Embedded Public Reasoning: A Response to Jonathan Haidt's The Righteous Mind

Michael J. DeMoor
The King's University, Edmonton, michael.demoor@kingsu.ca

Embedded Public Reasoning: A Response to Jonathan Haidt's The Righteous Mind

Abstract

Jonathan Haidt is a moral psychologist whose influential book, *The Righteous Mind: Why Good People are Divided by Politics and Religion,* explains the origins of our political disagreements. The aim of the book is to encourage understanding and civility in our public life. Deliberative democrats also have a significant stake in understanding the sources of our disagreements and see rational deliberation as the key to civility and democratic legitimacy. However, Haidt's empirical studies give reasons to suggest that the "faith" of deliberative democrats in reasoning may be misplaced, particularly as that faith tends be inflected in terms of a "Kantian" moral psychology.

This article analyzes four different explanatory "stories" that Haidt weaves together: (1) a "causal" evolutionary account of the development of morality; (2) a "causal" story about the psychological mechanisms explaining human action; (3) a "causal" story about the historical and cultural determinants of our political attitudes; and (4) a "normative" story about the grounds and justification of human action. The article then examines these stories to discern how deliberative democrats might rearticulate their conception of public reasoning, and their normative hopes for it, in light of Haidt's findings by introducing the "embedded" conception of public reasoning.

Embedded Public Reasoning:
A Response to Jonathan Haidt's *The Righteous Mind*

Introduction

It goes without saying that deliberative democrats are keen on reason. After all, it is to public reasoning that they turn as a basis for the legitimation of potentially coercive policies[1] and hence their political energies are directed toward creating the conditions and opportunities for such reasoning. These normative commitments to practical reasoning as the basis of legitimate and just governance, however, depend on certain "empirical" assumptions about plurality, disagreement, and reasoning as a social practice. If these assumptions are false, then the normative commitments of deliberative democrats can rightly be rejected as "utopian" and hence excluded from consideration as a basis for real politics. After providing a brief sketch of what I take these assumptions to be, I will consider the work of the moral psychologist Jonathan Haidt, whose empirical findings pose potential challenges to those assumptions and hence to the viability of deliberative democracy as a project. I will argue that, notwithstanding the importance and empirical strength of his picture of human moral psychology, deliberative democrats can nevertheless make good on their empirical assumptions, but only if they adopt a certain, broader view of (public) reason. If our conception and practice of public reasoning is limited only to explicit discursive practices, then Haidt's empirical conclusions do in fact show deliberative democracy to be utopian.

The crucial presupposition of deliberative democracy is that political disagreement is best managed by reasoning, and thus that reasoning is a potentially effective way of resolving disagreements. As I will discuss below, this is the point at which the conclusions of much recent moral psychology—Haidt's work in particular—casts doubt on deliberative democracy as both a theoretical and practical project. This presupposition, in turn, assumes that the disagreements that typify democratic societies are, at least in large part, *reasonable*. That is, the parties to the disagreement must hold their divergent opinions in a way that allows them to be held accountable for their positions in a free exchange of reasons, and that itself is adequately based on reasons that can be offered to explain or justify the position taken. In other words, citizens must be able and willing to reason together. Furthermore, we must show how it is possible that parties to such disagreements can see each other as reasonable even as they disagree. Finally, the existence of reasonable disagreements must be seen to be both an inevitable and a normative feature of free societies. That is, there must be some way to account for the fact that reasonable disagreement is a fact of a properly functioning democratic society, rather than a result of some defect either in the institutional arrangement of that society or in the virtues, dispositions, or powers of its citizens. Otherwise, there would be little point in turning to *reasoning*—rather than coercion, manipulation, or simple aggregation—as a way of addressing the disagreements and differences in our societies.

---

[1] See, for example, Joshua Cohen (1989) and John Dryzek (2011, pp. 21-22): "deliberative democrats generally believe that legitimacy is achieved by deliberative participation on the part of those subject to a collective decision" and that "this participation should have substantial influence on the content of the decision."

