then shape judgment, though it is a familiar feature of everyday intuitions that they can be substantially independent of what we would explicitly endorse or affirm. A functionalist in the philosophy of mind, for example, may retain an unaccountable intuitive sense that a box full of gears and wheels perfectly replicating the functional patterns of the human brain would simply lack consciousness; and an act utilitarian may still be unable to shake an intuitive sense that pushing a bystander off a footbridge to stop the runaway trolley would be wrong. When, below, we speak unqualifiedly of “intuition” and “intuitions,” it will typically be in this everyday, observational sense.

The final feature is that “intuition” (viii) is thought to guide spontaneous action, without need for conscious deliberation, decision, planning, or intention formation. A talented saxophonist can barrel into a riveting improvised solo that shifts key and rhythm fluidly to make a sat-
isfying expressive whole, without making decisions along the way. A wit can find a funny and all-too-true remark escaping her lips even before she has thought about what to say—sometimes to her regret. In such cases, individuals naturally describe themselves as acting “spontaneously” or “intuitively,” “without thinking about it.”

The felt compellingsness and spontaneous action-guidingness of intuition and intuitions would be nothing more than quaint, sometimes inconvenient, features of our psyches were it not the case that people—even hard-core analytic philosophers—so often are willing to attribute some degree of normative authority to the intuitive. What might account for this?

III. IMPLICIT COMPETENCE—A MODEL

The simplest explanation would be lack of alternatives—often we must make up our minds or act and have neither a well-considered judgment to fall back upon nor the time and information to form such a judgment. But my impression is that there is more to the story: intuition and intuitions generally have a good deal more street credibility than a “better than nothing” explanation would suggest. Often they play a pivotal role even in very thoroughly considered and well-informed judgments.

One explanation would be that people think that intuition and intuitions are, at least sometimes, manifestations of underlying competencies and implicit knowledge that cannot readily be brought to mind or articulated. After all, in those with little talent, experience, insight, style, or wit, “following intuition” seldom leads to a thrilling musical improvisation or a lightning-fast bon mot that beautifully melts a frigid social occasion into convivial laughter. Reliably good intuition in the observational sense seems to require good, reliable infrastructure.

Indeed, the popularity of the term ‘intuition’ in contemporary philosophy is partly traceable to the influence of a linguistic model, in which native speakers’ open-ended ability to produce and understand well-formed sentences spontaneously—without consulting the relevant rules—is attributed to the operation of an underlying linguistic competence. Linguists have argued that this open-ended ability cannot be accounted for by positing a trained repertoire of verbal behaviors—linguistic competence must be generative, a tacit mastery of rules or principles that can extend to indefinitely many novel cases.

The term of art for fluent speakers’ spontaneous assessments of grammaticality is ‘linguistic intuitions’, which linguists use to map the structure of rules and transformations in the imputed “depth grammar.” Linguistic intuitions behave very much like everyday intuitions of our observational sense. For fluent speakers, certain sentences will “sound wrong” or “feel right” nondeliberatively and nonvoluntarily, and it will
feel like a mistake to ignore this sense of rightness or wrongness even though they can’t formulate a grammatical rule or principle to explain it. Let us call the generative linguists’ picture of grammatical intuitions as surface manifestations of an underlying structure of rules and information a tacit-competency-based model of intuition.

In recent years, implicit competencies have been used to account for “inarticulable” or “intuitive” expertise or skill in a wide range of domains—music, medicine, business, sports, games of strategy, and so forth. Of course, in such domains there often is much more to expertise or skill than congruence with established usage, and organized bodies of information and projectible empirical generalizations play a larger role than rules. So what is borrowed from the linguistic model is not a template but a picture: open-ended capacities to respond fluently and adeptly to situational or expressive demands cannot be the result of a fixed repertoire of trained responses, so that some essentially generalizable capacities and information structures must be at work; and since an individual with these abilities typically can articulate only a fraction of these generalizable capacities or information structures, they must take the form of a tacit competence, which can be manifest spontaneously in thought and action.

IV. “NARROW” COMPETENCE IS OFTEN WIDE

The tacit-competency-based model of intuition and intuitions has largely been associated with domain-specific skills or expertise, such as linguistic competence, chess playing, or “baseball sense.” Yet we can easily overlook the breadth of competencies and knowledge involved in what might appear to be a specialized domain of human activity.

A tacit “grammar module” plus “dictionary” would hardly make me an effective participant in a linguistic community if I understood nothing of conversational norms, lacked any ability to “read” other people’s communicative intentions, had no insight into my own ideas and feelings, could not anticipate how my words would strike others, or couldn’t sense whether we were actually understanding one another. Here, then, is a complex set of general, generative skills and sensibilities—a psycho-socio-linguistic competence—drawing as much as grammatical competence upon powerful, possibly innate learning capacities, and requiring extensive “learning by doing,” even while remaining largely nondeliberative or “intuitive” in its operation.

V. AN EXAMPLE

A hypothetical example might help make clearer just how wide the implicit competencies or information structures involved in thinking or acting “intuitively” might be, even within a specialized domain, litigation.
Consider an attorney working pro bono defending an individual from a disadvantaged group against an accusation of murder. The trial is nearing a close, and in her professional judgment everything seems to be going her way. Jury selection went well, and she had deftly cast doubt on the slipshod forensic evidence. The prosecutor had bungled numerous times, and rulings from the bench have all been in her favor. But now, in these final days, a gnawing unease has grown within her. Leaving the courtroom on the next-to-last day of trial, having just succeeded at trapping the prosecution’s key witness—a petty criminal who has bargained his way to a lighter charge—in a blatant inconsistency, she feels defeated, not triumphant.

She keeps thinking about her client, who has been so sullen and uncommunicative throughout, and who only gave her the facts she needed to trap the witness as a result of series of tense, whispered exchanges during the cross-examination itself—over the course of which she had interrogated him almost as sharply as the witness. But she was new to pro bono work like this, and she tells herself that this sort of sullen, distrustful attitude is simply typical of the kinds of defendants one encounters in such cases. Who knows what they’ve been up to in the past that makes them uncomfortable in a court of law?

Still, she can’t talk herself out of the sense that something’s wrong. An image sticks in her mind. At the end of that day’s session, just as she was turning to thank her client for giving her the information she’d needed, she received an urgent text from her firm. A major case was breaking. “Sorry, I have to take this,” she said to her client, turning away. By the time she had turned back, he was already being led out of the courtroom in handcuffs. That evening, as she tries to work on her summary argument that night, she can’t keep focused and finds herself re-running the days’ events in her mind, always coming back to that image of her client, head bowed, looking away from her, as he was led out. “What’s with me?” she thinks, forcing herself to get back to work. “Well, at least tomorrow we’re done with the trial. I’ll be glad to have it over with.”

The next morning, as she approaches the courthouse, her legs feel like lead and her stomach is already tied in a knot. When her client is brought into the courtroom, she greets him and thanks him for yesterday, trying to be enthusiastic, but he just glances downward in silence, as he often does when she addresses him. But then he looks up and says with a slight smile she hadn’t seen before, “Yeah. We got it done.”

The session begins. Her summation is all prepared, and, still with no clue as to why she feels so uneasy, she resolves to stick to her tried-and-true formula: walk the jury through the evidence and the law step by step, making the logic of her case seem inescapable. “Win their minds and you will win their hearts,” she has always said.
But about a third of the way into the summation, hearing her own voice echoing back to her in the large courtroom and feeling a strange reluctance to look the jurors in the eyes as she paces back in forth in front of them, her unease becomes unbearable. She feels distant, disconnected, lost. She shoots a nervous glance at her client. Nothing—he’s staring down, expressionless. She tries to buy time, “But before I continue with the evidence, let me remind you of the details of the charge brought against the defendant. . . .” This feels wooden, preachy, hopeless. Her throat is going dry and she’s starting to flush deeply. It feels as if everyone must know that she’s struggling. Once she has repeated the charge, she breaks off again.

With no new idea, she tries to pull herself together, straightening her tensed body, taking a long breath, and walking slowly over to the jury box. As she does so, she can’t help taking another look over to her client, who is now staring at her, his brow deeply furrowed—with worry? anger? She turns abruptly and makes herself try to look the jurors in the eye. Focusing her energy as best she can, she resumes speaking, though no longer in her courtroom voice, “. . . But I know that you know the facts of this case backwards and forwards. So what is there left to say?” An awkward pause, her mind racing. More words come, “What’s left is to talk about what this case is really about—why, however heinous the crime, however much we might feel for the family of the victim, there will be no justice if we turn our anger on the wrong man. However uncertain you feel, there’s one person who knows for sure that my client is innocent—the murderer himself, who’s still out there, laughing at justice. My client just happened to be the right height, weight, . . . and color, and to have been near the scene of the crime. A convenient target for a prosecutor eager to get a conviction and able to produce a fake witness in return for copping a plea. But my client no more deserves life imprisonment than you or I. Throw the book at him and you’ll close the book on this case—and the real murderer will roam the streets as free as ever. We must keep the book open, and find the man who really deserves our anger, and our righteous justice.”

When she had stopped and approached the jury box, the jurors themselves had shifted in their chairs uneasily, looking away as she tried to catch their eyes. She was in unfamiliar territory, not seeing clearly where she was going, straining to keep her own concentration and composure. Having taken up the thread, she followed it, sentence by sentence, not stopping to think what to say. She moved her eyes from face to face as slowly as she dared—talking about the case as if talking to a friend late at night, her taught face and voice softening. One by one, the jurors stopped staring at their shoes or looking past her, and began to meet her eyes and fix on what she was saying. As she explained what moved her about this case, she felt the emotion rising in her throat and noticed that the eyes of
an older man the front row, who had sat steely faced throughout the trial, had started to tear up. Following where spontaneous thought and intense feeling had led her, she now senses that she must not drag this out. She must build to a conclusion.

But what is the conclusion?—she has no ringing phrases at the ready. Yet somehow, the words come. Barely controlling her voice, she says, “I have spoken to you from my heart. And I hope that I have reached your hearts. Because that is where you must search to find justice in this case. I know you will. Thank you. The defense rests.” The courtroom is dead silent. Her legs are shaking beneath her as she returns to her seat, but she feels that her work is done.

And she’s right. The jury votes to acquit. Pressed by a colleague afterward to say what led her to depart from her prepared summation and show her feelings, she draws a blank. “All I know is, at that moment I felt like I was dying out there. Talking to myself. Completely cut off. And I just couldn’t go on. I had to do something. Somehow, when I slowed down and concentrated on their faces, trying to speak to them, the words began to come. My head was pounding, but I just tried to make sure I never lost eye contact. It began to feel like, ‘Yes—keep going.’"

