

**TEACHING APPLIED ETHICS TO THE RIGHTEOUS MIND**

PETER MURPHY  
UNIVERSITY OF INDIANAPOLIS  
murphyp@uindy.edu

**ABSTRACT**

What does current empirically-informed moral psychology imply about the goals that can be realistically achieved in college-level applied ethics courses? This paper takes up this question from the vantage point of Jonathan Haidt's Social Intuitionist Model of human moral judgment. I summarize Haidt's model, and then consider a variety of pedagogical goals. I begin with two of the loftiest goals of ethics education, and argue that neither is within realistic reach if Haidt's model is correct. I then look at three goals that can be achieved if his model is correct; but each of these goals, I argue, lacks significant value. I end by identifying three goals that are of significant value and also realistically attainable on Haidt's model. These should be the focus of applied ethics pedagogy if Haidt's model is correct.

People who teach college-level applied ethics courses should design their courses around goals that are both realistically attainable and significantly valuable. Setting goals that are not realistically attainable is futile, and setting goals that are not significantly valuable is a poor use of resources. What goals then are both realistically attainable and significantly valuable for a college-level applied ethics course? Where, in other words, should people who teach these courses put their energy?

Consider what goals are realistically attainable. This is constrained by the nature of human minds. What can likely be achieved in a classroom of humans is not the same as what can likely be achieved in a classroom of chimpanzees or a classroom of aliens. This is enough to show that moral psychology needs to inform pedagogical strategies. In particular, knowing about the mental processes that produce moral judgments has implications for how applied ethics courses should be taught.

However, the nature of these processes is the subject of intense debate among psychologists and philosophers who work on this topic. Given this lack of consensus, one strategy is to proceed piecemeal and look at each of the leading theories of these processes and think through the pedagogical implications of each. Using this strategy, I will focus here on just one theory of these processes: Jonathan Haidt's Social Intuitionist Model. This model of human moral judgment is of special interest because it offers a particularly discouraging and gloomy view of human moral judgment. This then is an opportunity to think about what some of the worst news coming out of moral psychology might imply for how to teach applied ethics courses. Haidt's model, as we will see, clearly implies that some of the most familiar pedagogical goals of applied ethics courses are not realistically attainable. The challenge I take up is to determine whether there are some goals that are both sufficiently valuable and realistically attainable that applied ethics courses can be organized around should Haidt's model turn out to be correct.

The paper has five sections. In the first, I review Haidt's Social Intuitionist Model, and some of the support he offers for it. In the second, I argue that if this model is correct, then two of the loftiest goals of ethics education are not realistically attainable. In the third, I look at three other goals. Though each of these goals is within realistic reach if Haidt's model is correct, I argue that they are of little value. In the fourth section, I consider three other goals whose value lies between the lofty goals of the second section and the insufficiently valuable goals of the third section. I argue that these modest goals are of significant value, and that they can be achieved if

Haidt's model is correct. This supports my central contention that these goals should be the focus of college-level ethics courses if Haidt's model is correct. In the fifth section, I take up an important disappointment that some might feel in moving from a traditional applied ethics course that is organized around analyzing and evaluating arguments to an applied ethics course that is organized around these last three goals.

Here are a few things to keep in mind before proceeding further. First, I will not defend Haidt's model of the moral mind. For important criticisms of his model, and Haidt's various attempts to support it, see Fine (2006), Narvaez (2008), Liao (2011), Jacobson (2012), Haste (2013), Fry and Souillac (2013), and Maxwell and Beaulac (2013). Here I am going to simply assume that Haidt's model is correct, and focus on its implications for teaching applied ethics. If his model isn't the whole truth about the moral psychologies of students, then nothing that I say here will tell us what the appropriate goals of ethics education are. To make such a determination, we would need to know exactly where Haidt's model goes wrong and what the correct account of human moral psychology says instead.

Second, my focus will be quite specific. I will only consider college-level courses with a substantial applied ethics unit (for Haidt's views about early moral education of children, see Graham, Haidt, and Rimm-Kaufman, 2008). Such courses often include a unit covering standard ethical theories, including virtue ethics, feminist views, Kantian deontology, consequentialism (including utilitarianism), ethical relativism, and other ethical theories. My main focus is the applied ethics units in these courses. They might cover topics in the areas of sexual morality (e.g. gay

marriage), business (e.g. affirmative action hiring policies), our treatment of animals (e.g. in conducting scientific experiments), censorship (e.g. of various forms of pornography), punishment (e.g. capital punishment), the beginning of life (e.g. abortion), and the end of life (e.g. physician assisted suicide). Often an applied ethics unit begins by clarifying and posing a question. For example, is it morally permissible to have a first trimester abortion when there are no medical complications with the pregnancy, and the pregnancy resulted from consensual sex in which birth control was employed but failed? Typically students then study an essay that argues for an affirmative answer and an essay that argues for a negative answer.