Haidt's Challenge

In the book *The Righteous Mind: Why Good People are Divided by Politics and Religion,* Haidt brings together a significant body of empirical research in moral and social psychology. His intention is to use this research to help us understand "why we are so easily divided into hostile groups, each one certain of its righteousness" (Haidt, 2012, p. xii). He invites his readers to "step back, drop the moralism, apply some moral psychology, and analyze the game we are all playing" (Haidt, 2012, p. xvii) so we can come to "disagree more constructively" (Haidt, 2012, p. 274). More specifically, he seeks to uncover the psychological sources of the increasing divide, distrust, and even enmity between "liberals" and "conservatives" in contemporary politics. By being aware of these psychological sources, we can understand the views, attitudes, and fears of others. This, in turn, can lead to greater civility in our public life and possibly better public policy.

All of this is, of course, congenial to the normative goals of many, if not most, deliberative democrats. The idea of disagreeing more constructively is exactly what deliberative democrats hope to achieve by using public reason.[2] However, the way in which Haidt interprets his empirical findings seems to undermine the possibility that public reasoning or deliberation can provide hope for a more civil politics. Haidt's theory about the psychological sources of our disagreements depends on three fundamental claims: (a) that "intuitions come first, strategic reasoning second" (Haidt, 2012, p. [1]); (b) that "there's more to morality than harm and fairness" (Haidt, 2012, p. [93]); and (c) that "morality binds and blinds" (Haidt, 2012, p. [187]). These theses each pose challenges to the validity of the three parts of a deliberative democratic conception of reasonable disagreement as discussed above.[3]

Intuitionism

The first section of *The Righteous Mind* describes a shift in moral psychology away from the "rationalist" assumptions of figures like Jean Piaget and Lawrence Kohlberg for whom moral development is a process in which the end point is mature moral reasoners basing their decisions on rules or principles primarily oriented toward avoiding harm to other individuals. In the light of experiments that test the ways in which people in different cultures make moral decisions, Haidt says that it becomes increasingly clear that it is by no means universally true that this rationalist picture holds. First, subjects repeatedly make moral judgements based on considerations other than harm and fairness. Second, underlying these patterns of judgment is a tendency for the subject to attempt to explain their judgments by using moral reasoning as a *post hoc* justification for judgements made on the basis of intuitive responses (e.g., shock, disgust) rather than as the primary basis of the judgment (see Haidt, 2012, pp. 24-26).

As a result of these (and many other) findings, Haidt rejects the rationalist "Platonic" moral psychology according to which the mature moral reasoner predicates his judgments on rules or principles that have been discursively justified and in which emotions, passions, or intuitions are seen to be an impediment to solid moral judgment. He argues, "from Plato through Kant and

---

[2] Cf. Dryzek and Niemeyer (2011) on "meta-consensus" and List (2007) on "meta-agreement."

[3] Two caveats: first, Haidt is not deliberately assailing the fortress of deliberative democracy; in fact, he doesn't even mention it in his book. Nonetheless, it seems to me that his conclusions pose a real challenge for deliberative democrats and hence need to be answered. Second, I'm not challenging the veracity or methodological adequacy of his empirical findings, which are outside of my professional competence.

Kohlberg, many rationalists have asserted that the ability to reason well about ethical issues *causes* good behavior" (Haidt, 2012, p. 89). Instead he prefers what he calls a "Humean" model of moral judgment, in which reason "is and ought to be the servant of the passions" (Haidt, 2012, p. 30)[4] or, in his own metaphor, in which reason is the rider on the intuitive elephant: the rider can see where the elephant is going, perhaps help the elephant learn new skills, and can act as a "spokesman" explaining what the elephant is doing, but cannot really control or steer the elephant (see Haidt, 2012, pp. 44-47). Our moral judgements are ultimately led by our intuitions and the role of reason, at best, is to help intuition to lead us more effectively, never to replace or control intuition.[5]

