Teaching Philosophy
DOI: 10.5840/teachphil20251119234
Volume 48, Issue 3, September 2025: 545–567

What Are We Teaching When We Teach Moral Reasoning?

BRIAN HAMILTON
Florida Southern College

Abstract: If Jonathan Haidt is right that moral reasoning is typically a post-hoc defense of intuitive judgments, why teach ethics? In answer, I show first that the main conclusions of Haidt's "social intuitionism" are anticipated already in the classical moral psychologies of Socrates, Aristotle, and Augustine. Second, I suggest that these thinkers, though far from "delusive" about the power of reason to make us better people, still identify a richer, more constructive role for moral reasoning than Haidt does. Moral reasoning in the form of deliberative dialogue might be understood and taught as a kind of "spiritual exercise in common," as Pierre Hadot puts it—a means of confronting our own ignorance together with others and rendering our intuitions vulnerable to (re-)formation.

Key words: moral psychology, social intuitionism, rationalism, moral reasoning, ignorance

In his influential book *The Righteous Mind*, moral psychologist Jonathan Haidt sweepingly dismisses what he calls "the rationalist delusion" that reason ought to govern our moral lives.[1] He knows very well that in dismissing it he is opposing himself to the mainline tradition of Western moral philosophy, at least "from Plato through Immanuel Kant."[2] In fact, he revels in the opposition. He sides with Glaucon against Socrates, saying that morality is about reputation rather than genuine virtue.[3] Above all he sides with David Hume, who famously claimed that "reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them."[4] Haidt thinks the mainline tradition is at bottom a centuries-long exercise in self-justification—specialists in reason concocting elaborate theories that underwrite the need for specialists in reason.

Haidt does not think, however, that philosophers are uniquely devious or conspiratorial. He believes that self-justification is one of

the primary evolutionary purposes of reason—for everyone. Reason, to use one of Haidt's favorite analogies, is like a press secretary for our deeper emotional self.[5] It is usually called into action to defend already settled intuitions when they are challenged. In theory, human beings may be able to think their way to a moral conclusion, but that rarely happens in practice. We do what we do, says Haidt, on the basis of automatic intuitive judgments, and those judgments are not usually susceptible to reason's influence. If they can be influenced, Haidt suggests, it is usually by social pressure: we are exposed to the intuitions of others, and to the social consequences of our own intuitions, and that produces new intuitions in us. This, in brief, is Haidt's "social intuitionist" theory of moral reasoning.[6]

So Haidt would say that philosophers are using their reason to do what reason always does: to justify the intuitive judgments of the reasoner. It just so happens that philosophers have had an unusual amount of social success, and have therefore steered many others to the delusive conclusion that reason ought to govern the passions. What convinces Haidt that their conclusion is wrong is of course not a philosophical argument but an empirical one: when we watch the way in which people actually behave, we find that reason serves the passions rather than the other way around. The philosophers, therefore, have built their normative arguments on mistaken factual premises.[7] For true insight into our moral lives, we ought to turn not to philosophers but to psychologists.[8]

If Haidt's social intuitionist model of moral judgment is right, then moral philosophers might seem to be in significant trouble. The sort of reasoning we have made our stock in trade could look like little more than an instrument of self-justification or social manipulation. What's worse, most of us are engaged in teaching this sort of reasoning to impressionable young students. We like to think that it is good for our students to learn something about moral reasoning, that it may even make them better people. But what if we are only equipping them to defend their pre-existing moral intuitions more skillfully, regardless of whether or not those intuitions are true? What if we are only training our students in the art of sophistry, playing with words to make the weaker argument stronger—without making them better people? It's worth noting that Protagoras, an actual sophist, also claimed to make his pupils better people. He promised young Hippocrates that he would go home at the end of every day better than he was at its beginning.[9] And he would become better, Protagoras boasted, precisely because he would become more skilled in the art of *euboulia*, sound deliberation, in his private and public life.[10] Socrates, though, suspected that Protagoras was deceiving himself and his students. Haidt seems to share Socrates' skepticism—though he might be surprised to find himself on the side of such an avowed "rationalist."

I am not a psychologist, and I do not intend to challenge Haidt's empirical observations. I am, rather, a teacher of philosophical and theological ethics, and I want to think alongside Haidt about what it is I am trying to teach when I teach students moral reasoning. Given the strong evidence Haidt provides about the primacy of intuition in the actual experience of moral judgment, what useful role does explicit moral reflection play within moral education?

I agree with Haidt that explicit and open-ended moral reasoning plays only a modest role in our moral lives, even in the most virtuous lives. I agree too that moral reasoning ought to be recognized as an irreducibly social process. I therefore agree with the main provocations of Haidt's social intuitionist theory. But I want to suggest that the classical "rationalist" philosophical tradition—which I will represent here especially with reference to Socrates, Aristotle, and Augustine—already recognizes these facts about moral judgment, even if it established these facts on phenomenological rather than experimental grounds. Haidt underestimates the subtlety of the so-called rationalist tradition, and thereby denies himself of a useful set of conversation partners.[11] In particular, I want to suggest that these ancient rationalists identify three central goods of moral reasoning that Haidt neglects: namely, the good of communal dialogue itself, the good of naming and challenging our intuitions, and the good of confessing our ignorance. These uses of moral reasoning are fundamental to their own sort of rationalism, and should be fundamental too to the kind of moral reasoning that ethics teachers model and encourage in the classroom.

Although this essay focuses more on the theory than on the practice of teaching ethics, I do also intend to draw some practical lessons from the ancient rationalists I consider along the way. Socrates, Aristotle, and Augustine were ethics teachers too, after all, in their own often surprising ways. Common to all of their approaches, as different as they otherwise are, was a strong emphasis on conversation, and we might say more specifically, on *dialogical self-examination*. Their dialogical pedagogical style is directly related, I will suggest, to their understanding of the role moral reasoning plays in human lives.

Ancient Rationalism and the Passions

No ancient moral philosopher would have been surprised to hear that people typically make moral judgments on the basis of what Haidt calls intuitions, which he understands as the conclusion of an affect-laden automatic process.[12] The really hard question for them was whether those affect-laden, automatic judgments could be trained so that they were more nearly true, and if so, how. They wanted to know, in other words, whether virtue can be taught. The question occurs almost ob-

sessively in Plato's dialogues, without clear resolution. (This should have been Haidt's first clue that perhaps Plato is not quite so delusional about the power of reason as Haidt suspects.) Whether virtue can be taught sometimes becomes the *explicit* question under discussion, as in the *Protagoras* or the *Meno*, but it is *implicit* in the structure of all the dialogues inasmuch as Socrates is always attempting, through his winding cross-examinations, to prompt someone to reconsider an intuition they were not even aware they were acting on. Importantly—just as Haidt himself would tell us to expect—-Socrates almost always fails to convince anyone. Indeed, rather than be led by rational inquiry to a new moral judgment, his interlocutors usually double down on their initial intuition. Euthyphro goes on convicting his father; Meletus goes on convicting Socrates.