Third, I will only focus on the likely impact of these courses on students who take just one or two such courses in their college career.<sup>1</sup> As we will see later, Haidt himself acknowledges that his Social Intuitionist Model may not apply to people with *extensive* philosophical training in ethics.

Fourth, I will only consider applied ethics courses that mainly proceed by argument analysis. This means setting aside service learning courses, courses that mainly revolve around emotion induction, and other approaches to teaching applied ethics. The kind of course I have in mind focuses on identifying and evaluating arguments provided by proponents of positions on controversial issues in applied ethics. The arguments are evaluated to determine whether they commit any fallacies, whether they rest on plausible assumptions about what the correct moral theory is, whether any narrower ethical principles that they rely on are adequately defended, whether any judgments about individual cases that they rely on are adequately

defended, whether the objections directed at alternative views are cogent, etc. This is a very familiar way that contemporary philosophers teach applied ethics courses, especially philosophers who work in the tradition of analytic philosophy.

Fifth, I take it that it is ultimately empirical investigation that will tell us what goals can be realistically achieved in college-level applied ethics courses. What I offer here is intended as some plausible conjectures about whether some important goals are within realistic reach if Haidt's model is correct. I don't pretend to cover *all* goals. So think of these conjectures as good armchair bets about what further empirical investigations into some important goals of ethics education are likely to turn up.

### **Haidt's Model**

I begin with one piece of evidence that Haidt's model is designed to explain. A warning to readers who regularly teach applied ethics courses: this may remind you of discussions in your ethics classes. Subjects of a study were presented with vignettes like this one:

“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?” (Haidt, Bjorklund, & Murphy, 2000, appendix; Haidt, 2012, p. 38)

A high percentage, some 80% of American undergraduates, respond that it was wrong for Julie and Mark to make love (Haidt, Bjorklund, & Murphy, 2000). When asked why, they typically respond in one of several ways. Some say it will have a negative impact on their relationship. Others say Julie runs the risk of becoming pregnant with a child with birth defects. Subjects are then reminded that neither of these things is true in this case. Despite this, most continue to believe that it was wrong for Julie and Mark to make love (Haidt, 2012, p. 38-40). Haidt calls this *moral dumbfounding*. It occurs when people stick to their quick moral reactions to a case even when by their own lights they cannot convincingly support that reaction.

In support of his main thesis, Haidt offers this and other instances of moral dumbfounding, including other vignettes involving other harmless but offensive acts, like eating flesh from a human cadaver, eating one's dead pet dog, cleaning a toilet with one's national flag, and eating a chicken carcass after masturbating with it, which trigger moral dumbfounding (Haidt, Bjorklund, & Murphy, 2000, appendix; Haidt 2001; Haidt 2012, p. 3-4). He also offers, in support of his main thesis, findings from other social psychological experiments, plus evidence from neuroscience, anthropology, and primatology (Haidt & Joseph, 2004). His main thesis consists in

this slogan: “intuitions come first, strategic reasoning second.” (Haidt, 2012, chapter 2) It is crucial to understand, at least in broad outline, this slogan.

“Intuitions” is the word Haidt uses for the kind of quick emotional flashes that many people have when they read about Julie and Mark’s behavior. These episodes are characterized by being fast and effortless. And the processes that produce them are characterized by being unintentional, running automatically, not being accessible to awareness (we are only aware of the results of these processes), and not demanding the resources of attention (Haidt, 2012, chapters 2-3).

To explain people’s varying intuitions about cases, Haidt posits six individual settings that determine the kinds of intuitions a person will have when they react to different cases (Haidt, 2012, chapters 6-7). He suggests thinking of these as separate dials in the moral mind. One has to do with sensitivity to care and harm, another with sensitivity to fairness and cheating, another with loyalty and betrayal, another with authority and subversion, another with sanctity and degradation, and the last with liberty and oppression. Consider another example: when asked whether parents and teachers should be allowed to spank children who are disobedient, some people have quick intuitions that trigger negative judgments because they perceive spanking as cruel (Haidt, 2012, p. 124). These people are particularly sensitive to considerations having to do with care and harm – their care/harm dial is turned up. However others have quick intuitions that trigger a permissive judgment because they are sensitive to considerations having to do with authority and subversion – their authority/subversion dial is turned up. Where a person’s six moral dials are set, Haidt

claims, is a matter of complicated interactions between evolutionary factors, cultural factors, and socialization factors during childhood.