Later she learns indirectly that her trademark meticulous method had from the outset rubbed certain influential jurors the wrong way. She had come across as indifferent and remote—the words that got back to her were “cold,” “phony,” and “condescending”—while the prosecutor, legally outmaneuvered at every turn, nonetheless seemed to care strongly about the crime and its victim, and to treat the jurors as equals. They had come to distrust her and to resist believing what she was trying to establish. Her “demolition” of the prosecution witness felt to them cruel and overbearing. Sure, she could run circles around him, but what did that prove? She was indeed about to lose the case.

Later still, she realizes something more. She had in some degree been remote, even phony and condescending. Her contempt for the prosecutor, her smug pleasure at outmaneuvering him and the witnesses, and her evident sense that she, the high-powered attorney from a top firm, was the smartest person in the room—all came through to everyone present. Palpable to everyone, too, was the distance between her and her own client. Sure of herself from the start, social distance had made her wary of him. Sure of her plan of attack, she hadn’t listened carefully to what her client was trying to tell her all along. He had soon realized that he wasn’t being listened to and had become sullen and withdrawn. So she didn’t learn the crucial information that truly convinced her of his innocence until the next-to-last day. She had been unconvincing, she realized, because she herself was unconvinced. And he had seemed alienated in part because she was.
Working together that penultimate day for the first time, they had actually begun to trust one another at bit, despite her inquisitorial manner. Of course, she didn’t recognize this at the time—all she had felt then was a confused, nagging regret over her botched attempt to thank him the evening of the cross-examination. So her growing unease had a double source, and her formulaic, one-size-fits-all summation felt inauthentic to her, too. That afternoon she had, just a bit, broken through the barriers of class and race to connect to her client as an individual, and that lent a new urgency to everything else. The next day, delivering her routine summary felt wrong. And it was. No wonder she couldn’t go on.

VI. SPONTANEOUS VERSUS “AUTOMATIC” ACTION

Sometimes it is said that developing a skill or expertise is a matter of performing certain actions so frequently that they become “automatic,” no longer requiring the effortful, conscious attention or thought of a novice. The model here is inherited from animal learning, where repeated reinforcement is said to ingrain a motor response or habit, so that perception of the “conditioned” stimulus leads directly to behavior with minimal cognitive processing.

Although our trial lawyer did not act deliberatively in departing from her stock summation—indeed, she acted in violation of her explicit deliberative resolution—she certainly wasn’t acting “automatically” in this sense, or from habit. On the contrary, she had snapped out of her familiar role and routine, and her mind was working flat out, wholly focused and drawing upon all her resources, perceptual, intellectual, and emotional. As the trial had worn on, she had implicitly noticed certain signs—facial expressions, body language, and the like—that she was losing touch with jurors. These gave rise to the unease within her, undermining her confidence and sapping her motivation. Then, on the next-to-last day, the intense experience of working together with her client successfully under pressure had moved the two of them out of their accustomed frameworks of social distance. Her frustrated effort to thank him, only to see him led away, unacknowledged, put her in the position of sensing how things must look to him, and how he might have felt excluded and distrustful of the whole business. It was when it came home to her what an injustice would be done to him if the jury convicted, that her own sense of injustice was ignited. What might under ordinary circumstances have felt like nothing worse than flatness in courtroom—“doing my job” in a professional way—now felt like something she could not abide, a personal betrayal.

When she finally yielded to “intuition,” then, what spontaneously “felt right” was to step out of her distant, professional manner, to try to
reconnect with her client, and to reach across the gulf between her and
the jury, speaking to them as one human to another, letting her passion
and sense of injustice show. She had poured everything she had into it,
and become frank and eloquent in ways that surprised her, yet which were,
at bottom, expressive of her and her deepest values. However remote and
condescending her attitudes might have been, she was there, doing pro
bono work, because she does care about seeing justice done for the dis-
advantaged. One is never more present, alive to the world, and human—
ever less “automatic” in one’s thought and action, or inattentive to the
values at stake—than when one rises to the occasion in this way.

VII. GROUNDWORK FOR SPONTANEITY

Nonetheless, her turnaround that morning and eventual success should
puzzle us somewhat. With no clear insight into the source of her unease,
what could account for the unplanned yet effective coordination of all her
faculties in shaping how she eventually responded? No special-purpose
tacit competence or expertise could explain all that she went through that
morning or the day before, or enable her finally to find her own voice
when all appeared lost.

The groundwork for such spontaneous, coordinated response had
many layers. Nearest to the surface, conscious and unconscious elements
of her experience in the courtroom and with her client over the last few
days; somewhat deeper, her years of legal practice, honing skills of in-
terrogation, exposition, and persuasion; deeper still, her powerful mind,
“emotional intelligence,” and fundamental social and personal skills,
enlarged and refined over a lifetime. These abilities included a capacity
to read others’ faces or behavior, to feel something of what they feel, to
sense when she has someone’s trust, to overcome others’ defenses by
letting her own guard down, to keep focused despite high anxiety and
uncertainty, to persist in the face of frustration, and to stick up for what
she believes in. Hard at work, too, was a very high level of psycho-socio-
linguistic competence—having the words “come to her” that could ex-
press her heartfelt concerns while touching and, eventually, convincing,
her audience.

But pulling back somewhat, we can also see in her thoughts, feel-
ings, and behavior other, conflicting strains. Social preconceptions and
prejudices, reinforced by many years of living in a society divided by
class and race, inserted distance and distrust between her and her cli-
ent. Such implicit tendencies operated within her—as they operated
within her client, and indeed as they operate within all of us—even
against her own judgment and ideals, coloring her thoughts and actions
in ways that cannot be detected introspectively or controlled at will.
These implicit attitudes set the stage for their initial, failed interactions
and might have persisted unchallenged throughout the trial had not the critical cross-examination driven them to work together and broken down the set-piece character of their relationship—giving her a sense of distress about failing to thank him and bringing to his mind the ‘we’ in “We got it done.” This very incomplete summary of all the capacities and attitudes at work over the course of the trial suggests why undergoing this turnaround and salvaging the case by “acting intuitively” took everything she had.

The undramatic intuitions of everyday life—what strikes one as the right gift for a friend, or sounds like a bad idea, or smells fishy as an excuse, even though one is hard-pressed to say exactly why—also typically draw upon wider personal resources as well as more specialized competencies. Given the volume of sensory information we take in at each moment, the varied needs and goals we seek to meet, and the complex personal relationships and social environments we inhabit, the conscious, deliberate mind could hardly manage everything on its own. It must depend continuously upon the experienced discernment of nondeliberative perceptual, cognitive, and emotional capacities as we find our way through life.

Consider even so simple a question as when it makes sense to stop what one is doing to reconsider whether to continue, or to make a break and change course. Sometimes, as in the case of our trial lawyer in the middle of her summary, this is exactly what is needed; other times it is self-defeating or worse. We cannot give the main part of the task of monitoring whether to stop and rethink to deliberation itself—we’d forever be deliberating about whether to stop to deliberate about whether to stop to deliberate . . . . A brute disposition to stay on course would spare us such unproductive spinning of the wheels of thought, but it would also have led our lawyer to lose her case and to the conviction of an innocent man. What’s needed is a context- and evidence-sensitive “sense” that does not itself require active supervision—a tacit susceptibility to information about whether things are going well or ill for our current course of action, which can in turn weaken or strengthen our confidence in proceeding as we are. “Unsupervised” evidence-sensitive monitoring is not an impossible job description. For example, our tacit linguistic competence can, without benefit of supervision, generate a “sense” of grammatical anomaly that gets us to stop in mid-sentence to correct verb agreement even when we’re in the heat of an intense discussion.

VIII. CLASSICAL INTUITION AND INTUITIONS

We have now come, by a circuitous route, into the vicinity of the notion of intuition bequeathed to us by Aristotle and Kant, which we will call intuition in the classical sense. Classical intuition enters our mental lives
where something like discernment or judgment is needed, but deliberative thought won’t do.

In the *Posterior Analytics*, Aristotle points out that demonstrative science (*dianoia*) cannot stand on its own. In the first place, demonstration needs premises which, on pain of regress, demonstration cannot supply. Second, if science is to yield knowledge of things in this world, its abstract, universal concepts and principles must somehow be connected to concrete, particular events, objects, or experiences. Consider perception. Aristotle likens the problem of introducing universals into the flux of sensation to the problem of an army in rout, which can take a stand and avert disaster, but only if a defensive line can be formed even in the absence of a ruling authority, that is, *spontaneously*. The soldiers might all have been taught general rules of battlefield tactics and therefore know that only by forming a defensive line can they stop a rout, but if any actual individual soldiers are to “take a stand” and nucleate the formation of such a line, a sufficient number must come to “sense” that a particular time and place are right. Similarly, for a universal to “take a stand” in the onrush of unorganized sensation, Aristotle writes, the cognizer must have a spontaneous “discriminative faculty,” “intuition” (*nous*; also: “understanding”) that responds selectively to sensory sequences to introduce into the flux a “rudimentary universal,” so that organized cognition can begin to take hold.¹ In calling this discriminative faculty “intuition,” Aristotle is stressing that it must operate in a spontaneous, “immediate,” or nondemonstrative manner. Such a discerning sensibility constitutes a kind of understanding even though it cannot be spelled out as a set of principles or rules—just as combat-tested Athenian soldiers would have an intuitive “sense” or implicit “understanding,” impossible to articulate fully, of when the dynamics of a group of soldiers in retreat can be reversed if a few of them will turn to face the enemy. Fortunately, in the case of regrouping conceptually in the face of the onrush of sensation, “the soul is so constituted as to be capable of this process.”²

Kant put the problem in terms familiar from contemporary debates. Concepts, according to Kant, can be understood as general rules,³ so that applying a concept to an individual sensation or action involves determining whether the relevant rule applies. Can concept-governed reasoning do this? Only if we can determine whether those concepts apply, but “judgment cannot always be given yet another rule by which

---

2. Ibid., 100a12. A full account of the notion of intuition or understanding in Aristotle, and its roles in knowledge, lies well beyond the ambitions of this article. And, of course, the same is true of what I will say of intuition in Kant, below.
to direct its subsumption (for this would go on to infinity).” What is needed, on Kant’s view, is a nondeliberative way for knowledge to operate—a “sensibility” that is not regulated by concepts but which nonetheless can introduce form into experience by applying the categories of space and time, thereby giving deliberative thought a foothold. This sensibility he called “intuition” (Anschauung—a term whose root, like that of intuition, lies in “looking”). Even if I were innately equipped with the abstract, universal categories of space and time, I still would have no grasp of how they apply in my own experience and reasoning unless I had a capacity to sense that I exist at a particular location in space and time. Such a “direct apprehension” or a “singular representation” of being here now—a de se representation of the world and my location in it—is just the kind of particularistic, self-locating content universals cannot supply on their own. “Intuitions without concepts are blind,” he wrote, but “thoughts without contents are empty.”