Nonetheless, the very fact that Socrates proceeds in the way he does seems to show that he assumes philosophy is at least potentially effective in transforming intuitions. Otherwise, what's the point of the inquiry? This is the assumption that bothers Haidt. Even if Socrates recognizes that people typically *do* operate on the basis of intuitions, Socrates seems to think that people *ought* to operate on the basis of dispassionate reason instead. His mistake, in Haidt's view, lies not in the diagnosis but in the prescription. The delusion lies in the idea that people should, or even can, operate on the basis of reason rather than passion.

There's no denying that Socrates himself is a sort of rationalist, along with the tradition that follows him. Socrates seems to have believed that people always do what they think is best for them at the moment of the action, a view sometimes called "motivational intellectualism."[13] Much of Socrates' questioning, moreover, seems to presuppose that one must know what a virtue is, cognitively, before one can actually possess that virtue.[14] Plato does indeed say that the rational part of the soul ought to rule the spirited and appetitive parts.[15] Aristotle says the function of the human beings is a life of action in accordance with reason.[16] All the later Hellenistic schools, as Martha Nussbaum has shown, hold that "the diseases that impede human flourishing are above all diseases of belief and social teaching," and thus that philosophy—definition, dialectic, deliberation—is necessary to heal us.[17] Haidt is right that these thinkers are rationalists according to his own definition: they believed "that reasoning is the most important and reliable way to obtain moral knowledge."[18]

But to leave it there would give us a very lop-sided picture of these thinkers' moral psychology. That they all believed reason was a *sine qua non* of good moral judgment does not tell us all we need to know about how they understood the role reason plays in our actual experience of moral judgment—or whether other more "intuitive" or

"emotional" or "passionate" forms of engagement also play a role. Haidt tends to assume a flatly linear sort of reasoning in his critique of "the rationalist delusion." He means mainly that people do not dispassionately collect evidence and organize reasons when they make moral judgments, and he implies that rationalists "from Plato through Immanuel Kant" suppose that they do, or at least that they should. But that is not what ancient rationalism looked like.

Consider Aristotle, for example. Aristotle defines virtue as a settled capacity for rational decision,[19] and regards reflective choice (what he calls *prohairesis*) as, in Haidt's words, "where the action is" in moral judgment.[20] He is thus, according to Haidt's definition, a rationalist.[21] But Aristotle is also keenly aware of reason's limits. He knows very well that the sort of reasoning typical of his writing and classroom will not by itself lead anyone to virtue. On the contrary, he insists that the effectiveness of such reasoning depends on the proper *paidea* of the emotions having already been accomplished. We are taught from our youth to find some things pleasurable and others painful, and if that training has been bad, dialectic will be pointless. "On the whole," Aristotle says, "passion yields not to argument, but to force."[22] Thus parents and lawgivers train those in their care through praise and shame, reward and punishment, not through formal explanations. Aristotle is the first to admit that nobody who has been badly habituated will be saved by philosophical argument. That is why Aristotle saw no point in allowing any but the sons of the *aristoi* into the Lyceum—because these alone had reasonable odds of having already been socialized, *taught*, to feel the right pleasures, or in Haidt's terms, to make the right intuitive judgments.[23] Only those who had been taught well from childhood could be taught well by Aristotle.

Even as he teaches logic and dialectic, then, and even as he believes that good action depends in some sense on knowledge and judgment, Aristotle does not suggest that logic and dialectic will make his students good people. Here is another glimpse of an ethics teacher who, though a rationalist, does not fall prey to the rationalist delusion. I do not doubt that Aristotle thought that formal training in logic and dialectic could do some good for a person. As Martha Nussbaum points out, Aristotle saw the practice of moral reasoning with and alongside others—the sort of reasoning he modeled and encouraged in his own teaching—as part of the process of our habituation into good citizens.[24] It is not only children who can be habituated.[25] But Aristotle situates this dialogical formation within a much broader pedagogical context that extends all the way back to childhood and far outside the intellect.

Aristotle places so much emphasis on the non-rational dimensions of moral education and moral development, in fact, that it has been tempting for some readers to interpret him as an *anti*-rationalist in

his moral philosophy. Terence Irwin lays out a strong case for that interpretation in his magisterial history of Western moral philosophy, *The Development of Ethics*, building on the details I have already mentioned.[26] But he ultimately argues that "the case rests on a selective treatment of the evidence."[27] Even as Aristotle insists that virtue is a matter of habituation, and that habituation begins with various forms of non-rational training, he also insists that habituated virtue ultimately *includes* learning to think, to deliberate, clearly and well.[28] Although there is much more debate among scholars about Socrates' view on these matters, it is at least plausible to read him as agreeing with Aristotle that the basic groundwork of moral education is often accomplished through non-rational means (shame or even punishment[29]) but brought to its completion through rational reflection. A virtuous person doesn't just take pleasure in the right sorts of things; she also knows how to think about those pleasures and how to deliberate about how to respond to them.

Socrates and Aristotle are not unusual among their contemporaries in holding *both* that rational deliberation is essential to true moral judgment (the rationalist axiom) *and* that rational deliberation is largely impotent against deeply habituated passion. There is in fact broad agreement among the ancient Greek moralists about these basic ideas. They did not use the language of "intuition," but they did see the passions or emotions as performing a function very similar to what Haidt means by intuition—a kind of "fast reasoning" that delivers evaluative judgments or appraisals of value. As Martha Nussbaum says, ancient Greek moralists thought of emotions as "forms of intentional awareness" with "a very intimate relationship to beliefs."[30] My emotions depend upon and express beliefs about what is good for me, though the beliefs they express may be ones I have never considered or did not even know I held. My fear of death, for example, is rooted in a belief that dying will be bad for me.[31] Because my deeply habituated passions depend upon and reflect these long-held beliefs, I am likely to resist giving up those beliefs just because a new argument seems to show them to be in error; I am more likely to distrust the new argument than to distrust my old belief. But for the same reason, those passions are at least in principle subject to rational scrutiny. "Emotions may appropriately be assessed as *rational* or *irrational*," Nussbaum continues, "and also (independently) as *true* or *false*, depending on the character of the beliefs that their basis and ground."[32] So it is no contradiction to hold, with Haidt, that "intuitions [or passions] come first, strategic reasoning second" at the level of concrete experience, and yet, pace Haidt, that rational reflection provides trustworthy grounds for *criticizing* those initial evaluative judgments.