This remains true even in social relationships. Even in the give and take of argumentation with others, the rational arguments of others rarely spark a change in the rational judgments of an individual, at least in a way unmediated by that individual's intuitions. At most, the arguments of others can "challenge us, giving us reasons and arguments that sometimes trigger new intuitions, thereby making it possible for us to change our minds" (Haidt, 2012, p. 47). If our friends can only talk to our riders and not to our elephants, then there is little hope of changing our minds: "If you want to change people's minds, you've got to talk to their elephants. You've got to… elicit new intuitions, not new rationales" (Haidt, 2012, p. 48). In order to do this, they do not necessarily need to be more rational than us, but have to be able to empathize with us to see how we see things intuitively—to coax our elephants in different directions rather than giving our riders a new set of instructions on how to steer the elephant.

These conclusions pose a challenge to deliberative democrats because they challenge the assumption that citizens are rational in the sense that they can and do base their political judgments on reasons they are able to make explicit and publicly defend. Furthermore, it brings into question whether the practice of public reasoning is potentially effective as a means of reconciling or managing disagreements and divisions in public life. If explicit justifications primarily figure as *post hoc* rationalizations of intuitive judgments, then it would seem that asking citizens to deal with disagreement by offering one another explicit discursive reasons for their judgment holds out little hope of either changing their own judgments or having them change one another's minds. Though activities or forms of communication that appeal to the *passions* (i.e., emotions, intuitions, fears, hopes) of interlocutors are not necessarily a violation of the norms of deliberative democracy, its advocates are generally convinced that we are both rationally and morally obliged to prioritize appeals to the considered judgment of others, and prefer the "force of the better reason" to coercion, the deft manipulation of symbols, or the tugging of heartstrings. If Haidt is right, these commitments, assumptions, and hopes would seem to be founded on an erroneous "Platonic" moral psychology.

---

[4] Sharon Krause (2008) offers a "Humean" model of political deliberation that in some way parallels—albeit on different grounds—with my attempt to reconcile deliberative politics with recognition of the crucial role of sentiment, intuition, or passion in human moral psychology. Unfortunately, space does not permit me to interact at length with her arguments. See also Andrew Smith (2014b) for an attempt to use Adam Smith's theory of moral sentiments in a similar manner. I can only say for now that my concern with sentimentalist approaches (Hume's in particular) is that they run the risk of presuming a capacity for moral agency that their parsimonious approach to human nature (particularly the denial of natural teleology) undermines.

[5] Besides his own experiments (Haidt, 2001), Haidt cites a wide variety of studies and lines of research that back up his claim that emotion/intuition, rather than reasoning, is the primary causal agent in moral judgment (e.g., Damasio, 1994; de Waal, 1996; Wundt, 1907; Zhong & Liljenquist, 2006; Greene, 2008).

Moral Foundations

Haidt's second contention—that morality is about more than harm and fairness—poses yet another challenge, this one aimed at the Habermasian presupposition that citizens must be capable of recognizing their interlocutors as being *reasonable* even as they disagree.[6] Cross-cultural studies reveal that both the intuitive reactions and the *post hoc* justifications that people offer to explain them are not universal, but vary not only between people of different nationalities or religions, (Haidt, 2012, p. 110) but even between people of different social classes or educational backgrounds. What we feel moved, shocked, or disgusted by, and how we explain those reactions to ourselves and others is shaped by our socialization, our personalities, our experiences, and so on.

Haidt is at pains to resist explaining moral pluralism by appealing to a "blank slate" model of human nature. Although moral intuitions and justifications vary by culture, they vary across a field determined by our evolutionary development, which has given us "adaptations to long-standing threats and opportunities in social life" that "draw people's attention to certain kinds of events (such as cruelty or disrespect), and trigger instant intuitive reactions, perhaps even specific emotions (such as sympathy or anger)" (Haidt, 2012, p. 123). These adaptations are "modules" that consist of "triggers" that evolved in response to certain "adaptive challenges" and are set off by particular kinds of events, resulting in particular emotional responses to which certain characteristic virtues become attached. This description of the sources of our varied moral concerns and their varied expressions is called the "moral foundations theory" (Haidt, 2012, p. 123).