From the classical perspective, then, intuition is an integral, indispensable part of our capacity to respond rationally to the world, even though it must perforce not be the work of a faculty of declarative reasoning. As Kant put it, “Neither of these qualities or faculties is preferable to the other.” Intuition is at work whenever experience brings knowledge into play, and whenever knowledge in turn is applied in a particular context or action. Thus, for example, intuition is at work whenever tacit, rule- or concept-based competencies guide our thought and behavior—including our competencies with declarative reasoning itself.

Classically, then, intuition as a faculty mediates both the “pathway” from particularistic sensation to concept- or rule-based thought and the “pathway” from concept- or rule-based thought to particular actions. The classical view of intuition complements the tacit competency view: the ability to apply general concepts or rules in experience and action is part and parcel of acquiring generalizable informational structures or principles and bringing them to bear in actual thought and practice. Combining these two views of intuition we arrive at a picture similar to figure 1, in which intuition mediates two pathways between sensation and action. In one, shown by the solid line, there is a nondeliberative path from perception to action. This is the path of “intuitive” action, for example, in the spontaneous exercise of skill or fluent linguistic competence. In the other, shown by the dashed line, the path from sensation to action passes through discursive, concept-based reasoning. This is the path of “deliberative” action, for example, in thinking one’s way to a

6. Ibid., A51.
novel practical solution or mentally conjugating one’s verbs using rules one was taught last week in introductory Portuguese.

On both views of intuition, moreover, there is no assumption that we are aware of all of the operations of intuition, or that the operation of intuition would be fully transparent to us upon introspection. Sometimes, intuitive responses will be experienced as a “sense”—for example, our everyday “sense” of surrounding physical space or the passage of time, or a battle-tested warrior’s “sense” that now is the moment to turn to face the enemy. Sometimes, too, this “sense” may be accompanied by thought—“That was awhile ago” or “Now’s the moment!” Such a “sense” or thought typically has the features of intuition or intuitions in the observable sense. For example, it tends to arise effortlessly or spontaneously (ii), and to strike us as correct (i); it possesses some authority or generates some pressure toward action (v) that we are reluctant to ignore (vi), even if we cannot explain its basis or origin (vii); moreover, it tends to arise non-voluntarily (iii) and can be resistant to reasoning and judgment (iv); and it will tend to shape the way we interpret a situation and to guide attention, recollection, thought, feeling, and action nondeliberatively (viii).

IX. PSYCHOLOGY

Thus far we have described the observational, tacit-competency-based, and classical senses of intuition and how they might be related. But how well does this picture comport with current empirical research on the nature of mind and brain? If we consult the diagrams of mental processes or capacities found in contemporary psychology textbooks, we will find perception, attention, association, memory, cognition, emotion, motivation, and so on, but not intuition. Is figure 1 therefore a figment of the philosophical or commonsense imagination?
What if we were instead to ask a psychologist whether anything answers to the “job description” given by features (i)–(viii)? Then there would be no difficulty answering affirmatively, for (i)–(viii) describe functional characteristics of what is today called the affective system—understood broadly to incorporate reward, emotion, mood, affectively charged or valenced memory, as well as other elements of what was once called the “limbic” system and the regions of the brain that mediate connections between affective regions and the higher cortical regions that house explicit, declarative representations and reasoning. The main components of the broad affective system, as I will call it, have a long history in our mammalian ancestors and interact heavily with other brain regions in both aroused and default emotional states. Thus understood, affect appears to play a continuously active role in virtually all core psychological processes: perception, attention, association, memory, cognition, and motivation.

Recognition of the pervasive role of the broad affective system has in recent decades come to be called the “affective revolution” in psychology, comparable in scope to the “cognitive revolution” that displaced behaviorism in the 1960s and 1970s. Two hypotheses have been especially influential.

First, the hypothesis of affective primacy, first introduced by Robert Zajonc, according to which a new sensory input engages the affective system almost immediately, before higher-order declarative thought. Drawing upon association, current state, and memory, the affective system appears to code and sort incoming information in ways that reflect whether the information is more or less trusted, better or worse than expected, more or less urgent, relevant to the self or others, and so on. In humans, this coding and sorting typically takes place below the level of conscious awareness, shaping the way in which the perceptual information will be experienced at the conscious level and priming various implicit or explicit responses in thought and action.

Second, and partly stemming from the idea of affective primacy, psychologists have increasingly accepted a dual-process model of the mind, according to which the broad affective system is thought of as the substrate for a distinctive, “automatic” mode of processing (“System 1”), while the higher cortical regions are the substrate for a more “controlled” and “effortful” mode of functioning (“System 2”). These two modes interact, of course, as when the affective system’s coding and sorting of perceptual information frames deliberative thought and decision making, or when

deliberative thought regulates an emotional response in a “top-down” manner (e.g., figuring out that one’s toe was trod upon by accident will tend to moderate one’s anger). Since the two modes tend to run in parallel, however, they can also produce conflicts or interference in thought and action.

The resulting picture, very crudely drawn, should look familiar—see figure 2. Should we say, then, that intuition and intuitions are alive and well in contemporary psychology, and living in the broad affective system?

X. MORAL PSYCHOLOGY

Let’s return to the case of moral judgment, where affective primacy and dual-process models have played a key role in recent experimental and theoretical work by psychologists and philosophers. Not everyone accepts both hypotheses, to be sure, but acceptance is widespread enough to enable Jonathan Haidt to describe a “new synthesis” in moral psychology, of whose theoretical grounding in the “affective revolution” he writes: “Social psychologists have increasingly embraced a version of the ‘affective primacy’ principle . . . [in light of] evidence that the human mind is composed of an ancient, automatic, and very fast affective system and a phylogenetically newer, slower, and motivationally weaker cognitive system. . . . [The] basic point was that brains are always and automatically evaluating everything they perceive, and that higher-level thinking is preceded, permeated, and influenced by affective reactions (simple feelings of like and dislike) which push us gently (or not so gently) toward approach or avoidance.”

The natural suggestion in this “new synthesis”

is that what we have known as “moral intuition” or “moral intuitions” are to be attributed to the “automatic, and very fast” workings of the “ancient” affective System 1, while more self-conscious, principled “moral reasoning”—for example, applying the Golden Rule or impartially assessing costs and benefits—is the work of the “newer, slower, and motivationally weaker” cognitive System 2. Because System 1 is held to have “little understanding of logic and statistics,” it is thought to operate by means of “heuristics,” “frames,” “push buttons,” “biases,” “stereotypes,” or “blunt” emotions—mental simplifications, short-cuts, or “point-and-shoot” responses.

System 1’s “heuristics” or “pre-sets” are seen as the result of evolutionary design, selected because they are generally reliable guides in the typical range of circumstances Homo sapiens has encountered, though they can mislead or generate seeming incoherencies or other departures from decision-theoretic rationality—especially when stretched in ways human evolution did not anticipate. Fortunately, the effortful cognitive system can intervene and take charge, proceeding by controlled, stepwise, consultative reasoning to correct failings or “myopia” in the affective system. In the normal course of life, however, it is the broad affective system that constitutes our first response and does the bulk of the work getting us around in the physical and social world. Since individuals typically lack insight into the automatic workings of System 1’s processes, they often resort to post hoc rationalization if required to explain their behavior, a phenomenon known as “confabulation.”

Contemporary psychologists and philosophers have drawn upon affective primacy and dual processing to explain the peculiar blend we observe in human life of largely sensible spontaneous thought and action combined with some striking failures, which tend to follow certain patterns, reflecting the operation of “heuristics and biases” of the affective system. Such heuristics include the “availability heuristic,” in which information that is more recent, salient, familiar, or easier to process is automatically given greater weight; the “representativeness heuristic,” in which features thought to be “typical” are automatically treated as if they were highly probable, so that stereotypes often preempt evidence; and the “affect heuristic,” in which individuals take their positive or negative

emotional responses as proxies for what is credible or dubious, good or bad, right or wrong, true or false. In the case of moral judgments, psychologists have invoked this same interplay of rapid, automatic, affective processing versus slower, more effortful, cognitive processing to explain otherwise puzzling patterns in the “intuitive” verdicts people give in such well-known scenarios as trolley problems. For example, Joshua Greene and colleagues used functional magnetic-resonance imaging to observe metabolic activity in the brains of subjects making judgments in Footbridge versus Switch versions of the trolley problem and similar moral dilemmas. When the resulting observations were aggregated, it emerged that in Footbridge and other dilemmas involving “personal” infliction of harm (e.g., pushing a heavy person off a pedestrian bridge to entangle a runaway trolley, causing his death but sparing five transport workers further down the track), higher-than-control levels of activity were seen in brain areas associated with affect and lower-than-control levels in areas associated with cognitive processing, while in Switch and other dilemmas involving more “impersonal” infliction of harm (e.g., pulling a lever to shunt the trolley down a side track, saving the five transport workers on the main track, but killing one who happened to be on the side track) they found the reverse. This evidence suggests that subjects contemplating “up close and personal” harms underwent a fast, strongly negative affective response that tended to dominate slower, more impartial reasoning that would favor minimizing the loss of life—making them more likely to give a negative verdict. By contrast, subjects contemplating relatively “impersonal” harms experienced no comparable fast, negative emotional response, so that their slower, more impartial declarative reasoning predominated—making them more likely to give a positive verdict favoring minimization of the loss of life. Buttressing this analysis were reaction-time data suggesting interference when judgment went against emotion: the minority who gave the “emotionally incongruent” positive verdict in Footbridge took significantly longer on average to give their verdicts than those who gave the “emotionally congruent”—but not loss-minimizing—negative judgment.

These results had a galvanizing effect on thinking about moral judgment. Perhaps philosophers were chasing a will o’ the wisp in looking for an underlying, presumably deontic, principle or...


tacit moral competency to explain the asymmetries in stock examples like Footbridge versus Switch. Greene and colleagues have since strengthened and refined their case in a series of experiments involving carefully designed variations of the original trolley scenarios. They found, for example, that the extent of intuitive negativity in Footbridge-like cases varied with the involvement of “personal force”—roughly, whether sending the man to his death on the track below would be the direct or indirect effect of an impetus originating in one’s own muscular contractions (e.g., if one pushes the man with a pole vs. pushing a switch to open a trap door).¹⁵ This hardly seems like a difference that should carry weight in moral evaluation.