My point in all of this is that ancient rationalism did not presuppose, either descriptively or normatively, a dispassionate moral psychology; nor did it presuppose that people always make decisions on the basis of explicit deliberation. The rationalism of Socrates and Aristotle allows that even reflective action necessarily depends on and contributes to the peculiar configuration of implicit beliefs and prima facie desires that animate my "intuitions," my immediate evaluative judgments in a particular situation. Their view is thus compatible with Haidt's view that people do not typically stop to reflect at all, and with his view that if people *do* stop to reflect, it is typically to defend their intuitive judgments rather than to revise them. Again, after all, this is almost all that Socrates' interlocutors ever do.

Reason and Moral Knowledge

Yet Socrates and Aristotle nonetheless stubbornly cling to the belief that it is *better* to act reflectively than unreflectively, to the belief that reasoning gets them closer to something like moral knowledge. It is better to be Socrates than to be Euthyphro. Thus they remain rationalists rather than intuitionists.

The Greek moralists of the Socratic tradition believe that even many of the unreflective judgments we never have cause to question are in fact quite bad for us. The pleasures most of us have been socialized to feel and pursue—pleasures of wealth and honor and power above all—will ultimately harm us. Our only hope of happiness lies in retraining our pleasures and pains. Because they held that all emotions were ultimately rooted in beliefs about the world, as we have seen, they believed that such retraining was in principle possible, and possible precisely through philosophy, through the art of reasoning. Philosophical education can be the instrument for such modification, even if some long-held beliefs prove stubbornly resistant to change. And more importantly, it is only philosophical reasoning that has a hope of getting happiness *right*.

Haidt, by contrast, seems to think that a sufficiently extensive research program into how people *do* make moral judgments will itself give us all the answers we need about how they *should*. He is resuming, he says, a Humean project of "naturalized, scientific" moral philosophy grounded in real observation.[33] When he finally addresses himself directly to normative questions—in the last two paragraphs of the book's penultimate chapter—he grants that his "definition of morality was designed to be a descriptive definition; it cannot stand alone as a normative definition."[34] The normative position he finally prefers he calls "Durkheimian utilitarianism," meaning that the best moral system is the one that maximizes the public good (the utilitarian

part), keeping in mind that human beings are "groupish" creatures who have an innate need for social belonging (the Durkheimian part).[35] He does not argue for that position so much as simply state his own intuition (if I may call it that) that "there is no compelling alternative."[36] But it's clear that Haidt's main criterion for preferring Durkheimian utilitarianism is that it is the theory, of those he knows, that best fits his reading of the empirical evidence. Haidt believes he has access to enough data to identify the natural "taste receptors" of human moral experience, which he boldly names "moral foundations."[37] Even more important, he thinks he has "the most powerful tool ever devised for understanding the design of living things: Darwin's theory of evolution,"[38] which leads him to believe that the purpose of reason is not to arrive at the truth but to secure social alliances.

Haidt would not deny that rational reflection can, over time, influence our intuitions. Although he thinks that reason *usually* works as a press secretary, providing post-hoc justifications for our intuitive judgments, he also says (in another of his favorite metaphors) that our conscious reason is like a rider on the back of the elephant of intuition. When the elephant wants to go somewhere, the rider has little hope of steering it elsewhere. But over time, the rider might be able to *train* the elephant.[39] Haidt has long been a champion of cognitive behavioral therapy as a means of training the elephant, which rests on the assumption that people are capable of reflectively challenging their background beliefs and thus changing their feelings and behavior.[40]

Haidt's concession that reason can sometimes influence our intuitions might seem to rescue the rationalist. Perhaps, a rationalist might say, Haidt's empirical challenge to moral philosophy does not cut quite as deep as it initially seems to. Haidt shows us how *difficult* it is for reason to govern our behavior, but he does not say it is impossible. Can we not accept Haidt's warnings about the obstacles facing moral reasoning while still aiming to establish the truth by way of argument?[41] Is this not just what Socrates wanted us to do when he pressed us to examine our lives?

But Haidt's objection unfortunately cannot be so easily absorbed, because he denies that reason is *ever* capable of establishing moral truth. According to Haidt, what reason seeks is not moral truth but social alliances. "Morality," to put it another way, is not finally a kind of knowledge at all, as at least twentieth-century moral rationalism tends to assume,[42] but a form of social practice. Ancient Greek rationalism is arguably more capable of absorbing this challenge than modern rationalism, since in that context evaluative judgments were more closely tied to social performances than they later became.[43] Yet Haidt would press it further even than they do. Haidt seems to disagree that moral reasoning can do anything to clarify what true happiness consists in,

or what passions are and are not unhealthy. Perhaps, then, despite the allowances Socrates and Aristotle made for the power of the passions, it was not enough. Perhaps their confidence in the normative power of reason itself needs to be abandoned.

On this point it may be that another ancient thinker, the fourth- and fifth-century theologian Augustine of Hippo, proves a more promising conversation partner for Haidt. Augustine has often been read as a sort of voluntarist in his moral psychology—that is, as holding that human beings are possessed of a capacity for free judgment, a *will* (*voluntas*), that is undetermined by our beliefs and desires. It has even been suggested that Augustine *invented* the idea of the will, and thus the very possibility of voluntarism.[44] On this interpretation, Augustine does not seem like a helpful interlocutor at all. But Augustine's view turns out to be significantly more complicated.

There is some basis for the voluntarist interpretation in Augustine's early work, and especially in *De libero arbitrio* (begun in 388 CE and finished in the early 390s).[45] In *De libero arbitrio*, Augustine is trying to resist the idea that God is responsible for evil, and so instead he traces evil to each individual who performs it.[46] What makes our actions evil is desire, he says, and the only thing that can cause us to act according to desire rather than right reason is our own will and free choice.[47] He insists that it is wrongheaded to try to look behind the will to some other cause.[48] That can give the impression—especially given the way that this conversation developed in later centuries—that "the will" is some sort of undetermined, autonomous power in the person.[49]

But that impression is only partly right, even in this early work. As Han-luen Kantzer Komline explains in her meticulously comprehensive study of Augustine on the will, the young Augustine did understand the will as a hinge capable of swinging either towards the good or away from it by its own power.[50] That idea, taken alone, certainly sounds voluntarist, and does reflect a deep commitment to the view that one's will is fully within one's own power. "What is so much in the power of the will," he asks, "as the will itself?"[51] But actually, in *De libero arbitrio*, Augustine is better understood as a straightforward rationalist. Vice is at root a kind of ignorance, he says.[52] The mind is more powerful than desire and rightly rules over it.[53] If the mind does not rule us, that can only be because the mind has *willingly* become a slave of desire. Whether or not we avoid evil—whether or not we have a good will—is therefore entirely a matter of whether our mind chooses to exercise its rightful authority.