Along with the care/harm module, Haidt identifies the following: "fairness/cheating," "loyalty/betrayal," "authority/subversion," and "sanctity/degradation" (Haidt, 2012, p. 125) to which he later adds "liberty/oppression" (Haidt, 2012, p. 180). Taken together, these modules or dispositions constitute the moral foundations of our intuitions and judgments, and they are, to one degree or another, universal amongst human beings since they all have roots in our evolutionary development. However, within this common context, enormous variation is possible and seeing that is crucial for understanding political disagreements. For example, Western liberals have been socialized to be highly dependent on the care/harm, liberty/oppression, and fairness/cheating modules in making and justifying moral judgments, largely to the exclusion of the other foundations. Groups as culturally close to them as Western conservatives, however, make heavy use of all six modules, thus exhibiting a tendency to be morally "triggered" by things that liberals tend to view as morally irrelevant to explain their moral judgments in terms that liberals often see as irrelevant and even offensive (Haidt, 2012, pp. 181-184).

---

[6] For Habermas, this ability is in fact constitutive for communicative rationality as such: "Intuitively underlying the concept of communication rationality is the experience of the noncoercively unifying, consensus-promoting force of argumentative speech" in which "actors… change their perspective: they… shift perspective from the objectivating attitude of an actor oriented toward success who want to realize some purpose in the world, to the performative attitude of a speaker who wants to reach understanding with a second person with regard to something in the world. Without this switch to the conditions for the use of language oriented toward understanding, the actors would be denied access to the potential inherent in the binding and bonding energies of language" (Habermas, 1998, p. 220, 224).

Since, as discussed above, our moral judgments are primarily shaped by these intuitive responses to which "rational" justification functions primarily as a *post hoc* rationalization, these differences result in a great deal of mutual moral incomprehension and even distrust that is not easily overcome by calmly reasoning together. At the roots of our "disagreements"—even with people very culturally close to us—are enculturated differences in how we respond intuitively to events. Thus, not only do Haidt's findings bring into doubt the capacity of public reasoning to justly manage our disagreements, they also bring into doubt our ability to see those with whom we disagree as reasonable at all. For example, if I see sexual education as touching on the purity of my community and the sanctity of marriage, I am likely to see my interlocutor's overriding concern with sexual autonomy as not only missing the point, but morally monstrous. Reciprocally, my interlocutor is likely to see my concern with purity in terms of her concerns about liberty and oppression and thus to see it as my imposition of a regressive (and probably patriarchal) value system on not only my own but other's children as well. Moreover, these cross-purposes are grounded in our intuitive responses and those in the very soil of our evolutionary psychology and hence, they are likely to be very resistant to changes enacted through mutual reason-giving, however conscientious and well-intentioned.

The Hive Switch

Finally, the assumption of deliberative democrats[7] that disagreements must be rational in the sense that they emerge spontaneously within a well-ordered democratic society is challenged by Haidt's understanding of the sources and limits of our moral imaginations in his insistence that "morality binds and blinds." This view does not challenge the idea that moral disagreements will arise spontaneously, but that they do so as a result of *rational* differences and hence, are susceptible to rational management or adjudication. Rather, they arise in such a way that the factors that influence them (evolutionary imperatives of group formation) do not seem to be responsive or answerable to reason.

Since, like all human phenomena, morality evolved because it offered selective advantages to its bearers, our moral comportment is generally self-interested. Our moral judgments—including, perhaps especially, our intuitive judgments—and conduct are generally oriented by the desire to appear to be good rather than a sincere desire to be good. In this respect "we are 90% chimp" (Haidt, 2012, p. [187]). However, we are not only selfish, we are also "groupish," which means that "our minds contain a variety of mental mechanisms that make us adept at promoting our *group's* interests, in competition with other groups" (Haidt, 2012, p. 191). In this way, we are part bee as well as mostly chimp. This groupishness can, in some circumstances, override selfishness, causing us to exhibit what appears to be "altruistic" judgments or behaviors. Importantly, however, Haidt nowhere identifies a psychological mechanism that allows us to make *disinterested* moral judgments. Whether it is our own individual interest or that of a group, our moral judgments and conduct are fundamentally *interested* and shaped by competition (whether with other individuals in a group or between groups).