The second half of Haidt's slogan says that moral reasoning is strategic. By this Haidt means that rather than seek the truth, moral reasoning generates reasons for prior independently produced intuitions (Haidt, 2001, p. 819-825). Moral reasoning, he says, is like a lawyer: it looks for evidence to support a position that is independently assigned to it by another source, namely moral intuition. It differs then from a good scientist who acquires relevant evidence in a non-biased fashion, and lets that evidence dictate the position that she adopts. Moral reasoning functions like a lawyer, according to Haidt, for largely social reasons – most importantly to help us manage our reputations, build alliances, and win people over (Haidt, 2012, chapter 4). And the properties of this kind of reasoning contrast with the properties of intuition. Intuitions recall are produced quickly and effortlessly. Strategic reasoning, by contrast, is slow and effortful. Recall that the psychological processes that produce intuitions run unintentionally and automatically, are not accessible to awareness, and do not demand the resources of attention. By contrast, the processes underlying moral reasoning are intentional, under our control, consciously accessible, and do demand attention's resources.

That in sum is what Haidt says about the nature of intuition and moral reasoning. He gives the following helpful diagram, which situates these two ingredients in his larger Social Intuitionist Model:

*Figure 1*



(Haidt, 2001, p. 815; Haidt, 2012, p. 47)

We have already reviewed links 1 and 2. At link 1, intuitions produce moral judgments; and at link 2, moral judgments produce strategic reasoning. This makes A's moral judgments intermediaries between the quick affective flashes of intuition and strategic moral reasoning. Moral judgments are judgments about what is right and wrong: for example, the judgment that it was wrong for Julie and Mark to have made love, or the judgment that it is permissible for parents and teachers to sometimes spank disobedient children.

The model is rounded out with links 3 to 6 (Haidt & Bjorklund, 2008, p. 190-196). Links 3 and 4 are social: they capture dynamics across two people, A and B. Link 3 represents reasoned persuasion, and link 4 represents social persuasion. For our purposes, what is important about these links is that neither of them involve A influencing B by affecting B's evidence or by affecting B's analysis of evidence. Rather both directly target B's intuitions; so both operate prior to, and independently of, B's reasoning system. By contrast links 5 and 6 are intra-personal and are meant to capture processes within A. According to Haidt, these two processes are rarely utilized (Haidt, 2012, p. 69). Link 5 represents A reasoning her way to a moral judgment about some issue independently of what her intuition is about that issue, and even if this goes against her intuition about that issue. Link 6 represents A's reasoning in a way that influences the nature of A's intuitions.

### **Two Goals That Are Not Realistically Attainable**

Haidt's model, if correct, is bad news for traditional applied ethics courses that focus on formulating and evaluating arguments in support of familiar claims in applied ethics. Recall the relevant kinds of claims are about issues like whether consensual homosexual acts are morally permissible, whether first trimester elective abortions are morally permissible, whether capital punishment is morally permissible, etc. The focus on formulating and evaluating arguments typically goes hand-in-hand with the view that this is the best way for students to form and (if need be) revise their moral judgments. But if Haidt's model is correct, this gets things backwards. Moral argumentation is inert and does not have the power to change students' moral judgments. According to Haidt, moral argumentation issues from an inner lawyer that tries to strategically support positions that are independently assigned to it by moral intuition.

This central feature of Haidt's model bears directly on two traditionally popular and lofty goals of ethics pedagogy. One is *rational substantive changes in belief*. By this I mean changes in students' beliefs about the moral status of some actions, where these changes are based on appreciating various reasons to think *it is true* that the action in question has some particular moral status. Paradigmatically this involves moving from believing some claim to believing the denial of that claim; for example, going from believing that abortion is morally wrong in some range of cases to believing that it is not morally wrong in the same range of cases. This goal is not realistically attainable if Haidt's model is correct.

The use of *belief* here is non-technical. A belief is, among other things, the psychological state that a person reports when asked what their opinion is about something. So when someone is asked whether it was morally wrong for Julie and Mark to have had sex, and answers that they think it was wrong, this is excellent evidence that they believe it was wrong for Julie and Mark to have had sex. Of course, it is not absolutely conclusive evidence since the person answering might be lying, might have misunderstood the question, etc. But frequently we have good background evidence that no such thing is occurring.