This is as close as we are likely to come to the sort of delusion that Haidt ascribes to the entire tradition of rationalist moral philosophy. But Augustine, like Haidt, soon comes to judge this sort of rationalism as a delusion. Augustine's rationalist psychology is displaced, in the

mid-390s, by what historian Peter Brown famously called a "psychology of delight."[54] We are moved, Augustine comes to think, by what delights us. And crucially, Augustine comes to think that *we have no control over what delights us*. I cannot suddenly force myself to take delight in something previously found uninteresting or even repulsive just because I have decided that I ought to feel differently about it. As Augustine makes clear in the early books of the *Confessions*, he knows very well where his own delights came from: they came from the books he was made to read as a child, from the friends who wanted to steal pears from a tree, from the rhetors he emulated. The sources of his delights were *social*. And when those delights were transformed, the sources of the transformation were also social: the beauty and piercing intelligence of Ambrose's preaching, the stories friends told him about the saints, the persistence of his mother.

The main impetus for the sea-change in Augustine's moral psychology was theological: he wanted to insist that it is God, not I, who makes it possible for me to do good. Augustine's former, more nearly Stoic view is that the shape and direction of our *voluntas* is in our own power. Recall again the line from *De libero arbitrio*: "what is more in the will's power than the will itself?" But in *Ad Simplicianum*, the first work he penned as a bishop (in 396 CE), that old view crumbles.[55] First, he comes to see, under Paul's influence, that knowledge is not enough to change behavior. The law offers knowledge of the good but not the ability to do it. He says in the first question that "what remains to free choice in this mortal life" is not that we might actually do the just thing, but only "that one might turn to the one by whose gift justice might be fulfilled."[56] He thus takes the scope of our self-command and whittles it down to the ability to ask God's help. In the second question, he takes even that ability away. Now he maintains that we require God's grace even to *ask* for God's grace. We do not have the power, even if confronted by clear knowledge of what we ought to do (which he the biblical Torah supplies), even to *wish* we could will the good. To wish for it would be to find some delight in the idea of doing it, and our delights are not under our control. It remains true, Augustine says, that "wills are chosen." But the will is only moved to choose something if "something happens that delights and invites the soul—and it is not in a person's power to make such a thing happen."[57] "'Feeling,'" Peter Brown says, "has taken its rightful place as the ally of the intellect"[58]—or perhaps has overtaken it. By the time Augustine penned his first works against the Pelagians a decade and a half later, this had become a fundamental feature of his moral psychology: what his opponents failed to recognize, he said, was that to do the right thing we need more than the knowledge of what to do and the freedom to

do it; we also need the *desire* to do it, and the desire to do it is not something we can supply ourselves.[59]

So Augustine abandons the aspirational rationalism he had espoused in *De libero arbitrio* and moves in a direction that a "social intuitionist" like Haidt would recognize and appreciate. Augustine comes to recognize that human beings are moved to action, not by their "intuitions" exactly, but by their delights, their desires, their loves.[60] What Augustine calls "delight" is of course not identical to what Haidt calls "intuition," but they are both names for immediate and affect-laden evaluative judgments. Augustine comes to recognize too that we have strikingly little direct control over what we take delight in, and that instead it is our *relationships* that give shape to our delights over time. Augustine remains a rationalist in that he maintains the classical view that our affections are themselves expressions of implicit beliefs about the world.[61] He also remains a rationalist in Haidt's sense that reason is a better guide than intuition to true moral knowledge. But his mature view is that our capacity for free choice, far form being a locus of rational self-mastery, is in fact "derivative and has no independent power."[62] Our ability to reflexively evaluate our own motivations and actions, fundamental as it is to our psychology, is not what drives our action. It follows on and is bounded by our deeper motivational structure, and that whole motivational structure is what Augustine calls *voluntas*. Our loves, not our reason, are what gives our will its basic shape.

But Augustine might seem to make our work as ethics teachers even more pointless than Haidt does. For Augustine denies even that conscious reason can be counted on to "retrain the elephant." Consider Augustine's famous conversion experience in the garden in Book 8 of the *Confessions*. Augustine recognizes in himself a desire to join the church, and he wishes desperately that he could bring himself to act on that desire. But he also recognizes in himself an even stronger desire—so strong he experiences it as a binding necessity[63]—to stick with the marriage that promises him both social stability and an outlet for his sexual appetite. He recognizes "two wills" in himself—*duae voluntates meae*.[64] Although he reflexively prefers one over the other, try as he might, he can do nothing to force himself to do what he prefers. It is similar to the experience Aristotle called *akrasia*, except that Augustine's "divided will" is not a conflict between *boulesis* and *epithumia*, reason and appetite, but between two competing motivational orientations, both with their own relative integrity.[65] His efforts to reason away his inner conflict fail terribly, and the efforts of his friends fare no better. They fail despite the fact that they appeal to the elephant by telling stories about people Augustine admires. Those stories do succeed in increasing Augustine's desire to make a break with his old life, but they do not actually enable him to do it. The el-

ephant cannot be steered even with such tricks as these. The only real consequence of his friends' efforts is to make Augustine hate himself for being unable to do what he so admires in others.[66]

What Augustine remembers as resolving the conflict is not force of will, not a knock-down logical argument, nor even social-emotional appeals, but chance encounters with a child (*tolle, lege*) and a book ("put on the Lord Jesus Christ") that mysteriously and unpredictably re-arrange his desires.[67] We can hardly hope to orchestrate such encounters in our ethics classrooms. Augustine admits that this re-arrangement wasn't something he did or could have orchestrated on his own. It was luck—or rather, as he calls it, grace.

Augustine would not say that it is *always* impossible to actively train or re-train my desires. I doubt he would even say that it is impossible to retrain them by thinking. But an Augustinian will perhaps be even more pessimistic about this possibility than Haidt is. Haidt seems to believe that if our riders are psychologically sophisticated enough—wise enough in the ways of elephants—we can nudge our intuitions in just about any direction we want. (It is hard not to suspect that there is a version of the rationalist delusion still lurking in his theory, if now of a more behavioralist sort.) Perhaps we can sometimes re-train our desires. But we cannot do so *consistently* or *reliably*. We must at least admit that. Nor can we count on our reason even to establish what we should desire, for reason itself is beholden to our loves. If for the ancient Greeks, philosophy was the only sure medicine for the illness of bad desire, Augustine does not trust even this medicine.