---

[7] Rawls (1996, p. 37), for instance, insists that "reasonable pluralism" must be seen as "the work of free practical reason within the framework of free institutions" rather than some unfortunate and eradicable flaw of our societies. I argued in my introduction that this idea is basic to the conception of "reasonable disagreement" that deliberative democrats assume in their normative orientation toward deliberation as a way of managing plurality.

Morality, then, is an artifact primarily of the groupish part of us. We have a "hive switch" (Haidt, 2012, p. 223) that, when activated, allows us to "transcend self-interest and lose ourselves (temporarily and ecstatically) in something larger than ourselves" (Haidt, 2012, p. 223). This moral ecstasis *binds* us in a common identity and common mission with others in a group. But this same switch simultaneously *blinds* us to the moral claims of, or solidarity with, other groups or other individuals therein. In sum, Haidt offers this account of why we disagree: "The explanation is that our minds were designed for groupish righteousness. We are deeply intuitive creatures whose gut feelings drive our strategic reasoning. This makes it difficult—but not impossible—to connect with those who live in other matrices, which are often built on different configurations of the available moral foundations" (Haidt, 2012, pp. 317-318).

In *Political Liberalism*, John Rawls offers an account of the basis of "reasonable pluralism" based on what he calls "the burdens of judgment" (Rawls, 1996, pp. 54-58). These are inescapable facts about the evidentiary incompleteness, normative complexity, semantic indeterminacy, and limited scope of our moral judgment that result in reasonable people coming to different moral conclusions reasonably—and, indeed, being able to recognize each other as doing so. Our disagreements are reasonable because the factors that lead to these disagreements are fundamentally *cognitive* or *conceptual* and are hence amenable to rational deliberation, if not necessarily decisive rational adjudication.[8] Haidt's account of the basis of our disagreements, in contrast, pushes them below the cognitive (or at least *discursive*) level by locating them in our "hivish" intuitive evolutionary psychology. The conclusions of this instinctive groupishness can only be "justified" *post hoc* and even then, at best partially and in terms of the particular configurations of our moral foundations that are shaped by our communities, never in a genuinely disinterested fashion. Grounded disagreements thus are extremely resistant to discursive ("rational") deliberation, management, or reconciliation.

Haidt's Four Stories

My argument in the remainder of this article will be that, notwithstanding Haidt's challenges, deliberative democrats need not despair in their commitment to public reasoning as a means of managing doxastic plurality and grounding the legitimacy of coercive legislation. This is so because, even if we accept more or less at face value his empirical results, Haidt has not shown that moral intuition, as he conceives it, is unresponsive to reason. Therefore, we do not have to abandon the conception of reasonable disagreement that, as I argued in my introduction, is assumed by deliberative democrats. If that is so, then Haidt's challenges to that threefold conception is not fatal to the project of deliberative democracy. The price for doing this, however, is that we must accept a "broader" conception of reason than a merely discursive practice, what I shall call an "embedded" conception of public reasoning.

---

[8] These burdens are fully compatible with agents conscientiously exercising *disinterested* moral judgment. Though I will side with Rawls in insisting that a capacity for disinterested judgment is essential, I should note that he errs in offering only cognitive factors as the basis of our disagreements. In wishing to show how disagreements can emerge even when people are morally blameless, he fails to account for the fact that our disagreements are everywhere and also tied up with factors to which moral blame properly attaches (i.e., self-interested volition, prejudice, lack of sympathy). In this sense, Haidt is indeed the more acute psychologist. Unfortunately, he fails in not recognizing the ways in which those "moral" (rather than "cognitive") factors are accountable and responsive to reasons.