Of course what we believe does not just matter for what we say, it also matters for how we act, what we feel, and how we reason. Our beliefs really matter, in part, because they have such a wide range of effects. This is partly why beliefs need to be an important focus of education. But by no means are they the only focus. The kind of moral education that people try to accomplish in ethics classes can involve other goals too: for example, increasing students' sensitivities to other people's circumstances, needs, and views; motivating students to act in certain ways; imparting knowledge of various moral principles; etc. But emphasizing these alternative goals and claiming that this is markedly different from changing students' beliefs is problematic. Consider some examples. To help Michael become more sensitive to the views of women, it might be crucial that he drop his belief that women are less informed than men about important issues, and replace it with the belief that women are just as likely (and often more likely) are more informed about important issues than men are. Similarly, if a student comes to know that treating a person as a mere

means is *prima facie* wrong, this will involve their coming to believe that this is so, as well as coming to believe the supporting evidence for this principle. After all, knowledge requires belief. And knowledge requires having evidence, which often requires having further beliefs. This is not to say that altering beliefs is all there is to changing a person's sensitivities; for example, there are obviously affective components to our sensitivities too. The present claim is simply that emphasizing these other pedagogical goals, while thinking that this allows one to leave the arena of beliefs, is an error. Increasing students' sensitivities to the circumstances, needs, and views of others, motivating them to act in certain ways, and imparting knowledge of moral principles to them frequently involves addressing their beliefs.

There is a second important goal that is not realistically attainable if Haidt's model is correct: *rational substantive changes in behavior*. An example of this is going from never donating money to an organization that helps the poorest people in the world to regularly doing so, where here too the change is brought about by appreciating reasons to think that *it is true* that one is morally required to share in this way. If Haidt's model is correct, people teaching applied ethics courses should not expect to achieve either of these goals. This is because non-strategic moral reasoning – that is, moral reasoning in which moral judgments are informed by reasons for and against claims – is, on Haidt's model, almost always impotent with respect to both of these goals.

Notice I am not claiming that Haidt's model implies that these goals *cannot* be achieved in applied ethics courses. That assertion is too strong to be warranted; and

Haidt himself would disagree with it for just this reason. In the experiment involving the Julie and Mark case, 17% of subjects did change their judgment under questioning. And in an important endnote, Haidt says he is open to the idea that extensive training in philosophical ethics might cause someone to make moral judgments in ways that significantly differ from those identified in the Social Intuitionist Model (Haidt, 2012, p. 329, en. 44). The claim I am presently defending is weaker. It is the claim that rational and substantive changes in belief and behavior are goals that are *sufficiently unlikely* to be achieved in a single applied ethics course that instructors are not warranted in organizing their course around them.

But these two goals seem to be common ones in applied ethics courses.<sup>2</sup> This is the most natural interpretation of the familiar claim that applied ethics courses change, or at least should change, students in meaningful ways. To see the plausibility of my conjecture that most people who teach applied ethics courses aim to achieve these two goals, consider *the examined life test*. It poses two questions to people who teach applied ethics courses. First, are you aiming, at least in part, to help your students lead a more examined life? And second, do you think that leading a more examined life has effects on what one believes and how one acts? My conjecture is that the majority of people who teach applied ethics courses will answer both of these questions in the affirmative. They *do* aim to have their students lead a more examined life. And they *do* think the examined life impacts both what one believes and how one acts.

Of course instructors aim to achieve other goals as well. They may also aim to help students become more morally sensitive to the interests of others, improve their skills in evaluating moral arguments, recognize how ethical principles apply to cases, and achieve many other familiar goals. What I will say later has some implications for these other goals. The point for now though is simply that many instructors do aim, among other things, to have their students examine, and then sometimes revise, their moral beliefs as well as how they act. Notice also that aiming to achieve these goals is consistent with exercising special care in how one tries to achieve them. In particular, instructors can try to achieve these goals without encouraging their students to adopt any specific beliefs or to act in any specific ways. Doing the latter would be objectionably biased and violate ideals of intellectual neutrality, as well as other pedagogical norms.

But if Haidt's model is correct, these goals are no more than pie in the sky. One way to try to counter this pessimism emphasizes links 5 and 6 in Haidt's model. Recall that at these links, reasoning is in control and either directly determines one's moral judgments (link 5) or does so indirectly by first influencing one's moral intuitions (link 6). How probable are the prospects for reconfiguring students' moral minds in a semester or two so that these links are strengthened, and the links at 1 and 2 are weakened? This is obviously a question that must be answered with empirical data.