In a joint paper with Haidt and in his subsequent book, Greene suggested that the origin of rapid, negative affective response in cases of “personal force” might be grounded in an evolved, automatic emotional alarm response to a ME HURT YOU schema—a “violatio[n] a chimpanzee can appreciate”—or a “myopic module” that is triggered by contemplating the use of personal force to harm another. Once the alarm is triggered, mental evaluation of the action is cut short, and so exhibits “blindness to harmful side effects.” It simply “can’t see” the fact that failure to push allows the five workers down the track to be crushed, while pushing would save four lives. Hence the failure of “intuition” to weigh costs and benefits in the way System 2 would have done, had it come into operation.¹⁶

In a similar way, various theorists have proposed that the human affective system has an evolved, “automatic” disgust response to acts of kinds inimical to increasing “inclusive fitness,” such as eating contaminated food, close contact with others’ body fluids, and incestuous or homosexual sexual behavior. For example, psychologists Debra Lieberman and Adam Smith found that the strength of self-reported disgust at the thought of incestuous activity varied in proportion to the degree of biological relatedness of the imagined partners.¹⁷ Akin to the ME HURT YOU schema or “myopic module,” this “automatic” disgust response is thought to be relatively insensitive to less direct countervailing infor-


Hearing the following vignette, Haidt proposes, triggers just such a quick, outcome-insensitive “flash” of disgust:

**Julie and Mark.** Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to make love?18

What Haidt found was not simply that experimental subjects think Julie and Mark’s lovemaking “not OK,” but that, when pressed to explain their judgment post hoc, the justifications they cited were harm-related features of incest in general, such as genetic abnormality or psychological trauma, even though these negative consequences were plainly excluded by the details of the vignette. When this discrepancy was pointed out to them, however, many subjects apparently did not withdraw their original judgment but insisted upon it despite being unable to offer an alternative rationale, a phenomenon Haidt and colleagues labeled “moral dumbfounding.”19 Haidt explained this phenomenon by the introspective inaccessibility of the “automatic” disgust reaction, and its resistance to ancillary information.20 Mark and Julie’s act simply “felt” wrong to subjects in a way that seemed like a mistake to ignore, whether or not they could supply a justification.

In what has become a very large literature, researchers in moral psychology and experimental philosophy have argued that a better understanding of the distal (evolutionary) and proximal (affective) origin of “intuitive” moral judgments might show them to be something other than manifestations of underlying moral competencies or principles. Greene, among others, has drawn the normative conclusion that “moral intuitions” might therefore deserve less deference, and greater critical scrutiny, than they characteristically receive in philosophical and commonsense moral thought.21 So: while intuition and intuitions are indeed

alive in contemporary psychology, a significant body of research suggests that they might not be as well as one might have hoped.

Traditional moral philosophers have tended to object to this method of inferring normative conclusions from psychological research. It is, after all, the “genetic fallacy” to claim that a dubious historical or causal origin precludes normative legitimacy. The fact that an intuitive judgment issues in the ordinary psyche from the affective system does not tell us whether it is true or false, or whether a principled justification can be found for it. That said, however, our intuitions do not bear upon their sleeves the seal of validity, so our reliance upon them should take place within an overall method of wide rather than narrow reflective equilibrium—we cannot afford to ignore whatever evidence can be found about why we might have, or find plausible, certain intuitions rather than others.

XI. AFFECTIVE ATTUNEMENT

It is precisely in this catholic evidence-gathering spirit that I would like to raise a different sort of question about the “dual-process” picture of moral judgment as sketched thus far. I won’t be questioning the quadripartite model of the mind in figure 1, which seems to me to have impressive philosophical and psychological backing. Neither will I be questioning affective primacy or the role of the broad affective system as represented in figure 2. Rather, I’ll be questioning the picture painted in Section X above of the nature and operation of the affective system or “System 1,” according to which it is an “automatic,” “relatively inflexible,” “point-and-shoot” system of “simple likes and dislikes,” “flashes” of emotion triggered by simple schemas, or “gut feelings” that cannot handle statistics and are incapable of “deriving conclusions about individual cases from properties of categories and ensembles,” and therefore congenitally lacks System 2’s capacity for balanced, forward-looking cost-benefit decision making.

A review of the last decade’s work in affective and behavioral neuroscience will, I believe, yield a very different portrait of nature and function of the broad affective system. Increasingly, it has come to be seen as a flexible, experience-based information-processing system quite capable of tracking statistical dependencies and of guiding behavioral selection via the balancing of costs, benefits, and risks. Indeed, this appears to be its core function—what it was “designed” to do as it evolved

over the course of our mammalian and primate past. This role for the affective system co-evolved with its distinctive capacity to exercise a direct, systematic effect on virtually all aspects of mind and body. As Randy Nesse and Phoebe Ellsworth write: “Emotions are modes of functioning, shaped by natural selection, that coordinate physiological, cognitive, motivational, behavioral, and subjective responses in patterns that increase the ability to meet the adaptive challenges of situations that have recurred over evolutionary time.” The emotions, thus, are seen as responses to internal and external stimuli evolved to attune the organism systemically with respect to the challenges and opportunities it faces in meeting its needs and achieving its goals in changing physical and social circumstances. They serve, inter alia, to situate the organism, code its sensory input, inform its interpretation of its situation, prime its inferences, and shape its dispositions to act. Intriguingly, these are the very roles envisaged for classical and tacit-competency-based notions of intuition.

Research in affective neuroscience, rather than uncovering a system confined to simple heuristics, blunt or “holistic” emotions, or “pre-sets” and “push buttons,” have found that the broad affective system encompasses a complex set of interactive subsystems, containing groups of neurons whose firing rates and patterns of interaction covary systematically with absolute versus relative gains or losses along a range of biological and social dimensions, variance in gains and losses, probability of a gain or loss conditional upon a certain stimulus or action, occurrence of better or worse than expected outcomes, and consequent expected values of various actions with reference to the individual’s current and anticipated needs or goals. The firing rates and interaction patterns in these subsystems are updated through experience via “discrepancy-reduction” learning processes that continuously generate expectations, compare these expectations with actual outcomes, and use this informa-

24. Such a primary function is quite compatible with capacities for other, more “stimulus-bound” or “stereotyped” functions. For example, I am told that seeing a spider or a snake excites immediate negative affective responses even when it is known that the spider or snake is harmless or no more than an image. And animal behavior that is “overtrained” owing to lack of variation in the reinforcement regime can become “automatic” in the sense that it is reenacted in a mechanical, stereotyped way and can be relatively insensitive to devaluation of reward.


tion to produce a neural “teaching signal” that guides forward revision of expectations. Such discrepancy-reduction learning tends with experience to attune expectations to actual outcomes—their frequency, magnitude, and variability—in a manner that approximates Bayesian updating. Moreover, these continuously updated expected-value functions predict behavior in choice tasks.27

Much of this work is quite recent, and there are a number of important differences among researchers over the contribution of particular regions and neurotransmitters in the brain. Further, the most fine-grained research has been done using single-neuron recording for rodents and primates rather than humans. However, the same underlying neural structures are found in more developed forms in humans, and coarser-grained noninvasive brain-imaging studies of humans engaged in learning and choice tasks have found patterns of metabolic activity consistent with the fine-grained patterns of neural activity found in animals engaged in similar tasks. It emerges, then, that the overall picture of the broad affective system in animals and humans is remarkably congruent with our philosophical understanding of the operation of rational procedures for learning and decision making.28 Increasingly, the affective system has come to be viewed, not in contrast to representational, cognitive, evaluative, and decisional capacities, or as disruptive


of forward-looking cost-benefit calculation, but as an integral part of them.29

The “affective revolution” thus did not simply revolutionize our idea of the role of the affective system in our psyche, it revolutionized our idea of the affective system itself. The “ancient” affective system we inherit from our animal ancestors appears to have been shaped by evolution for efficient and effective collection and utilization of information to generate an expectation-based evaluative landscape that implicitly guides thought and action as the individual navigates its way through its physical and social environment.30 This, presumably, gives us an explanation of how foraging mammals are able to learn through experience nearly optimal patterns of behavior in gathering and exploring for food, and in selecting mates and cooperation partners.31

In explaining the “myopic” responses of the affective system we inherit from animal ancestors, Greene writes: “For nearly all animals in nearly all contexts, it makes sense to consume calorically rich foods as soon as they’re available.”32 Yet foraging animals can pursue much more complex strategies than this, for example, regularly forgoing consumption even when hungry in order to cache food—in some species, keeping track of the specific location and changing condition of cached food—so that it can be accessed at times of greatest caloric need.33 Efficient foraging requires repeated decisions about whether to leave a given food patch before it is exhausted in order to explore for potentially richer or less risky patches or to achieve better nutritive balance. It is noteworthy that foraging birds and rats can manage to make such trade-offs of actual versus potential costs and benefits in highly efficient ways and execute corresponding sequenced, “plan-like” behavior—without, one presumes, conscious deliberation and planning.34 That this is possible suggests how it might be that humans are able to act intuitively—“without stopping to


32. Greene, Moral Tribes, 138.


think,” as our lawyer might have put it—in intelligent, organized, “plan like” ways that extend over time and are aptly responsive to a changing array of costs and benefits. Indeed, humans placed in experimental environments that simulate foraging tasks have shown a comparable capacity to learn implicitly from experience to make apt trade-offs of risks, costs, and benefits in the acquisition of monetary, informational, or social value.\(^{35}\)

What might guidance by an “evaluative landscape” look like when it is not carried out through conscious deliberation and planning? Historically, rats in mazes were treated by experimental psychologists as “stimulus bound” creatures, driven by conditioned motor habits. If a rat had been rewarded for turning right at a given choice point, then coming upon this juncture in the maze as it trundled along would “automatically” trigger an acquired right-turn motor routine. Early on, however, experimentalists Karl Lashley and Edward Tolman observed that, given the chance or required to do so by altered circumstances, rats could quickly improvise novel but effective ways of getting to their destinations—taking shortcuts or executing motor patterns “spontaneously” for which they had never been reinforced.\(^{36}\) The idea began to take hold that rats develop not just a raft of local stimulus-response connections, but a “cognitive map” of their surroundings, which can then be flexibly deployed when novel possibilities present themselves.\(^{37}\) Much more recently, neuronal recording has provided evidence that, as an animal explores an environment, specialized place cells and grid cells in the hippocampus build up both relational and grid-like geometric representations.\(^{38}\) When a rat is trained in a maze it acquires such spatial representations, and, when it reaches a choice point, activation in these representations spreads ahead of the rat down the alternate paths, sweeping first one branch, then the other, in rapid succession. Learned information about the comparative magnitude and probability

---


\(^{37}\) Tolman, “Cognitive Maps.”

of reward integrates with these sweeps and leads to the rat’s choice to turn right or left. If the rat takes the “wrong” turn, and thus fails to hear the usual clicking sound of the food tray opening as it approaches the feeding station, neuronal activity quickly spreads backward to the choice point and down the path not taken, before the rat itself reverses direction. Quite literally, then, rats appear to build up and follow an evaluative landscape by learning the lay of the land, representing it in centered and coordinate neural arrays, acquiring reward- and risk-related expectations for places and paths in these arrays, and acting via the comparison of these expectations in real time.