Before I can make this clear, it is necessary to disentangle four "stories"[9] that Haidt weaves together in his moral psychology as outlined above:

Story 1: A causal story of the evolutionary development of morality in human nature

Story 2: A causal story about how human beings make moral judgments

Story 3: A causal story about how history and culture affect our moral judgments

Story 4: A normative story about what grounds and justifies our moral judgments

Most of Haidt's empirical studies focus on Story 2 and, to a lesser extent, Story 3 and he interprets these in terms of Story 2. He frequently demurs from offering any fleshed-out version of Story 4—notwithstanding the "therapeutic" intent of the book—on the grounds that "in psychology our goal is descriptive. We want to discover how the moral mind a*ctually* works, not how it *ought* to work" (Haidt, 2012, p. 120). The "Haidtian" argument against deliberative democracy that I made in the earlier section amounts to the claim that a deliberative democratic account of Story 4 is incompatible with an intuitionist account of Story 2 which is overwhelmingly supported by the empirical (experimental) evidence.

My argument in response will be that a deliberative democratic account of Story 4—in the form of a broadly Kantian view—can be compatible with Haidt's telling of Story 2 as long as we either reject his particular ("consilient") version of Story 1 or give a more fleshed out version of his attenuated approach to Story 3, and preferably both. Before I can make that argument clear, I will say a little bit more about each of these stories.

First, Story 1, as mentioned, serves primarily as an interpretive framework, grounding Haidt's take on Story 2 and Story 3 in a comprehensive narrative and a conception of human nature. It is in light of this grounding that he is able to claim that the features of our moral psychology (i.e., primarily intuitionism, moral foundations theory, and the hive switch) are permanent features of human nature: we are "born to be righteous" in that "human nature is not just intrinsically moral, it's also intrinsically moralistic, critical, and judgmental" (Haidt, 2012, p. xii-xiii). For Haidt, this approach is a working out of E.O. Wilson's explanatory strategy of "consilience" in which "the study of ethics would… be taken out of the hands of philosophers and 'biologicized' or made to fit with the emerging science of human nature" (Haidt, 2012, p. 2) which is "refounded as the interpretation of the activity of the 'emotional centers' of the brain" (Haidt, 2012, p. 67). For Haidt, a biological evolutionary psychology is the proper study of human nature, and everything to be included in an adequate account of human nature has to be cashed out in terms of the narrative articulated by that science.

Second, I have already said a good deal about Story 2. The key features of this story are intuitionism, moral foundations theory, and the hive switch. Human beings make moral judgments

---

[9] I use the word "stories" here largely for lack of a better term, since not all of them are "theories," nor are they all "assumptions" or are all explicitly argued for in the book. In a sense, they are "stories" because in each case they are all accounts of how something develops, has developed, or how something comes about. Furthermore, this fits with the narrative framing that Haidt himself gives the various major divisions of the book. "Stories" remains, though, a deliberately inexact term in my use of it here and little in the way of narrative theory should be read into it.

fundamentally on the basis of intuitive (i.e., emotional, immediate, "gut-level") responses to situations, whereas discursive or "rational" justifications come after these responses and function primarily to explain or rationalize them both to one's self and (primarily) others.[10] These intuitive responses are clustered into "modules" focused on different areas of moral concern of which care/harm and fairness/cheating or only two of six. Furthermore, these responses, though generally self-interested can sometimes assume the form of altruism when circumstances trigger our tendency toward strong identification with our group(s). The vast majority of Haidt's empirical studies focus on establishing this story, and I will grant that they do so adequately. Furthermore, the challenges to deliberative democracy described above primarily come from the three main features of this story. One part of this story that goes beyond what the empirical studies justify, however, is his tendency to cash it out in "mechanical" metaphors of "switches" and "triggers," which is motivated by his attachment to a consilient version of Story 1, which I shall have more to say below.