We should not hold out much hope though that one or two college courses will permanently affect these links in a sufficient number of students to warrant

making these central goals of a course. The biggest obstacle here is exposure. A standard college course has roughly 35 hours of class contact. If we add the optimistic assumption that students spend two hours thinking about the material for every hour of class contact (which is doubtful – see Arum and Roska, 2011), this adds 70 more hours, for a total of roughly 105 hours of exposure for a one-semester course. On the assumption that students are awake 16 hours per day, or 5840 hours in a year ( $365 \times 16 = 5840$ ), this is roughly 1.8% ( $105 / 5840 = 1.8\%$ ) of a student's waking hours in a year. Over the four years that a student spends in college, this is a little less than one-half of one percentage point of their waking hours ( $105 / (5840 \times 4) = 0.45\%$ ). Pit this against the hours, perhaps for many students far more than 105 hours, spent on activities that reinforce links 1 and 2, and it looks like one or two semesters of analyzing arguments in applied ethics is probably not enough time. It is even possible, depending on an instructor's pedagogical techniques, that many of the 35 hours of class contact do more to *reinforce* links 1 and 2 than to weaken them.

Here is a bit of anecdotal evidence. For two semesters in a row, I asked students in my medical ethics course to identify any substantive changes in belief that they underwent because of the course. Each semester 25 or so students were enrolled in the course; and each semester, we covered about 15 applied topics. That makes for roughly 750 opportunities for substantive changes in belief. Not one student reported a single substantive change in belief!

Another attempt to counter this pessimism appeals to links 3 and 4 in Haidt's model. Recall these are social links. At link 3, B's intuition is affected by A's

reasoning; and at link 4, B's intuition is affected by A's judgment. Perhaps there would have been more substantive changes in belief in my classes if greater time was spent having students share their views with one another. In fact, much of our time was devoted to this, with the result that I just reported. More importantly though even if any substantive changes in belief had resulted, these would have been instances of *non-rational persuasion*. It is not any old substantive changes in belief that instructors typically tend to think are valuable. Such changes can be accomplished in all sorts of non-rational and even irrational ways: with regular hypnotism sessions, by finding out what celebrities have vouched for various ethical positions, etc. Recall that links 3 and 4 are non-rational since they *bypass* B's reasoning. So even if B made some novel moral judgment due to influence along link 3 or link 4, that judgment would be no more rational than a judgment that showed up in a moral dumbfounding case. In neither instance is there a rational check on the source of the moral judgment, whether that source is the words of one's interlocutor or the six dials within one's own moral mind. For this reason, substantive changes in belief that are triggered by activity along links 3 and 4, or changes in behavior that are mediately caused by such activity, lack the real cognitive value that we prize.

### **Three Goals That Are Not Sufficiently Valuable**

I turn next to three goals that can be realistically achieved if Haidt's model is correct. However I will argue that each of these goals fails to meet our other condition: they are not sufficiently valuable to be key organizing goals for applied ethics courses. So



if achieving these goals is the best that can be done in applied ethics courses, people should probably stop teaching these courses and do something else with their time.

One goal in this category I call *who said what*. The goal here is to have students memorize what various thinkers have said about different ethical issues. While this might be useful for an appearance on *Jeopardy!*, it does nothing to improve cognitive skills other than memorization. This is of little value: memorization is already the focus of many college courses; and, as students often point out, memorizing facts is not necessary when those facts are as close as their cell phones.

A second goal in this category is *ammunition supply*. The goal here is to equip students with more, and better, reasons that they can offer in support of their intuition-fueled judgments. Haidt's theory predicts that ethics courses will achieve this goal since it says that humans naturally cherry pick their way through moral considerations, heavily weighing what they agree with and discarding the rest. Like a lawyer, moral reasoning has no use for reasons that favor opposing views. Here too there is a serious problem with value. Fostering these kinds of thinking habits violates the basic norm of rationality that says people should form beliefs on the basis of all of the evidence that is available to them, and accurately weigh each piece of evidence. In short, a course organized around *ammunition supply* is an extended exercise in the fallacy of confirmation bias.

A third goal in this category is *self-knowledge*. The goal here is to have each student discover where their six moral dials are set. This could be accomplished by exposing them to a wide range of cases that will trigger their moral intuitions. They

can then identify their intuitions, and see what their intuitions imply about their own individual settings. While this kind of self-knowledge is undoubtedly of some value, it is false advertising to make this a central goal of an ethics course. Ethics courses by their nature are supposed to focus on certain normative questions about the morality of actions. They are not supposed to stop at mere description of people's varied reactions, without critically evaluating those reactions. At best this kind of self-knowledge might be the first step in a longer process of critical reflection. For an ethics course though, it is not an appropriate endpoint.

Regardless of how frequently applied ethics courses are organized around these goals, there are good reasons to think that each of them lacks significant value.