While far short of the sort of propositional representation necessary for conscious evaluation, deliberation, and choice, such a landscape of expectation values can guide action selection intelligently in ways that meet the animal’s immediate and middle-term needs in a changing environment. It thus constitutes a proto-form of a tacit practical competency. Moreover, because it is a learned information structure rather than a set of stimulus-response connections (for example, it separately encodes and updates value, risk, expected value, and relational and absolute space), it can properly be spoken of as more or less accurate, complete, reliable, grounded, or experience-tested. It thus has the necessary features to constitute a proto-form of implicit practical knowledge. Indeed, the shoe might be on the other foot. It is thanks to our capacity for such nonpropositional, bodily centered, grounded mental maps and expectations that we are able to connect human propositional thought to the world via de re and de se beliefs and intentions.

It might seem preposterous to imagine that a foraging rat, for instance, carries about in its tiny brain an “expected-value map” of its environment—and follows it when foraging or seeking mates and cooperation partners, updating these expectations continuously as it makes its choices and experiences their outcomes. And yet evolution had millions upon millions of generations to build a more efficient and effective system for animal guidance, and it appears to have hit upon processes similar to those developed via a priori investigation in philosophy and control theory. The result is a “tiny” rat brain with some 15–20 million neurons.

One reason for calling such a learned information structure a representation is that it exhibits information-value-sensitive processing that is substantially independent of current experience or context. During episodes of REM sleep after a day of training in a maze, a rat’s acquired spatial representations of the maze can be observed to be repeatedly reactivated. Neuronal recordings reveal greater activation in areas less

explored during the day—an actively “information-seeking” rather than passively habitual or associative pattern of cognition. Gaps are filled in and novel pathways begin to emerge, “shortcuts” that correspond to no actual path the rat has ever seen, much less been rewarded for taking.40 Performance in the maze the following morning is faster, and the presence of such shortcuts in the rat’s self-improved representation of the maze might help explain why Lashley’s rats could scamper directly to the food box once they’d escaped to the top of the maze.41

In humans, activation in the “default network” of the brain increases when task demands are reduced, and this activation has been hypothesized to function in a way analogous to REM sleep, drawing selectively upon acquired expectations and episodic memory to simulate and evaluate possibilities not yet encountered—a process called “prospection.”42 Like the hard drive on your computer, it seems, the brain takes advantage of lulls in task-based demands to enrich relational links in stored and recently acquired information, updating conditional expectations so that when a new task requires information, it will be able to draw upon a rich, connected structure of expectations, a model, rather than myriad stored sequences of particular facts. Elite athletes in competitive sports, for example, do not appear to owe their edge over very good athletes to more practiced motor skills—both have reached high levels of physical proficiency and draw selectively upon a wide array of well-developed motor programs. Their distinctive advantage seems to lie rather in the greater accuracy, informativeness, and accessibility of the tacit “forward models”—evaluative landscapes linking circumstances, actions, and outcomes—they develop and spontaneously follow in training and competition.43

XII. ATTUNED INTUITIONS

These action-guiding forward models fit the job description laid down by tacit-competency-based models of “intuitive” action guidance. They provide a potential mechanism by which individuals can become attuned to their physical and social environment and its demands: a “good, reli-

41. Lashley, Brain Mechanisms.
able infrastructure” to explain how skilled individuals and experts can have “reliably good intuitions.”

How might such implicit information structures manifest themselves in awareness? Recent years have seen the development of a conception of conscious emotion as providing deliberatively accessible “feelings” that correspond to underlying representations of relevant information about one’s situation and prospects. For example, A. D. Craig has proposed that the anterior insula receives continuous streams of information about one’s bodily condition, homeostatic motor functions, environmental conditions, reward state, motivational state, and goal attainment to generate conscious feelings of well- or ill-being that evolve over time and shape how we are disposed to think and act. Other familiar “feelings” such as interest, preference, desire, aversion, enjoyment, trust, uncertainty, anger, fear, surprise, suspicion, disgust, satisfaction, and frustration convey more specific dimensions of information—so that the “intuitive feelings” by which we often choose need not be simple and univocal. Often they are complex and mixed, reflecting perhaps an implicit sensitivity to a wider range of positive and negative considerations than we could consciously hold in mind at any one time. In a recent choice experiment, for example, people seemed better at matching their preferences in consumption choices with more than four desiderata when they were forced to choose without deliberation—by “intuition” or “feel”—rather than allowed ample time to ponder the choice. Moreover, because the affective system enters early into the perceptual stream to orient attention, color cognition, and prime action, the Aristotelian notion of intuitive action as guided by “evaluative perception” of a situation and its demands or prospects might not be so difficult to understand after all: a situation will look promising, a person will appear dangerous, a remark will sound wrong.

XIII. ANTECEDENTS

I have been using terms like ‘value’ and ‘evaluative’ in the psychologists’ descriptive sense, to designate a kind of informative and regulative function, not a conceptualization as good or bad, or as reason giving. Philosophers may react with skepticism to the thought that affective states, even when they function evaluatively, could actually constitute appropriate representations of value, or an element of practical knowl-

45. Craig, “Anterior Insula and Human Awareness.”
edge that could guide action “in the right way” to make it responsive to reasons. It is therefore worth noting several important philosophical antecedents for the idea that well-attuned affective states can be fitting responses to value, or part of practical knowledge or understanding.

One place to look is in theories of virtue, for, as Aristotle writes, “Virtue is concerned with feelings and actions.” He explains: “I am talking here about virtue of character, since it is this that is concerned with feelings and actions, and it is in these that we find excess, deficiency, and the mean. For example, fear, confidence, anger, pity, and in general pleasure and pain can be experienced too much or too little, and in both ways not well. But to have them at the right time, about the right thing, towards the right people, for the right end, and in the right way, is the mean and best, and this is the business of virtue.”

Through training and experience, “the courageous person is the one who endures and fears—and likewise is confident about—the right things, for the right reason, in the right way, and at the right time.” To be sure, the fully virtuous warrior will have a theoretical understanding of strategy and tactics, but Aristotle believes that no strict rules can guide him reliably in situ. Rather, he must have an affectively attuned practical appreciation of the risks involved, the potentials for taking action, and the values at stake—the “courageous person feels and acts in accordance with the merits of the case,” grounding him in his concrete situation in such a way that “what is true appears to him.” Well-attuned fear is not a mere belief about the evaluative landscape, and yet we can speak of this fear, like belief, as “reasonable,” “accurate,” or “justified.” It thus can form part of the soldier’s “practical intelligence” or “intuitive understanding,” permitting apt translation from perception to action, even when conscious deliberation is impossible and no rule can be found.

Unlike Aristotle, Kant thought there was a perfectly general rule for the determination of whether a given action is permissible or impermissible. Yet he, too, believed that a being could know this rule without being moved to act in the right way at the right time. A “merely theoretical cognition” of the moral law must be accompanied by respect for the law, which makes the law a determining ground of the subject’s choice and an incentive for her action. To avoid problems of regress in rule following akin to those we discussed in Section VIII, Kant ar-

49. Ibid., 1115b17–19.
50. Ibid., 1115b19.
gued that respect is “not a judgment about an object that would be a duty to bring about or promote” but an “antecedent predisposition[n] on the side of feeling.”\footnote{Ibid., 6:402.} Within Kant’s philosophy as a whole, it is affect that registers appreciation of value or the sublime and that moves us—“Respect . . . something that comes nearer to this feeling is admiration, and this as an affect.” Furthermore, as a form of affect, it is involuntary, “Respect is a tribute we cannot refuse to pay to merit, whether we want to or not . . . we still cannot help feeling it inwardly.”\footnote{Immanuel Kant, \textit{Critique of Practical Reason}, trans. M. J. Gregor (Cambridge: Cambridge University Press, 1996), 5:76–77.} The virtuous person is distinct from the individual who simply acts in the way duty requires, precisely because the virtuous person is moved to comply with duty in situ by the “moral feeling” of respect for the moral law and for others as ends in themselves—a form of “mental attunement,” as Kant puts it, akin to an experience of the sublime in nature.\footnote{Immanuel Kant, \textit{Critique of Judgment}, trans. W. S. Pluhar (Indianapolis: Hackett, 1987), 5:267.} Without this appreciative susceptibility underwriting the “mind’s receptivity to concepts of duty as such,” obedience to the moral law’s requirements constitutes mere “legality,” not virtue.\footnote{Kant, \textit{Metaphysics of Morals}, 6:399, 219.}

If Aristotle and Kant are our guides, then, locating a response in the affective system does not disqualify it from constituting an apt practical appreciation of the relevant moral considerations as such, or from being a constitutive part of virtuous character or practical intelligence. On the contrary, one might say. As Kant puts it, a “human being [who] is entirely without moral feeling . . . would be morally dead,” even if he had theoretical knowledge of duty’s requirements.\footnote{Ibid., 400.}

XIV. SOCIAL LEARNING, SIMULATION, AND EMPATHY

Moral assessment has distinctive features, like a measure of independence from the personal standpoint, that seem alien to affect as many philosophers and social psychologists understand it. Can the affective system really attune us to the kinds of nonpersonal social or moral values with which Aristotle, Kant, or our trial lawyer is concerned? Empathy—understood as a capacity to imagine and understand the attitudes of others from their point of view, not our own—has recently taken a central place in the literature on social cognition, moral development, and prudence. Experiencing a mild shock, anticipating the arrival of such a shock, watching another undergo such a shock, and imagining inflicting such a shock upon another, all appear to activate exten-
sively overlapping regions of the affective system. A leading hypothesis is that, as we engage in social life, we spontaneously simulate the mental states of those around us, deploying our own affective system as a “test bed.” These simulations are “off line” in the sense that they do not lead directly to behavior—perceiving someone as angry does not automatically make me behave angrily—but they nonetheless contribute essential information about the social situation and its prospects. So can my “off line” simulation of future possibilities—contemplating getting or giving a shock does not hurt me, but it registers that hurt within me.