The Modest Goods of Moral Reasoning

This brief tour through ancient moral psychology has been meant to show, first, that there is much more common ground between the philosophers' "rationalism" and Haidt's "intuitionism" than one might expect. Socrates, Aristotle, and Augustine all deny that reason alone, apart from our passions or desires, can or should move us to act. All of them recognize that more "intuitive" processes (i.e., automatic, affect-laden evaluations) underlie and shape the things we think about and the way we think about them. All of them recognize, moreover, that those intuitive processes are socially-formed. We are ill-served, I want to suggest, by dismissive dichotomies between rationalism and intuitionism. The dividing lines are not so stark.

But my goal is not merely historical or typological. It is also practical. I have been trying to set Socrates, Aristotle, and Augustine before us as alternative models for teaching moral reasoning. What makes these figures especially pedagogically interesting is that although they agree with Haidt on those psychological points, *they go on teaching*

moral reasoning anyway. They continue to think it is important to teach moral reasoning. Their view of what moral reasoning can do is considerably more modest than the view of the rationalists Haidt opposes. But a modest view of moral reasoning can still be a constructive view. Recognizing that these rationalists are not as psychologically naive as they are sometimes taken to be opens up the possibility of learning something from them about what it might look like to teach moral reasoning well.

For those of us who teach moral reasoning, it is important to be clear about what the modest goods of moral reasoning are. Admittedly, the modest goods I want to focus on here might seem far *too* modest to some rationalists. Some rationalists will no doubt want to defend, against Haidt, the notion that sound logic can reliably circumvent the cognitive biases that the passions can bring. Some rationalists will want to defend the idea that we are sometimes able to act rationally in the face of our passions. I have not tried to offer an argument against such rationalists in this paper. Instead, by showing that some rationalists *grant* those points and yet remain rationalists, I hope to broaden our sense of what teaching moral philosophy might do.

So let us grant for the moment that we should avoid suggesting to our students that they should set their emotions aside and act on pure reason; let us grant that this is a "rationalist delusion." What then should we tell students that moral reasoning is for? What role do Socrates, Aristotle, and Augustine preserve for it? I will focus on three goods: the good of communal reflection, the good of criticizing intuitions, and the good of confessing ignorance. The first fits well with Haidt's social intuitionism; the second two challenge or develop his framework.

Communal Reflection

Haidt agrees, as we have already seen, that moral reasoning (understood as slow, strategic reasoning) has a constructive role to play in our moral lives. He believes that such reasoning is typically undertaken to justify our intuitions to others; very rarely, if ever, does our reasoning lead us to modify our own intuitions. But when we engage in conversation with others, sometimes other people's reasoning *does* cause our intuitions to change.

Socrates, Aristotle, and Augustine would agree with him on this. They may disagree, however, on what to make of the social character of moral reasoning. Haidt suggests that moral reasoning is social mainly because it is something we do to justify our intuitions to others, and thus to build alliances. He thus sees moral reasoning as good mainly insofar as it aims at social cohesion. But there might be other goods worth keeping in mind.

The good of communal reflection is clearest in Socrates, who famously practiced his philosophy through actual in-person dialogues, usually with others but occasionally with himself (in the *Crito* and the *Symposium*, for example, he produces a dialogue partner out of his own imagination or memory and thus examines himself). Those dialogues almost always arose from and were demanded by a concrete question about responsibilities. What does Euthyphro owe his father? What does he owe the slave who died by his father's neglect? What should Hippocrates expect of a teacher, and what should Protagoras promise his students? The point of these dialogues was not, however, to create more social cohesion by winning people over to his intuitions, as Haidt might suspect. He provoked his conversation in the agora because he thought it was good in itself to be thoughtful and self-aware about what we are doing to whom and why, and because he recognized, like Haidt, that I can reach that goal better by thinking with others than by thinking alone. Conversation with others was a means to self-knowledge.

We must admit—and we should press our students to consider—this sort of dialogical self-examination does not always yield "moral truth." It may be that the person I am talking to is a bully or a demagogue or simply mistaken. I might end up more wrong than I was when I began. (Again some rationalists will feel frustrated, since reason is supposed to be able to provide us a means out of such a situation. Again I ask: grant for a moment that it cannot, or at least cannot always.) But we might still say that such dialogical self-examination is a *prerequisite* to truth. Socrates is never quite sure if he has the right answer, but he knows that the only way he can even hope to get a right answer is by examining himself and others, and by pressing others to examine him in turn. And what's more, we might say that the experience of communal reflection is itself a good, apart from its product.

In the classroom, we should therefore prioritize actually giving students reason and time to speak with one another. If the goal is to give students experience examining themselves and others, there is no substitute for real conversation. But both reason and time for conversation are difficult to come by, so giving those things to students requires real work. Socrates gave his interlocutors reason to speak by asking them questions about the things they were doing right at that moment (going to court, choosing a teacher, and so on). That is an option for us too, though it requires knowing something about our students' lives and their being willing to discuss those things publicly. Somewhat more accessible, given our own mores, is the Aristotelian approach: pointing out puzzles in received wisdom and asking students to try to resolve them. The difficulty of finding time, on the other hand, is more a product of our own pedagogical expectations than anything else. Prioritizing conversation means deprioritizing lectures, deprioritizing

"getting through the readings," deprioritizing a tidy and testable finish line. It means letting go of a certain amount of control over what the students "get out" of a class session.

Criticizing Intuitions

Moral reasoning also offers a particular way of engaging with our own intuitions—namely, self-critically. I have argued that these ancient rationalists would have granted Haidt that "intuitions" were a fantastically powerful force in our moral psychology, far more powerful as a rule than rational deliberation. But they also insisted that these intuitions were typically *bad*. They were bad in the sense that acting on them would undercut our own flourishing. So they encouraged an essentially self-critical posture towards our own intuitions. Haidt certainly recognizes that our intuitive judgments are often bad for us, which is why people need tools like cognitive behavioral therapy to try to correct them. But, since this is a normative rather than empirical question, he does not dwell on it. The ancients do dwell on it.

I mentioned earlier that Socrates' mode of questioning *seems* to imply that he thinks clear and certain knowledge of virtue is a prerequisite to actually acting virtuously, but I suspect that this is a misunderstanding of Socrates. At least in the *Apology*, Socrates says he knows only that he is ignorant, and he wants others to recognize the same thing about themselves. His questioning is designed to produce the experience of *aporia*, of puzzlement. His use of dialectic is not meant to secure an intuition (as Haidt has it) but to destabilize it, in hopes that we might be able to curb their destructiveness.