Thirdly, we have seen that Haidt points to historical and cultural development and difference (Story 3) to explain how the picture of universal human nature interpreted in terms of Story 1 and empirically justified in Story 2 nevertheless admits a substantial variation between different groups. He connects this to Story 1 by reflecting on the ways in which culture and genetics interact in the process of human evolution (see Haidt, 2012, p. 210ff), which simultaneously provides a way of making his consilient version of Story 1 less reductionistic and shows the "naturalness" of Story 3 by grounding it in the science of human nature par excellence (viz., biologistic evolutionary psychology). The price that Haidt pays for doing so, however, is that he is extremely sketchy on how this process of cultural learning and differentiation happens and, crucially for my purposes, whether it can be described as a rational process.

Finally, Haidt repeatedly insists that the goal of moral psychology is not to give a normative account of moral judgment (Story 4), but rather is restricted to a descriptive story (Stories 1-3). However, fact and value are never that easy to keep apart, and elements of a normative story are clearly present in the book. In the first place, the "ought" emerges when Haidt makes a suggestion for what the actual purpose of moral reasoning might be, albeit after showing that moral reasoning plays little causal role in individual moral judgment. He says: "We do moral reasoning not to reconstruct the actual reasons why *we ourselves* came to a judgment; we reason to find the best possible reasons why *somebody else ought to join us* in our judgment" (Haidt, 2012, p. 44). Our moral judgments have, if you will, objective normative purport: we make them assuming that what we judge to be right is in fact right and hence, binding on the judgments of others regardless of their subjective preferences or views.

Furthermore, *The Righteous Mind* itself has such normative purport in so far as the intent of the book is therapeutic: to contribute to the remedy of a pathology in our political culture (viz., our inability to disagree constructively and conduct our public affairs with civility). Haidt hopes that his work can do this by enabling us to "understand ourselves, our divisions, our limits, and our potentials" and to "step back, drop the moralism, apply some moral psychology, and analyze the game we're all playing" (Haidt, 2012, p. xvii). The idea, presumably, is that some self-reflection,

---

[10] This appeal to others is itself grounded primarily in our "chimp-like" (or "Glauconian" to employ another of Haidt's metaphors) desire to *appear* to be good (serving the interests of the group) regardless of whether we are actually being good.

guided by the truths revealed in the book, will temper our tendency to rush to the judgment of others and to view one another as acting from either base (usually culpable or invincible) ignorance or from morally dubious motives. That this hope is shared by deliberative democrats[11] is significant. They wish to ground that hope on human capacity for practical (moral) reason. Haidt wants to ground it on our ability to take moral lessons from scientific moral psychology. These views are not so far apart; the question will be which approach to grounding this hope is more adequate. We have seen that Haidt gives reasons to wonder whether the deliberative democratic grounding is sufficient; the goal of the remainder of this article is to argue that it is, notwithstanding Haidt's empirical conclusions. A secondary goal will be to suggest that some of Haidt's assumptions make him unable to give an account of human psychology that can ground his own hope. This particularly applies to his telling of Story 1 and his failure to offer a compelling version of Story 3, which together constitutes an account of human moral psychology largely devoid of meaningful agency.

The argument will proceed in two steps. First, I will argue that a broadly Kantian, "rationalist" version of Story 4, which I will take to be affirmed by most deliberative democrats, does not rely on what Haidt calls a "Platonic" moral psychology. Second, I will argue that we can make sense of the compatibility of a rationalist Story 4 and Haidt's intuitionist Story 2 if we see the rationality of moral judgment being grounded in either a "rationalist" account of Story 3 or a "rationalist" account of Story 1. Doing so will mean going beyond Kant in a "Hegelian" and/or "Aristotelian" direction but in ways that deepen and ground that Kantian rationalism.

Reason and Moral Judgment

I begin with the assumption that deliberative democrats are, as a rule, rationalists about Story 4, at least with respect to public life. What I mean by this is that theorists such as Habermas believe that moral judgments or decisions made in public are legitimized and justified by their rationality.[12] In other words, actions, statements, and decisions in the public sphere ought to be normatively evaluated, at least in part, according to the degree to which they are rational. This means that acceptable[13] reasons can be and are usually given for why these decisions are right, and that these reasons appeal to facts, principles, values, or beliefs that are themselves rational in the same sense and are, usually but not necessarily, widely accepted by the various interlocutors in the public sphere. For the purposes of this article, I will call this view "broadly Kantian," without meaning