For social beings such as ourselves, it is perhaps unsurprising that our capacity for first-personal prospection is of a piece with our capacity to project “what it is like” for others—having a reliable idea of the outcomes of my actions for myself often depends upon having a reliable idea of its effects for others. Certainly, this was the situation of our trial lawyer, into whose thinking the information had to penetrate that, however she judged the excellence of her case, it was not succeeding in winning the minds or hearts of the jurors. To function properly, then, empathy must have an “impartial” character—representing others’ attitudes and actions in their own right, not as a projection of what we think appropriate or hope to see. Our affective response in a social situation, to the extent that it is “emotionally intelligent,” takes place on several levels and in several registers at once. Whether I feel anxiety or confidence in the presence of another depends upon how I simultaneously represent the others’ (perhaps quite different) mental states. Since these representations of others are spontaneous and nonvoluntary, their emotional effect within me can be “recalcitrant” to judgment—as was the case for our trial lawyer’s rising anxiety. This spontaneous, nonvoluntary, and potentially judgment-resistant character is also to be found in the perceptual system, and for the same reason—they are part of the price for a capacity to receive feedback from experience that keeps us in touch with a reality that often is at odds with what we now believe or wish for. “Good social intuition” requires this capacity, and individuals with high IQ but profound deficits in empathic capacities find such “intuitive” knowledge of others’ mental states—or even their own future mental states—exceptionally difficult to acquire or act upon fluently.

“Good moral intuition” likewise requires an ability to “feel” what things are like from points of view that are not one’s own. Given the

inevitable influence of our own assessment of a situation upon self-conscious moral deliberation, spontaneous in situ empathic simulation can be a valuable counterforce in thought and action. For example, it can inhibit self-righteous violence. The theories of moral development that dominated the early days of cognitive psychology emphasized levels of abstraction and generality in moral reasoning, but measures of reasoning capacity failed to show a systematic connection to differences in moral conduct. An ability to empathize, and activation or inhibition of this ability in situ, appear by contrast to be important predictors of moral sensitivity and behavior.

XV. PRACTICAL DYSFUNCTION

If the broad affective system is pervasively involved in attuning our thoughts and actions to the physical and social world around us, then disorders of this system should tend to put individuals “out of tune” practically. Individuals who suffer depression, mania, or schizophrenia, or who experience damage to the interface between affective regions and the higher cortices, show systematic problems in learning, deliberation, and decision making. Psychopathy is thought by some to involve a profound deficit in empathy—and perhaps specifically emotional rather than cognitive or motor empathy—resulting in impulsivity, disinhibition of violence, instrumentalization of others, and difficulty in forming stable, reciprocal relationships. Patients suffering frontotemporal dementia and dementia due to Alzheimer’s disease typically retain normal declarative knowledge of norms of socially appropriate behavior, but those with frontotemporal dementia, which involves selective degeneration of brain regions associated with the generation and comparison of social emotions, exhibit much more serious and systematic disturbances in their ability to apply and follow social norms than those with Alzheimer’s disease. In puzzling “dissociation disorders,” it appears that the absence of the normal “feeling” of familiarity can lead one

to see close kin as “impostors,” or to lose one’s “sense” of being present here and now, or of exerting genuine agency—a radical loss of practical situatedness that can be thought of as a failure of intuition in the classical sense. Finally, since the affective system functions as a statistical learning system as well as a regulative system, we might expect those suffering neural deficits or lesions that disrupt the normal role of affect in cognition to have greater difficulty in detecting, for example, risk-benefit relationships in their environment, and some evidence suggests this.

XVI. LIMITATIONS

Of course, statistical learning systems have characteristic weaknesses as well as strengths. For example, Amitai Shenhav and Joshua Greene present evidence, consonant with the main line of argument above, that moral judgments involve domain-general evaluational capacities that have been shaped evolutionarily by such tasks as foraging for food, avoiding harm, or forming or maintaining social relationships. In such tasks, they argue, operating at the level of individuals or small groups, there is a diminishing marginal value for most goods, and so it would make sense if our “intuitive” evaluation functions discount additional units of a given resource as the total number of units of that resource gets large. Their experiments suggest that people’s spontaneous moral assessments carry this logic into areas where it does not make good sense, such as discounting additional lives saved as the number of lives saved rises. Fortunately, as they point out, self-conscious, rule-based reasoning (“count every life equally”) can be used to counteract this limitation of our more intuitive thought processes.

Statistical learning and empathy are not magic—like perception, they can afford only prima facie, defeasible epistemic justification. And—again, like perception itself—they are subject to capacity limitations and can be only as informative as the native sensitivities, experiential history, and acquired categories or concepts they can bring to bear. If these are impoverished, unrepresentative, or biased, so will be our statistical and empathic responses. Discrepancy-reduction learning is good at reducing the effects of initial biases through experience (“washing out priors”), but not if the experiences themselves are biased in the same ways—as,

66. Ibid. I am grateful here to comments by Fiery Cushman.
arguably, is often the case in social prejudice. It therefore is of the first importance to epistemology and morality that we are beings who can critically scrutinize our intuitive responses. My defense of the practical intelligence that can be embodied in intuition is not meant to encourage uncritical acceptance of “intuitions.” Rather, it seeks to identify actual processes of learning and representation that would enable us to see more clearly where our intuitions might be expected to be more reliable, but also where they might be expected to be less.

An adequate epistemology for intuition will parallel an adequate epistemology for belief generally—indeed, if the classical and tacit-competency-based notions of intuition are to be accepted, then belief formation, assessment, and application themselves are inextricably intermingled with intuition and intuitions. There might be one subject matter here, not two. And in that joint subject matter, there is a role for reflective scrutiny, considerations of consistency and coherence, norms of evidence, theory development and testing, and the application of theory to our own epistemic practices, intuitions included. Recall once more Kant’s remark, “Neither of these qualities or faculties is preferable to the other.”

XVII. SUMMING UP THE ARGUMENT THUS FAR

We have considered several notions of intuition and intuitions—observational, tacit-competency-based, and classical—and surveyed some findings of recent work in psychology and neuroscience that suggest how the affective system, broadly understood, might in fact provide a psychic basis for them. What, we asked, might this tell us about the credibility of intuition or intuitions?

Summarizing and simplifying what we’ve discussed thus far, the predominant tendency in much current research on moral judgment has been to understand the affective system as “automatic,” “point-and-shoot,” “button pushing,” and heuristic based, with “little understanding of logic and statistics” and responses that take the form of “simple likes and dislikes.” Seen this way, affective reactions would be poor candidates for the status of tacit practical intelligence or knowledge, even if they are often useful as “shortcut” stand-ins for such knowledge and understanding.

I have been arguing for a different conception of the affective system, as a system designed to inform thought and action in flexible, experience-based, statistically sophisticated, and representationally complex ways—grounding us in, and attuning us to, reality. It presents to

us an evaluative landscape of the physical and social world capable of tacitly guiding perception, cognition, feeling, and action in ways anticipated by Aristotle and Kant, among others. Seen in this alternative way, affective responses are prima facie plausible candidates for the status of elements of practical intelligence or knowledge. We must of course be careful here: in this case, as in intelligence or knowledge generally, the quality of this candidacy will reflect limitations in the system’s capacities and in the experience it receives. As no more than a philosophical observer, and therefore hardly one qualified to have an authoritative opinion, the general lesson I draw from the past decade or so of research in affective science is that our capacity for spontaneous yet apt responsiveness to reasons for belief and action has at its core the operation of implicit affective processes. We must keep in mind that the rigorous scientific study of mind and brain is still in its early stages, but at least thus far, it appears to have identified a mental architecture within which “intuition” in the classical or implicit-competency-based sense finds a natural place and could possess prima facie authority.

With this cheerful, if preliminary and oversimple conclusion in mind, let us quickly revisit some of the well-known examples used by moral philosophers and psychologists to elicit, analyze, and assess “moral intuitions.” How would this preliminary conclusion alter our understanding of these examples and associated “intuitions”?

XVIII. JULIE AND MARK AND JANE AND MATTHEW

Return to the example of Julie and Mark’s incestuous evening, discussed in Section X. Let us suppose now that our trial attorney is asked what she makes of the scenario. She begins reading:

**Julie and Mark.** Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them.

Reading these lines, she has already begun spontaneously to imagine their situation, engaging tacit competencies in causal attribution and “theory of mind” to understand the developing narrative and the way Julie’s and Mark’s minds are working. Whatever “automatic” aversion or disgust she has at the mere thought of incest will be triggered as soon as she reaches “making love,” but then so will be a rapidly spreading activation across connections in her tacit social and psychological competencies and information structures. She is already forming a “model” of Julie and Mark as individuals with certain thoughts and feelings,
and her own mind is racing ahead, projecting possible next steps of the story, sweeping out some likely outcomes of their behavior. This in turn activates more connections with stored evaluative information about the risks, benefits, and costs of these outcomes, linked to her own experience of familial relations and to what she has seen, personally or professionally, of actual incestuous behavior. Very quickly, a large amount of negative expected value is being “loaded up” on their decision to make love, with little on the positive side except their mutual accord and their thought that making love would be “a new experience” that might be fun. All of this has gotten under way, almost entirely tacitly, within the short amount of time elapsed since she began to read the story. As a result, she already has a prompt, effortless “intuitive” strongly negative affective response, but it is more than a “flash of disgust” or “simple feeling” of dislike.

She goes on to read:

. . . Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to make love?

Updating on this information, our lawyer’s strongly negative evaluative representation might soften a bit—at least, improbably, they seem to have gotten off lightly, and that’s a good thing. Even so, would reading this do much to change our lawyer’s initial assignment of high negative expected value to their actions—would she now see their decision as having been “OK”?

Not really. Their decision was a highly risky, poorly motivated idea—none of this is changed by a fortunate outcome. The vignette shows Julie and Mark to have been insensitive to, and insufficiently motivated by, the lasting harm they might have caused to each other—taking such a decision so readily, and for the sake of something that might be “interesting and fun” or, at least, a “new experience.” Our lawyer’s “Not OK” intuition stands because her affective system—like commonsense judgments of right and wrong, smart and dumb, reasonable and unreasonable generally—pays attention not just to outcomes, but to persons, motives and values, and the ways in which they decide and act. We might compare:

**Jane and Matthew**: Jane and Matthew are brother and sister. They are hiking together in Alaska on summer vacation from college.

One night they are staying alone in a cabin in a high pass. They decide that it would be interesting and fun if they tried playing Russian roulette with the revolver they are carrying with them for protection from bears. At very least it would be a new experience for each of them. As it happens, the gun does not go off, and neither suffers any lasting trauma from the experience. They both enjoyed the game, but decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that, was it OK for them to play Russian roulette with a loaded revolver?