Augustine is an even clearer example of this self-critical posture. Augustine was deeply, painfully aware of all the ways that his socialization had trained him to desire some things more than others, and how his desires often kept him from thinking clearly about what would really be good for him or for those around him. After Paul convinced him to break with the naive rationalism of his younger days, he no longer held out any hope that he could think his way to better desires. But that did not mean he gave up thinking. Nor did his recognition that our relationships largely govern the direction of our moral lives lead him to believe that the options were only to manipulate or not to manipulate. Instead, he set to work reflecting on his relationships. Contemporary philosopher Margaret Urban Walker has suggested that "we have an urgent need for *geographies of responsibility*"—[68]that is, a clearer map of who we usually assume is meant to do what for whom, and when, and why. This is exactly what Augustine sets out to produce for himself in his *Confessions*. In drawing his personal map, he becomes more aware of *why* his desires, his "intuitions," are what they are. As Haidt rightly insists, it is our relationships that shape our

intuitions over time. By becoming more aware of the social causes of my intuitions, I might gain some fresh perspective on them. Some I might end up trusting more; others I might end up trusting less.

Becoming self-critical about my intuitions is no guarantee that I will end up with right ones. I might end up challenging the wrong intuitions and becoming worse than I was. But holding dogmatically to my intuitions is a pretty good guarantee that I will end up wrong. By encouraging students (and ourselves) to develop a critical self-awareness about our intuitions and their social sources, we are creating a space for their transformation.

In one of my classes recently, after we had spent some time reading and discussing ancient Jewish wisdom literature, I gave the students this exam question: "The book of Proverbs paints a picture of the wise person and the foolish person. Which are you?" The goal of the question was turn the students' critical attention back on themselves, even as they continued to wrestle with the text we were studying. Ethics teachers are often reticent to ask questions that require students to put their own views, or certainly their own lives, on the dissecting table. Students are understandably reticent to answer such questions. But if a self-critical posture is one of the central goods of moral reasoning, we cannot escape actually asking students to adopt that posture.

Confessing Ignorance

Haidt has been especially interested in cases where people have clear and strong intuitions about what to do. He helped to run a series of experiments, for example, designed to produce what he calls "moral dumbfounding"—experiences in which I have a strong intuition about something, like the badness of incest or of flag-burning, that I cannot justify and that seems impervious to contrary reasons.[69] People's responses to these situations illustrate, he thinks, the primacy of intuitive judgment; we cling to our intuitions even as we freely admit we have no reason to maintain them. I suspect there are other ways of accounting for such experiences that would be more consistent with a rationalist framework, especially if you hold, as we have seen that Socrates, Aristotle, and Augustine did, that the emotions give expression to deeply held beliefs.[70] But here I want to emphasize a different point: *our intuitions are not normally so strong*. Haidt knows this, of course; his own experiments showed that only certain kinds of stories were able to produce the experience of dumbfounding. But as before, the ancient rationalists saw a different normative significance in this than Haidt does.

Aristotle famously said that all philosophy begins with wonder, by which he meant that it begins by recognizing one's own ignorance. "The person who is puzzled and amazed considers himself to know nothing"

(*Meta*. 982b). In all his philosophy, including his moral philosophy, he thus begins by noticing puzzles or contradictions in common opinion and setting out to resolve them. In this he follows the lead of his teacher's teacher, Socrates, who claimed to know only that he knew nothing. The oracle at Delphi called him the wisest of everyone, he insisted, only because he understood that his wisdom was worthless (*Apol.* 23a). "Surely it is the most blameworthy ignorance to believe that one knows what one does not know" (*Apol*. 29b). And so he goes around examining those who think they know something precisely in order to show them that they do not.

The point of showing people that they know nothing is not just to undermine their credibility, though Socrates' opponents clearly experienced it that way. A frank confession of one's own ignorance is the first part of taking care of oneself and others. Socrates believes that those who go around thinking they know what they do not know—those, we might say, who cling fiercely to the truth of their intuitions—end up doing tremendous damage to their own souls and to the whole city. Those who have brought him before the court, Anytus and Meletus, think that by putting Socrates to death they are protecting the children and protecting all of Athens; they are so confident that they know what children need, what it takes to educate them for the good, what the gods require, that they willing to kill for it. In fact, Socrates says, their own testimony shows that they are only playing at seriousness (*Apol.* 24c). What they are really trying to do is to avoid giving an account of their own lives (*Apol.* 39c). A strong dose of epistemic humility would have served them all far better.

So in response to Haidt's challenge, we ethics teachers might finally say that when we teach moral reasoning, we are (or should be) teaching students the skill of recognizing what they do not know. Just as we press them to be self-critical about their intuitions, we should also press them to be self-critical about their claims to knowledge. It is pedagogically useful in this regard to imitate Haidt's experiments and produce instances of moral dumbfounding for our students directly. Those experiences force students to confront the disconnect between their "moral knowledge" and their moral reasoning. It is more difficult, though, to help them see that their reasoning often fails even in those situations where it seems to line up with their intuitions. One strategy I have found useful is to give students examples of twisted moral justifications from real people—mass shooter Elliot Rodger's manifesto justifying his "war on women,"[71] theologian John Howard Yoder's rationale for his abusive "sexual experiments"[72]—and ask them to read and discuss them. Counterintuitively, as Augustine came to see, moral reasoning can be perhaps most useful in undermining its own delusive claims to certainty and finality.

Conclusion

Perhaps, then, Jonathan Haidt's social intuitionism is less opposed to the rationalist philosophical tradition than it might seem, and less a challenge to the work of ethicists than might be feared. Haidt is right, I think, that ethics teachers ought to give up the boast of Protagoras: that by teaching students to deliberate well, we might send them home better people every day. But even if we grant that people typically act on the basis of the "fast reason" of intuitive judgments rather than "slow" reason of linear thought, slow reason might still have some important role to play in our moral lives beyond self-justification. We might even think, without being empirically naive, that moral reasoning has some advantage over our passions in making our lives *good*. On the view I have tried to draw from Socrates, Aristotle, and Augustine, reasoning helps us not mainly by settling on the right answer and then commanding our passions to follow it, but by providing us a way to notice, in conversation with others, that neither our intuitions nor our reason can claim to have the right answers. If we can make our classrooms spaces of honest confusion and vulnerability, we give students an opportunity to confront our ignorance and perplexity together with others. This experience itself might even prove capable of forming our passions, though with Augustine, I suspect we must admit that exactly how our passions will be formed is beyond our control. We will have to wait on luck, or grace, to stumble on the truth.