### **Three Modest Goals**

All is not lost. There are at least three goals that are both attainable and worth attaining in applied ethics courses if Haidt's Social Intuitionist Model is correct.

#### *Beyond Demonizing*

The first is a goal that Haidt himself emphasizes: *beyond demonizing* (Haidt, 2012, chapter 12). The goal here is to help students critically evaluate and subsequently abandon any view that they might have to the effect that anyone who disagrees with them about a controversial moral issue does so because they are stupid or evil. Sheer stupidity or moral depravity might explain why *some* people hold the views that they do on a moral issue. But this will not explain why *all* people hold an opposing moral

view, especially when the views concern controversial moral issues like the ones mentioned earlier. Getting beyond demonizing is a valuable cognitive achievement since it replaces inaccurate caricatures of one's opponents with a more accurate picture. It also has real social value since jettisoning these caricatures is good for civil society (Musschenga, 2013).

A powerful strategy for achieving this goal is to generously borrow from traditional applied ethics courses that focus on formulating and evaluating arguments for opposing positions. These activities can help students move beyond demonizing since they involve critically reflecting on the relative merits of various positions, including the strengths of positions that differ from their own, and weaknesses in their own position. In this way, the resources of more traditionally focused courses can be repurposed to help achieve this goal.

Important resources from other traditions of moral education might also be helpful for achieving this goal. For example, borrowing from the cognitive developmental tradition of contemporary neo-Kohlbergians like Rest and colleagues (Rest, Narvaez, Bebeau, and Thoma, 1999 and 2000), a tradition extending back to Kohlberg (1984) and Piaget (1932) might also be helpful. Here exercises using Kohlberg-style dilemmas can help students take up perspectives other than their own. Doing this, and emphasizing the cogency of the reasons that are available to support alternative perspectives, can be a powerful tool for helping students to get beyond simply demonizing opposing viewpoints.

Notice too that this goal can be reached even if rational substantive changes in belief and behavior are unlikely to be achieved. Students who move beyond demonizing their opponents need not convert to their opponents' views. They can continue to think of their opponents as mistaken. But they will now do so with an explanation that is more accurate, complex, and charitable than the demonizing one.

### *Downgrading Intuitions*

A second goal (and one related to the first) that is both realistically attainable and worthwhile if Haidt's model is correct is *downgrading intuitions*. By eliciting Haidt-style intuitions from students, having them scrutinize the reasons that they offer in support of these intuitions, and then having them reflect on whether they might be guilty of moral dumbfounding, instructors can help students become aware of how arbitrary and indefensible it is to take their own quick reactions as authoritative. This goes beyond the earlier self-knowledge goal of simply knowing where one's own six dials are set to critically appraising those settings and reflecting on how likely it is that one's own settings are reliable at detecting what is really morally right and wrong. One's own settings are probably not reliable in this way since Haidt-style intuitions are highly variable across individuals and basically insensitive to evidence (as reflected in the phenomena of moral dumbfounding). Knowing that one's own settings are probably not trustworthy is clearly a worthwhile cognitive achievement.

But is this goal within realistic reach on Haidt's model? Haidt himself presents no empirical evidence about what happens when a person has their own moral

dumbfounding repeatedly pointed out to them and is repeatedly challenged to discuss the implications of their moral dumbfounding. Applied ethics courses are a unique opportunity to do this. Perhaps doing this in certain ways will help students to downgrade their own intuitions.

We need to be modest though. This downgrading, even if it is achieved, cannot be expected to result in substantive changes in belief or behavior. Remember these last two goals must be abandoned if the Social Intuitionist Model is correct. The resulting picture is reminiscent of David Hume's claim that even if skeptical doubts are correct, they are idle (Hume, 1738). So reflective downgrading will not be causally efficacious in the way we might want. That is, knowledge that the source of one's moral beliefs is not trustworthy may not make that source any less prolific in generating moral beliefs. Still the basic Socratic value holds: it is better to have knowledge of your shortcomings than it is to be ignorant of them; and this is so even if this kind of *critical* self-knowledge does nothing to alleviate those shortcomings.

One final point here. Like the previous goal, this goal can be achieved by carrying over readings, class discussions, and assignments that focus on formulating and evaluating arguments for various positions. These tools are indispensable for bringing out the strength of opposing positions and the weaknesses in one's own positions, and thus for reflectively downgrading the accuracy of one's own intuition system. And here too the prospects for achieving this goal can be bolstered by using resources like those from the cognitive developmental tradition mentioned earlier.