Was it? Whether or not the thought of Russian roulette triggers a “flash of disgust” in us, we also immediately appreciate as we read the story what a senseless loss it would have been had the gun gone off—and what a world of pain this would have left behind. Jane and Matthew carelessly put all this at risk for the sake of a potentially “interesting and fun” evening. Not OK, despite the fortunate outcome. In my experience, students who have been given the case of Jane and Matthew before Julie and Mark have little trouble identifying a rationale for their negative verdict in the latter case that is not obviated by its fortunate outcome. Perhaps these students are simply seizing on this fig leaf to cover their gut disgust. No doubt some are. But others might be understanding their own negative response a bit more deeply, and why it resists hearing of fortuitous consequences.

Julie and Mark played Russian roulette with their psyches, arguably with more than a one-in-six chance of serious harm. The fact that experimental subjects had such harms uppermost in their minds when queried about their disapproval need not show mere confabulation, since running the risk of these harms is relevant to the question whether their conduct was “OK.” Myself, I would have been more worried about the quality of intuitive moral assessments if Haidt had found that people, once reminded that no lasting harm resulted, simply lost any “sense” that Julie and Mark’s decision and action were “Not OK.”

In this connection, it is noteworthy under what conditions “immediate” intuitive assessments of this kind can be reversed. One of the key insights of recent research on affect is that the affective system recruits whatever information it can—from sensation, memory, imaginative simulation, past thinking, and so on. Even a seemingly basic affective response such as disgust at a smell involves a complex synthesis of information from multiple senses and memory. Whether smelling volatile carboxylic acid disgusts or delights depends upon whether one is opening the door of one’s favorite cheese shop or one’s little-used gym locker.

Social change can effect just such a change in the information people receive. The psychological literature on moralized disgust has long
included attitudes toward homosexuality, and various experiments have found that increased disgust or sensitivity to disgust goes along with increased disapproval of homosexual relations. While there has been a tendency in the literature on stereotypes and prejudice to assume that tacit affective responses are fairly stable and resistant to evidence, a growing body of evidence suggests otherwise. The “evolved function” of disgust, like all affective states, must be combined with a dynamic, “information processing” perspective.

Indeed, because statistical learning systems operate by discrepancy reduction, incongruous information receives special attention in implicit processing. Eight-month-old infants pay greater attention to sequences of phonemes they have not yet heard and update their conditional expectations accordingly. Exposing individuals to subliminal images of admired individuals in stigmatized groups can result in less “automatic” stereotyping, even prior to any change in explicit or conscious bias. And working together successfully with someone from a stigmatized or “alien” group on a meaningful project to which each contributes also reduces such implicit bias—as our trial lawyer and her client discovered in the final days of the trial.

Such evidence is especially intriguing in light of the “natural experiment” over the last decade of changing attitudes on gay marriage,


where rates of disapproval versus approval of permitting gay marriage virtually inverted, so that now a majority favors permission. The increased openness of individuals about their own sexual orientation had the effect of dramatically changing the information most people have about gays and gay relationships—over the last decades, many heterosexuals have realized that a number of close kin and other individuals they know, admire, love, or work with collegially, are gay. Our affective system will use the information it gets, and with wider and more representative experience, it can work tacitly to shift biased “intuitions,” even those that have persisted in the population for centuries and been backed by powerful religious or cultural institutions. Here again we see collusion of the epistemology of moral intuition and the epistemology of belief generally.

XIX. IN THE EXECUTIVE SUITE

But what if things are the other way around? If, instead of experience-based underlying causal models and social competencies helping to guide tacit moral assessment in ways responsive to evidence, our tacit moral assessments themselves tend to preempt the role of evidence in our social attributions? This challenge to the “lay scientist” model of social cognition has been posed sharply by Josh Knobe’s influential work relating moral evaluations to the imputation of intentionality in individual actions.75

Knobe introduced a striking pair of examples.76

**Boardroom 1:** The vice-president of a company went to the chairman of the board and said, “We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.” The chairman of the board answered, “I don’t care at all about harming the environment. I just want to make as much profit as I can. Let’s start the new program.” They started the new program. Sure enough, the environment was harmed.

Did the chairman intentionally harm the environment?

Most say, yes.

**Boardroom 2:** The vice-president of a company went to the chairman of the board and said, “We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.” The chairman of the board answered, “I don’t care at all about helping the environment. I just want to make as much profit as I can. Let’s start the new program.” They started the new program. Sure enough, the environment was harmed.

Did the chairman intentionally harm the environment?

75. For a review and discussion, see J. Knobe, “Person as Scientist, Person as Moralist,” *Behavioral and Brain Sciences* 33 (2010): 315–29.

76. Ibid., 317.
profit as I can. Let’s start the new program.” They started the new program. Sure enough, the environment was helped.

Did the chairman intentionally help the environment?

Most say, no.

Yet what is the difference between the two scenarios except for replacing ‘harm’ with ‘help’? Knobe concludes that “people’s moral judgments can actually influence the intuitions they hold both in folk psychology and in causal cognition”—the “lay scientist” is a moralist, bending her attributions of intentionality to fit her values.77

Knobe has since gone on to extend his interpretation to a range of intentional idioms. But do such paired examples show that evidence and tacit causal and social competencies are not at work? Imagine:

**Goat 2**: You are a goat herder living on a hillside above a remote village. Your uphill neighbor has an orchard of olive trees. One day you are browsing the shelves in the agricultural supply store in the village, when you overhear your neighbor being told by the man at the register, “Yes, you could use this spray on your trees, and it would kill all the bugs, but when the rain comes it will wash off and run down onto your neighbor’s fields—killing the bugs in her grass and helping her goats.”

You listen intently for his answer:

“I don’t care at all about helping her goats, I just want to kill the bugs on my olive trees. Give me the spray.”

How does this feel? Your neighbor is about to use a spray that will help your goats, but he doesn’t care at all about that. Not a great attitude—a selfish indifference toward helping you is a problem of sorts in a neighbor. But on a scale ranging from good will to ill will, with neutral in the middle, it is somewhere near the center—live and let live. If you imagine complaining to others in the village of your neighbor’s lack of interest in whether his spraying to protect his olive trees will benefit your goats—“my main source of livelihood!”—could you expect much alarm, outrage, or sympathy?

Now imagine:

**Goat 1**: You are a goat herder living on a hillside above a remote village. Your uphill neighbor has an orchard of olive trees. One day you are browsing the shelves in the agricultural supply store in the village, when you overhear your neighbor being told by the man

---

77. Ibid., 315.
at the register, “Yes, you could use this spray on your trees, and it would kill all the bugs, but when the rain comes it will wash off and run down onto your neighbor’s fields—harming her goats when they eat the grass.”

Again, a keen interest on your part will attach to your neighbor’s next words.

“I don’t care at all about harming her goats, I just want to kill those bugs. Give me the spray.”

What sort of neighbor have you got? On the scale from good will to ill will, is his attitude located at the same point as the neighbor in Goat 2? Is not caring at all whether one harms a neighbor a relatively neutral, live-and-let-live attitude? Of course, your neighbor didn’t set out to harm your goats—he says he’s thinking only of his trees. But this is clearly a case of one thought too few. And if you imagine telling others in the village of your neighbor’s indifference to whether his spraying harms your goats—“my main source of livelihood!”—do you more easily imagine garnering some alarm, outrage, or sympathy?

Goat 1 and Goat 2 suggest that Boardroom 1 and Boardroom 2 are less symmetrical than they first appeared. Intuition makes use of whatever evidence it can, and given our experience, someone who “doesn’t care at all” about whether he harms his neighbors is, happily, statistically rare, while someone who “doesn’t care at all” about whether he helps his neighbors is, perhaps regretfully, much more common. Statistical learning systems pay special attention to anomalies, since they carry more information than events that are more predictable. The attitude of the chairman in Boardroom 1 and the neighbor in Goat 1 attract such attention, and in both ordinary life and clinical diagnosis these attitudes would indeed be viewed as more “antisocial” than the attitude of the chairman in Boardroom 2 and the neighbor in Goat 2, calling for greater than normal vigilance on the part of others. 78 Since an action involving knowingly, willingly, and with indifference harming others as a completely foreseeable side effect is commonsensically understood to be “antisocial,” the actions in Boardroom 1 and Goat 1 are seen as congruent with the rather “antisocial” attitudes expressed by the chairman and neighbor. It is natural, therefore, for “intuition” to treat these actions as to that extent intentional. By contrast, the indifference toward the completely foreseeable side effect of helping others expressed by the chairman in Boardroom 2

78. Indifference to causing harm to others, and willingness to violate the rights of others without remorse, are key diagnostic criteria of Anti-Social Personality Disorder. See American Psychiatric Association, *DSM IV-TR*; and R. D. Hare, *Without Conscience: The Disturbing World of the Psychopaths among Us* (New York: Guilford, 1993).

This content downloaded from 18.7.29.240 on Fri, 8 May 2015 09:47:12 AM
All use subject to JSTOR Terms and Conditions
and the neighbor in Goat 2 is not commonsensically seen as signaling the presence of a relatively “prosocial” attitude on their part. So an act with the “prosocial” side effect of helping others is not particularly congruent their attitudes, and to that extent is seen “intuitively” as less intentional. What may matter in such intuitive social attributions of intent with respect to side effects is the fit of the action with the causal-attitudinal-intentional model of the agent we tacitly construct in light of his or her behavior—for example, as “antisocial” versus “self-concerned” versus “prosocial”—rather than the moral quality of the side effect itself (for experimental evidence that evaluation of the chairman’s underlying character mediates the “Knobe effect,” see the work of Chandra Sripada and colleagues on the “deep-self concordance” model). 79

As we noted above in discussing the “impartiality” of empathy in generating implicit causal-attitudinal-intentional models of social situations, it is vitally important for us to be able to gauge others’ attitudes and intentions accurately even—perhaps especially—when they are at odds with our own. It would be surprising, then, if our causal-attitudinal-intentional modeling of others were skewed so strongly by our own moral values, rather than tracking evidence of the values of others—and thus, crucially, of their likely behavior.

XX. BUS AND TROLLEY

As we saw, Greene and colleagues concluded that the asymmetry in intuitive moral acceptability in the well-known trolley cases arises because in Footbridge, but not Switch, “the force that directly impacts the victim is generated by the agent’s muscles (e.g., in pushing).” 80 Yet consider the following scenario, which I presented to my students in a large introductory course in moral philosophy, just before giving them Switch and Footbridge:

**Bus:** You live in a city where terrorists have in recent months been suicide-bombing buses and trains. The terrorists strap explosives to themselves under their clothing, and, at busy times of the day, spot a crowded bus or train and rush aboard, triggering the bomb instantly to avoid being stopped. You are on a very crowded bus at 5:10 pm, and are struggling to get to the door at your stop. The doors are starting to close and you won’t be able to get off unless you jostle the slow-moving obese gentleman trying to exit at the same time.