Socrates believed that the greatest good for human beings is to talk with others every day about virtue (*Apol.* 38a). This is, happily, what ethics teachers get to do—if we are willing to accept the challenge. He believed it was the greatest good not because he thought we might settle what virtue finally looks like, but because the dialogue itself forms people in a certain way. For Socrates, Pierre Hadot has argued, dialogue was a sort of "spiritual exercise practiced in common."[73] To enter into dialogue with oneself or with others is integral to wisdom and thus to happiness—not because it *secures something for us* but because it *makes us into a certain kind of people*. What counts in Socratic dialogue, Hadot says, "is not the solution to a particular problem but the path traveled to get there, a path on which the interlocutor, the disciple, the reader, forms their thought, renders it more fit to discover on its own the truth."[74] Dialogue forms the thought of the people involved not by giving them the right answers they can then use to command themselves, as Haidt seems to suppose, but by giving them opportunity to realize what they do *not* know and to seek the truth of it. Through dialogue they have the opportunity to become not sophists, who think they already possess the truth, but true philosophers, who love truth and long for it, knowing that they do not have it.

Notes

1. Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon Books, 2012), 88–91. Some of the core insights of Hadit's "social intuitionist" theory of moral judgment were anticipated in his important article, Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (October 2001): 814–34. The main arguments of *The Righteous Mind* were reprised more briefly in Jonathan Haidt, "Moral Psychology for the Twenty-First Century," *Journal of Moral Education* 42, no. 3 (September 2013): 281–97.

2. Haidt, *The Righteous Mind*, 2012, 28.

3. Haidt, 74.

4. Cited in Haidt, *The Righteous Mind*, 2012, 25.

5. Haidt, 78–81.

6. Haidt, 46–48.

7. Haidt, 73–74.

8. Not, of course, that psychologists always get it right, either. On the contrary, Haidt's critique is aimed especially at the cognitivist or rationalist biases among other psychologists, represented paradigmatically in moral psychology by Lawrence Kohlberg. See Haidt, "Moral Psychology for the Twenty-First Century," September 2013, 282–283.

9. *Protagoras* 318a–b. All translations of Plato are my own.

10. *Protagoras* 319a.

11. Although I am focusing on Haidt's work in this essay, he is certainly not unique in, from my perspective, underestimating the value of pre-modern moral psychologies. On the contrary, Haidt gives older philosophical and theological traditions far more attention and credit than psychologists usually do. Though I think his interpretations of those figures are sometimes mistaken, I am nonetheless grateful that a psychologist has even taken the time to read them closely.

12. In Haidt, "The Emotional Dog and Its Rational Tail", Haidt defines a moral intuition as "the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion" (818).

13. See Thomas C. Brickhouse and Nicholas D. Smith, *Socratic Moral Psychology* (New York: Cambridge University Press, 2010) for a good overview of the scholarly debates over what exactly Socrates meant by this.

14. But I agree with Alexander Nehamas, *Virtues of Authenticity: Essays on Plato and Socrates* (Princeton, NJ: Princeton University Press, 1999), ch. 2, that "Socrates' belief in the priority of definition is much less radical than we have often tended to suppose."

15. Plato, *Republic* 441e.

16. Aristotle, *Nicomachean Ethics* [EN] 1098a. All translations of Aristotle are my own.

17. Martha C. Nussbaum, *The Therapy of Desire* (Princeton, NJ: Princeton University Press, 1994), 34.

18. Haidt, *The Righteous Mind*, 2012, 7.

19. Aristotle, EN 1106b–1107a.

20. Haidt, "Moral Psychology for the Twenty-First Century," September 2013, 288. Aristotle, in EN 1111b, says that *prohairesis* is "most proper to virtue." See Haidt, 288, for the "where the action is" language.

21. Aristotle himself is in some ways a critic of Platonic rationalism, inasmuch as Aristotle believes that knowledge always begins from "things known to us" (EN 1095b)—in this case, practically, from common beliefs about happiness and virtue that seem to have some argument for them—and not from some universal and independent "form of the good" (EN 1096a). But in the broad sense that Haidt gives to rationalism, Aristotle is no less a rationalist than Plato is. What Aristotle is criticizing in those opening pages of the *Nicomachean Ethics* is not Plato's rationalism per se but his belief that the goodness of various actions is reducible to some single property.

22. EN 1179b29.

23. See Nussbaum, *The Therapy of Desire*, 54–56, for a sketch of the selection criteria governing admission to Aristotle's classroom.

24. Nussbaum, 58.

25. Wouter Sanderse, "Does Aristotle Believe That Habituation Is Only for Children?" *Journal of Moral Education* 49, no. 1 (January 2020): 98–110.

26. Terence Irwin, *The Development of Ethics*, v. 1, §§87–93.

27. Irwin, *Development of Ethics*, 167.

28. Thus Irwin: "Though Aristotle's description of the early stages of habituation emphasizes—not surprisingly—children's training in pleasure and pain, he does not take this training to be all that is needed for virtue of character. At the later stages of habituation, it is up to agents to elect different kinds of action" (*Development of Ethics*, 181–182).

29. On the role of non-rational means of moral suasion for Socrates, see for example Jessica Moss, "The Doctor and the Pastry Chef: Pleasure and Persuasion in Plato's *Gorgias*," *Ancient Philosophy* 27 (2007): 229–49, and Paul Woodruff, "Socrates and the Irrational," in *Reason and Religion in Socratic Philosophy*, ed. Nicholas D. Smith and Paul Woodruff (London: Routledge, 2000), 130–50. For a narrower rationalist interpretation of Socrates, see Terry Penner, "Socrates," in *The Cambridge History of Greek and Roman Political Thought*, ed. Christopher J. Rowe and Malcolm Schofield (Cambridge: Cambridge University Press, 2000), 164–89.

30. Nussbaum, *The Therapy of Desire*, 80. In Martha C. Nussbaum, *Upheavals of Thought: The Intelligence of Emotions* (New York: Cambridge University Press, 2003), Nussbaum explores this "cognitive-evaluative" view of emotions in a more constructive mode.

31. This is what Socrates chides the jurors about in his *Apology* (29a): they assume he should be desperately afraid of death, but if they would admit their ignorance about whether death is in fact bad for us, they might not think it worth fearing.

32. Nussbaum, *The Therapy of Desire*, 80–81.

33. Haidt, *The Righteous Mind*, 2012, 114–16.

34. Haidt, *The Righteous Mind*, 2012, 271. His definition: "Moral systems are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and

evolved psychological mechanisms that work together to suppress or regulate self-interest and make cooperate societies possible" (270).

35. Haidt, 272.

36. Haidt, 272.

37. Haidt, *The Righteous Mind*, 2012, ch. 7.

38. Haidt, 30.

39. Haidt draws this metaphor both from the Buddha and from Plato, both rationalists, and the image sits uncomfortably, it seems to me, with his rejection of the reason-as-ruler paradigm.