*More Ethical Framing*

A final goal that is both achievable and of significant value is *more ethical framing*. The goal here is to have students more frequently frame decisions, both their own and other people's, as ethical decisions. Or, since decisions can be framed in more than one way, the goal is to have students more frequently frame decisions as at least ethical ones, even if they also frame them as some other kind of decision as well – for example, as financial, career, or health decisions. This too is a cognitive improvement since many decisions people face are in part ethical decisions. In fact, according to some ethical theories (most obviously, those that recognize the category of morally optional actions, as many standard forms of utilitarianism and deontological ethics do), every option that a person ever faces can be morally evaluated, thus making *all* decisions that we ever face (in part) ethical decisions. By having all of the important dimensions, both ethical and non-ethical, of decisions before our minds, we have a fuller understanding of our own decisions, and other people's decisions. And having a fuller understanding of a decision is clearly of cognitive value.

Might this also have social benefits? When we frame a decision as an ethical one, does this increase the probability that we will make a better decision? Certainly a better decision is not *guaranteed* to result (though for reasons to be cautiously optimistic, see Bazerman & Tenbrunsel, 2011). And this is a very complex topic (see Kern and Chug, 2009; Kellaris, Boyle, and Dahlstrom, 1994; and Petrinovich and O'Neill, 1996). Pessimism though is probably what is ultimately warranted here if Haidt's model is correct since it predicts that even when a decision is framed

ethically, it will not be informed by a non-biased collection and analysis of the evidence. As usual, our intuitions will rule and they will function in a lawyer-like way, bringing with them all of their acute biases.

Still even if better decisions will not be a probable consequence of framing decisions as ethical decisions, we might expect people to more often take responsibility for their decisions if they frame them this way. For example, framing a decision about how to invest my money as only a financial decision about various means that might make me rich will plausibly dispose me to not take seriously a charge that I am partly morally responsible for an economic injustice that resulted from that decision. By contrast, if I frame this decision as at least in part an ethical one, for example by thinking of it as a decision that will impact other people's legitimate interests, this will plausibly dispose me to take such a charge more seriously. And taking such charges seriously, even if I go on to disagree with them, is better for my personal development and for my relationships with other people. This is a second benefit of framing more decisions as ethical ones: it may help us to more frequently take responsibility for our decisions.

### **Dealing With Disappointment**

Many people who teach traditional applied ethics courses that focus on formulating and evaluating arguments and do so with the aim of prompting students to undergo some substantive rational changes in their beliefs and actions will think that reorienting a course around these last three goals is a serious step down. If Haidt is

right, the best that teachers of applied ethics courses can do amounts to an odd combination of taste-based subjectivism and dogmatism. Each of us has our own individual dials set some way; and though each of us has no probative case for thinking that our dials are set correctly (or for thinking that other people's dials are set incorrectly), our dial settings go right on producing the same moral judgments – and knowing all of this about ourselves does nothing to change it. Hearing this news, it is natural to be disappointed and to lament that substantial rational changes in beliefs and actions are not in the offing.

I end with three things to help blunt this disappointment. First, ethical subjectivism, here thought of as roughly the view that at bottom we have no better guide to what is morally right and what is morally wrong than our moral reactions, sentiments, and other quick processes, may be *true*. This is a highly disputed view of course, but something along this line has had formidable defenders like David Hume, Adam Smith, Charles Stevenson, and others (Kauppinen, 2014). Especially powerful versions of this view constrain the origin of the quick reactions and sentiments to normal ones, excluding for example those that were born of processes shaped by highly stressful events in childhood. I cannot adjudicate these views here –no less than a book is needed to properly do that. But this is just additional evidence that, at least relative to the goal of having students believe what is intellectually credible, and not believe what lacks intellectual credibility, it may not be so bad to have students come out of an applied ethics course as ethical subjectivists.



Second, even if this kind of ethical subjectivism is false, subscribing to it might have beneficial effects on people's attitudes to one another. For example, it might soften their attitudes towards other people by helping them become more tolerant of those who oppose their views, more tolerant of unfamiliar lifestyles that they disapprove of, and more tolerant of public policies that differ from the ones they favor.

Third, recall the earlier point that applied ethics courses that are organized around *beyond demonizing* and *downgrading intuition* can achieve these goals by coopting much of the material of traditionally oriented courses, especially material that focuses on formulating and evaluating ethical arguments. A course constrained by Haidt's model of the moral mind would be premised on the prediction that this material will not help students undergo rational conversions in their ethical beliefs or affect their ethically relevant actions. But as we saw that material can be used in a course that is constrained by Haidt's model in much the same way that it is used in a traditionally oriented applied ethics course, only now with the aim of getting students to go beyond demonizing and to downgrade their own intuitions. In this way, the basic toolkit of a course constrained by Haidt's model can borrow heavily from traditional courses that focus on the analysis and evaluation of arguments. Moreover, as we saw, similar points apply to some of the resources from the cognitive developmental tradition.