---


Suddenly you notice a man rushing up to the bus and forcing his foot into the doorway, wedging it between the fat man and the door frame. He is reaching with one hand under his coat and a gap between the buttons reveals to you what look like explosives strapped around his chest. You can’t reach this man, but if you push the corpulent gentleman beside you hard in his direction right now, he will fall directly on top of the seeming bomber and both will end up on the empty sidewalk, while you fall backwards into the bus as the doors snap shut.

—So, if you push hard, and this man is not a bomber, then the bus will leave behind two very annoyed men on the sidewalk, and you will be left on the bus, covered with embarrassment. But if he is a bomber, the bus will be spared, and you with it, but the fat man killed as the bomber explodes underneath him.

—On the other hand, if you simply squeeze off the bus alongside the corpulent gentleman and do nothing more, and the other man is a bomber, then many people on the bus will be killed while you and the corpulent gentleman are safe on the sidewalk. But if this man is not a bomber, then no one on the bus will be hurt and you simply will have jostled a corpulent gentleman while exiting a bus, and you can apologize to him on the sidewalk.

Whatever happens, you will not be killed if there is a bomb and it goes off—you will either be on the bus when it explodes on the sidewalk, or on the sidewalk when it explodes on the bus.

Should you (a) shove the corpulent gentleman hard right now, or (b) squeeze off the bus, jostling the corpulent gentleman but doing nothing else?

When I ran this uncontrolled and unscientific experiment, I received the following results (with thanks to Warren Harold):

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Push the corpulent gentleman?</th>
<th>Pull the lever on the switch?</th>
<th>Push the “fat man” off the bridge?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus</td>
<td>67% Yes</td>
<td>33% No</td>
<td></td>
</tr>
<tr>
<td>Switch</td>
<td>72% Yes</td>
<td>28% No</td>
<td></td>
</tr>
<tr>
<td>Footbridge</td>
<td>29% Yes</td>
<td>71% No</td>
<td></td>
</tr>
</tbody>
</table>

What might this pattern of judgments tell us? Why might the exertion of “personal force” in Bus garner almost as much acceptability as pulling a lever in Switch? It might tell us that morally irrelevant details distinguish Bus from Footbridge in the intuitive mind—buses and bombs versus trolleys and wheels, say. Or it might tell us something rather different.

Let’s again try the thought experiment of asking how our broad affective system might be responding to the scenarios. First, try to imagine yourself on the bus, pushing the large gentleman onto the seeming bomber. Vividly—under what feels to you like life-or-death pressure, actually feeling the heft of his body as you shove with all your might, aiming as best you can squarely at the seeming bomber? Now, try to imagine
yourself pushing the “fat man” off the bridge. Vividly—under what feels to you like life-or-death pressure, actually feeling the heft of his body as you shove with all your might, aiming as best you can to hit the track below squarely? I am guessing that for many of you, the answer to the first question is that you can credibly simulate this imaginatively, but in the second case you cannot—even if you are inclined to judge that the heavy man should be pushed in Footbridge. When I try to imagine heaving the man off the bridge, I find I can at best feel myself doing it mechanically, with almost physical reluctance. Not so in the case of Bus. What might this tell us about the kind of motivational capacities pushing requires in the two cases?

Try now to imagine the social setting. Imagine those on the bus who see the apparent bomber’s approach and see your situation in the doorway, separated from the apparent bomber by a large, oblivious man. Can you also imagine them shouting to you, “Push him, for God’s sake!”? By contrast, imagine the five track workers seeing the trolley hurtling toward them and seeing you on the bridge next to the large man. Can you also imagine them shouting up to you, “Push him, for God’s sake!”? Now try to imagine the aftermath of pushing in each case, and imagine that in Footbridge you succeeded in stopping the trolley and that in Bus the bomber was real and the explosion took place on the sidewalk. You now have to bear the awful fact of having pushed an innocent man to his death. In which case, if either, can you imagine bearing this more easily? And is the mixture of feelings involved—guilt, regret, shame, disgrace—the same in each case? For which, if either, can you imagine that others, including members of the innocent man’s family, might show more acceptance, understanding, or forgiveness of your action? Or, imagine that you failed to push the man in each scenario. In Bus, can you imagine feeling haunted by guilt or regret? Can you imagine that others would understand if you did? Now imagine not pushing in Footbridge. Would this haunt you in the same way? And would others understand equally well if it did?

If you are like me, you will feel asymmetries between Bus and Footbridge when taking up these different perspectives, asymmetries that parallel the original felt asymmetry between Bus and Footbridge. They are all of a piece. This suggests, if I am right, that the asymmetry at work here is relatively standpoint independent—it has to do with an implicit, shared social understanding of the situations and has little to do with the distinctive perspective of the person who pushes. Both Bus and Footbridge involve personal force and fit the “me hurt you” schema, yet imagining these two cases results in a systematic, inter-

connected set of asymmetries in "reactive attitudes." These imaginative structures suggest that there is more at work here than a "blunt biological response" or "myopic module" that sounds an alarm but ignores collateral effects.

I’m sure I cannot fully articulate the implicit, shared social knowledge that makes pushing someone to a gory death in Bus virtually indistinguishable in moral acceptability to my students from pulling a remote lever in Switch, and so different from Footbridge. Some factors leap to mind. In Bus one begins by imagining oneself among the passengers who would be killed if the bus were bombed, and this might do more than the scenario in Footbridge to “equalize” one’s empathic sensitivity to the fates of those one might save with one’s empathic sensitivity to the fate of the man who would be pushed. I suspect, too, that given widespread awareness of terror bombing in recent decades, Bus engages a sense of social solidarity and of the need for collective self-defense. This would make Bus closer than Footbridge to another staple example, in which one might be forced to smother a crying baby to save a group of people in hiding from a genocidal attack—even if the attackers are known to spare babies while slaughtering all the rest.

The questions to ask might be: What aspects of the stage setting of classic trolley problems make it unlikely for them to engage a strong sense of social solidarity or collective self-defense? And how does the stage-setting selectively affect our empathy for potential victims? Perhaps Footbridge has narrower relevance in assessing overall moral theories than many philosophers imagine, and cases like Bus deserve more attention. This might be a different way of reading the import of studies suggesting that individuals exhibiting lower than normal levels of empathy—for reasons of psychopathology, brain damage, or social distance—exhibit less asymmetry in their judgments of Footbridge and Switch. So I am not arguing that intuition “gets it right” in either Footbridge or Bus, only pleading for closer attention to what these intuitions might be telling us about what we represent, or fail to represent, in assessing these scenarios intuitively.

83. Greene, Moral Tribes, 236–40.
84. For example, even if we accept a negative verdict in Footbridge, this hardly would tell us that consequentialism is unacceptable. We’d have to balance this evidence with evidence from cases like Bus or smothering the crying child. And rule- or motive-utilitarianism might be able to draw upon facts about typical human motivation and in extremis social requirements to yield the result that sacrificing an innocent person would be consistent with the best rules or best motives in cases like Bus and Switch, but not in cases like Footbridge.
In any event, I’ll soon be way over my head if I wade any further into the deep waters of trolleyology and intuition testing. More and better-worked-out examples, and more sophisticated data analysis, would be needed before drawing any conclusions about what cases like Jane and Matthew, Goat 1 and 2, and Bus might tell us about the source or credibility of the moral intuitions they elicit. Unlike those psychologists and philosophers who have carried out recent decades’ developments in empirical moral psychology, I myself have performed no controlled experiments and tested no real hypotheses. From an empirical standpoint, all I have offered are anecdotes.

These anecdotes, however, share certain features. They suggest that Hume might have been right in saying that intuitive moral evaluation of behavior is tightly bound to our concern with underlying attitudes and motivation. What an act means for us and for the community depends not only upon its specific, perhaps fortuitous outcomes, but upon what it might say about the evaluative landscape of potential good or ill we face going forward. Family, friendship, trust, cooperation, and social solidarity are the most important ways we have of ameliorating that landscape, and for these to flourish people should not: be willing to engage in incest with a loved one for the sake of amusement, even with consent; be indifferent to the harm they cause others, even when it is a side effect; have little resistance to pushing an innocent bystander to his death, even to prevent an accident; or be unwilling, in extremis, to try to overcome such resistance to carry out necessary acts of collective self-defense. These are not matters of global moral character, I suspect, but of contextually more specific dispositions to feel and act. And intuitive moral evaluation appears to be complex enough to be sensitive to the fine-grained differences in context, attitudes, intentions, and behavior that make all the difference to us as social beings.86

With the help of anecdotes, supplemented by some evidence from genuine research done by others, I have made a few, tentative suggestions about when intuitive moral assessments might be expected to have greater credibility—even when they oppose one’s own considered judgment: for example, when individuals have wider and more representative experience, a better-developed ability to imagine what things would be like from the standpoints of others, a better “feel” for the underlying dynamics in personal and social situations, or greater foresight in imag-

ining alternatives. These are also, I think, characteristics of those people whose intuitive moral responses we especially value or trust. What is it about these people that gives their intuitive responses greater authority for us? Is it that they hold moral principles we share? Many people who share our principles are decidedly not individuals to whom we would turn in difficult decisions. I suspect that we seek out people who strike us as having well-developed implicit social and emotional competencies in virtue of which they are better attuned to the evaluative landscape of concerns, values, risks, and potentialities inherent in the actual, messy situations we face. These are individuals whose intuitive assessments are, by our own lights, likely to be more reasons-responsive than our own.

Philosophers have long seen the necessity of postulating implicit competencies, sensibilities, or "faculties" that can attune us nondeliberatively to reasons to think and act. And in recent years there has begun to be impressively detailed empirical evidence that "intuitive attunement" might have psychic reality after all. Intuition appears to be alive and well, and living in the affective system, broadly understood. And that is the rational tale of the affective dog.

Better understanding of this tale, and what it tells us of the strengths and weaknesses of the affective system in keeping us in touch with the realities and possibilities of physical and social worlds around us and the psychic worlds within, can afford moral philosophers something we long have needed: a firmer theoretical and evidential basis upon which to explore the nature and informativeness of the "moral intuitions" upon which we rely extensively and inevitably. In short, a better-grounded epistemology of moral intuition, enabling us to assess more critically and confidently intuition’s diverse roles in moral thought and action. Pioneering work by psychologists, neuroscientists, and experimental philosophers has opened the way, and now we must see where it leads. Moral philosophy will never be quite the same.