40. In Jonathan Haidt, *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom* (New York: Basic Books, 2006), ch. 2, he introduces cognitive behavioral therapy as one of three ways (alongside meditation and Prozac) to change your "affective style." More recently, Greg Lukianoff and Jonathan Haidt, *The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure* (New York City: Penguin Books, 2019) refers continuously to cognitive behavior therapy as the main cure for the youth mental health crisis, even including a how-to guide as an appendix.

41. My thanks to an anonymous reviewer for pressing this point.

42. See Margaret Urban Walker, *Moral Understandings: A Feminist Study in Ethics* (New York: Routledge, 1998), 36–39, for an argument to this effect.

43. Alasdair MacIntyre suggests, for example, that in Homer "a man is ἀγαθός ['good'] if he has the ἀρετή ['virtue'] of his particular and specific function" (Alasdair MacIntyre, *A Short History of Ethics* (New York: Macmillan, 1966), 8). To "know" the "good" in that case is just to know what characterizes a particular social role—a wife, a king, a soldier. It does not require special access to any detached moral truths.

44. Hannah Arendt, *The Life of the Mind* (New York: Harcourt, 1981), pt. 2, p. 84, for example, famously calls Augustine "the first philosopher of the Will." Albrecht Dihle, *The Theory of Will in Classical Antiquity* (Berkeley, CA: University of California Press, 1982), 144, called Augustine "the inventor of our modern notion of will."

45. Not only Arendt but also more sensitive genealogists like Michael Frede rely heavily on this book in their treatment of Augustine's moral psychology. Michael Frede, *A Free Will: Origins of the Notion in Ancient Thought*, ed. Anthony A. Long (Berkeley, CA: Univ. of California Press, 2011) calls *De libero arbitrio* "the authoritative exposition of his view" of the will (159). It is worth noting too that Frede, responding mainly to Dihle, disagrees that Augustine's view of the will was innovative and suggests that the source of the modern idea of will is considerably older. Augustine's "view of the will is pervasively Stoic," he thinks, "but embedded in a Platonist notion of the world" (155).

46. *De libero arbitrio* 1.1. All translations of Augustine are my own, based on the Latin of the *Patrologia Latina*, but a convenient English translation can be found in Augustine, *On Free Choice of the Will*, trans. Thomas Williams (Indianapolis, IN: Hackett, 1993).

47. *De lib. arb.* 1.11.

48. *De lib. arb.* 3.17.

49. Risto Saarinen, *Weakness of the Will in Medieval Thought: From Augustine to Buridan* (New York: Brill, 1994) shows how Augustine's insistence that all sin is voluntary gives rise, at least as early as Anselm, to a stronger notion that no one ever acts against

their own will (see 45–48) and later, in Peter of Poitiers, to a full-fledged voluntarism (see 65–66).

50. Han-luen Kantzer Komline, *Augustine on the Will: A Theological Account* (New York: Oxford University Press, 2020), 28–31.

51. *De lib. arb.* 1.12.

52. *De lib. arb.* 1.1.

53. *De lib. arb.* 1.10.

54. Peter Brown, *Augustine of Hippo: A Biography*, new ed. (Berkeley, CA: University of California Press, 2000), 148.

55. Paula Fredriksen, "Beyond the Body/Soul Dichotomy: Augustine on Paul Against the Manichees and the Pelagians," *Recherches Augustiniennes* 23 (1988): 84–114 says that "Augustine's views change more drastically between 394 and 396/8 that between 398 and 430" (89), leading particularly to his abandonment of a classical view of "moral autonomy" (95). Komline, *Augustine on the Will*, more cautiously, calls *Ad Simplicianum* a "benchmark work" (62) in which Augustine first develops a "more chastened view of the condition of the fallen human will" (88).

56. *Ad Simpl.* 1.14.

57. *Ad Simpl.* 2.22.

58. Brown, *Augustine of Hippo*, 148

59. *De spiritu et littera*, paras. 6–20.

60. These terms are not entirely interchangeable for Augustine, as they tend to have different connotations. See Jesse Couenhoven, "Augustine's Moral Psychology," *Augustinian Studies* 48, no. 1 (2017): 23–44, 28. But Augustine is not fussy about distinguishing them, and they all point to the same basic motivational structure.

61. Couenhoven, 33.

62. Couenhoven, 39.

63. Komline, *Augustine on the Will*, ch. 2, argues that if the dominant metaphor of Augustine's early thinking about the will is that of the hinge, capable of swinging towards or away from the good, he comes to think of the *fallen* will as a chain that binds him to sin.

64. *Conf.* 8.10.

65. Aristotle's discussion of akrasia is in EN 7.1–10. Couenhoven, "Augustine's Moral Psychology", 36, puts it helpfully: "the most fundamental internal division of which Augustine speaks is not hierarchical [i.e., between higher and lower faculties] but horizontal, a fracturing of the mind between good and evil."

66. *Conf.* 8.17.

67. *Conf.* 8.29

68. Walker, *Moral Understandings*, 99.

69. Jonathan Haidt, Fredrik Björklund, and Scott Murphy, "Moral Dumbfounding: When Intuition Finds No Reason" (Unpublished Manuscript, 2000); Haidt, "The Emotional Dog and Its Rational Tail," 817; Haidt, *The Righteous Mind*, 2012, 36–40; Jesse Prinz, *The Emotional Construction of Morals* (New York: Oxford University Press, 2009), 29–32.

70. See Matthew L. Stanley, Siyuan Yin, and Walter Sinnott-Armstrong, "A Reason-Based Explanation for Moral Dumbfounding," *Judgment and Decision Making* 14, no. 2 (2019): 120–29, for an experimental attempt to offer such an alternative explanation.

71. Ian Lovett and Adam Nagourney, "Video Rant, then Deadly Rampage in California Town," *New York Times*, May 24, 2014, https://www.nytimes.com/2014/05/25/us/california-drive-by-shooting.html.

72. Rachel Waltner Goossen, "'Defanging the Beast': Mennonite Responses to John Howard Yoder's Sexual Abuse," *Mennonite Quarterly Review* 89 (January 2015): 7-80.

73. Pierre Hadot, *Exercices spirituels et philosophie antique*, nouvelle édition revue et augmentée, 41 (Paris: Michel, 2002), 31; translation mine. Hadot holds that this conception of dialogue was not uniquely Socratic, but widely shared in ancient philosophy.

74. Hadot, 35.

Brian Hamilton is Associate Professor of Religion at Florida Southern College. brian@bdhamilton.com.