## **Conclusion**

In sum, applied ethics courses can be worthwhile even if Haidt's Social Intuitionist Model accurately describes the moral minds of the students in these courses. There was bad news: on the assumption that Haidt's model is accurate in this way, applied ethics courses should not be organized around getting students to engage in rational and substantive revision of their own moral views or of how they act. But in the end, there was good news. These courses can be organized around goals that are both significantly valuable and realistically achievable: namely, helping students to move beyond demonizing their opponents, helping them to reflectively downgrade their own Haidt-style intuitions, and helping them to cultivate the habit of framing decisions as ethical decisions.<sup>3</sup>

---

<sup>1</sup> It is not clear that Haidt's model is intended to apply to judgments in metaethics or normative ethical theory.

<sup>2</sup> Unfortunately I have been not been able to locate any empirical surveys about the percentage of applied ethics instructors that aim to achieve these goals.

<sup>3</sup> Many thanks to Jonathan Evans, Cliff Guthrie, Dana Harrison, Alan Preti, Louis Tietje, Michael Slote, and an audience at the 2012 meetings of the Society for Ethics Across the Curriculum for helpful comments.

## References

- Arum, R. & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. (Chicago, University of Chicago Press).
- Bazerman, M. & Tenbrunsel, A. (2011). *Blindspots: Why we fail to do what's right and what to do about it*. (Princeton, Princeton University Press).
- Fine, C. (2006). Is the emotional dog wagging its rational tail, or chasing it? *Philosophical Explorations* 9, 83-98.
- Fry, D. & Souillac. (2013). The relevance of nomadic forager studies to moral

- foundations theory: moral education and global ethics in the twenty-first century. *Journal of Moral Education* 42: 346-359.
- Graham, J., Haidt, J., & Rimm-Kaufman, S. (2008). Ideology and intuition in moral education. *European Journal of Developmental Science* 2: 269-286.
- Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition find no reason. Unpublished manuscript, University of Virginia.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment, *Psychological Review* 108, 814-834.
- Haidt, J., & Joseph, C., (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus* 133: 55-66.
- Haidt, J. & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Cognitive Science of Morality* (pp. 181-217). Cambridge, MA: MIT Press.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion* (New York, Pantheon Books).
- Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education* 42: 281-297.
- Haste, H. (2013). Deconstructing the elephant and the flag in the lavatory: promises and problems of moral foundations research. *Journal of Moral Education* 42: 316-329.
- Hume, D. (1738). *A treatise of human nature*. Mary Norton and David Norton ed. (New York, Oxford University Press).
- Jacobson, D. (2012). "Moral dumbfounding and moral stupefaction. *Oxford Studies in Normative Ethics* 2, 289-316.
- Kauppinen, A. (2014). Moral sentimentalism, *Stanford Encyclopedia of Philosophy*
- Kellaris, J., Boyle, B., & Dahlstrom, R. (1994). Framing and situational ethics. *Marketing Letters* 5: 69-75.
- Kern, M. & Chugh, D. (2009). Bounded ethicality: The perils of loss framing. *Psychological Science* 20: 378-384.
- Kohlberg, L. (1984). *Essays on moral development: Volume 2. The Psychology of Moral Development*. (San Francisco, Harper & Row).
- Liao, M. (2011). Bias and reasoning: Haidt's theory of moral judgment. In T. Brooks (Ed.), *New Waves in Ethics* (pp. 108-127). New York, Palgrave Macmillan.
- Maxwell, B. & Beaulac, G. (2013). The concept of the moral domain in moral foundations theory and cognitive development theory: Horses for courses? *Journal of Moral Education* 42: 360-382.
- Musschenga, B. (2013). The promises of moral foundations theory. *Journal of Moral Education* 42: 330-345.
- Narvaez, D. (2008). The social intuitionist model: Some counter-intuitions. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Cognitive Science of Morality* (pp. 233-240). Cambridge, MA: MIT Press.
- Petrinovich, L. & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology* 17: 145-171.

- 
- Piaget, J. (1965). *The Moral Judgment of the Child*. M Gabain trans. (New York, Free Press).
- Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach*. (Mahwah, NJ: Erlbaum).
- Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (2000). A neo-Kohlbergian approach to morality research. *Journal of Moral Education* 29: 381-395.