---

[11] Mutz (2008, p. 523) summarizes deliberative democrats' hopes in a list of desired outcomes of deliberation: "more public-spirited attitudes; more informed citizens; greater understanding of the sources and rationales behind public disagreements; a stronger sense of political efficacy; willingness to compromise; greater interest in political participation; and, for some theorists, a binding consensus decision." Though Mutz famously argues that empirical research suggests that many of these hopes stand in tension with one another, this remains a pretty good summary of the hopes I have in mind here (cf. also Gutmann & Thompson, 2004, pp. 10-13).

[12] Habermas's version of this is his "discourse principle": "Just those action norms are valid to which all possibly affected persons could agree as participants in a rational discourse" (Habermas, 1996, p. 107). Naturally, deliberative democrats will vary widely in the way in which they express such a principle and many will take issue with the categorical nature of it, but the fundamental point remains that almost all deliberative democrats are committed to the idea that political actions norms are legitimated by social practices of reasoning (cf. DeMoor, 2014).

[13] Though some deliberative democrats display this tendency (e.g., Cohen, 1989), it is not necessary for them to be too parsimonious about what counts as "acceptable" reasons, at least not in ways that require them to be strictly secular or universally acceptable (cf. Smith, 2014a).

to suggest that Kant himself would endorse it in precisely these terms or to deny that there are non-Kantian (say, for example, civic-republican) ways of articulating, grounding, and endorsing it.[14]

In spite of the fact that Kant himself seems to endorse a "Platonic" moral psychology,[15] the broadly Kantian public morality I sketched above is compatible with the idea that, in the individual judger, passion or intuition causally precedes moral judgment and gives direction to any subsequent course of moral reasoning as long as intuition is not thought to be an intrinsically non-rational "faculty."[16] Someone might be said to be acting rationally, which the rationalist normative story requires, even if their judgment is caused by intuition if three conditions apply. First, the actor is rational if the intuition that caused her action is (prima facie) in accordance with what reason does in fact require, even if the agent herself is not able to demonstrate that connection. One might also say, by analogy to epistemology, this is a partly "externalist" account of moral justification. This externalism will be further unpacked in my discussion of an "Aristotelian" take on Story 1. Second, it is rational if, when the intuition does not fully align with what reason requires, it is responsive to reasons to bring it into alignment. Thirdly, if our intuitive reactions are grounded in our human nature and yet culturally shaped, then they might be rational if these background conditions are themselves rational. In other words, even if reason is not prior to intuition in the consciousness or act of the judger, that act might yet be rational if reason is prior in the situation which grounds and conditions the judgment. If these conditions obtain, then intuition itself can be said to be rational (at least potentially)[17] and hence, judgments causally preceded by intuition can be as well even if the judger is not able to adequately make the reasons behind the judgment fully explicit. If, on the other hand, they do not obtain, then my broadly Kantian public morality is in trouble: If our judgments are not and cannot be grounded in, conditioned by, and responsive to reason, then it is frankly unreasonable to use reason as a norm by which to evaluate them, at least if "ought" implies "can."

It is crucial, however, to see that this way of showing the compatibility of a rationalist Story 4 and an intuitionist Story 2 depends on uncoupling "rationality" from "explicit discursive awareness." All of these conditions show how a judgment can be rational in a way that the judger himself is not fully or explicitly aware of—the rationality of the judgment is "embedded" in other cognitive attitudes (i.e., emotion, "gut feeling," social expectation, aesthetic taste) than "knowing-that" a certain principle is right or true and consciously applying it to the case at hand. Haidt's experimental results do in fact strongly suggest that moral judgment rarely, if ever, takes a Platonic form and, even if it does so sometimes, it is never to the exclusion of it being mixed with those other cognitive attitudes. Thus, for example, even a very conscientious attempt to base one's stance on, say, euthanasia on moral first principles (say, human dignity, the place of autonomy in the moral life) will be mixed with a felt identification with some groups and an intuitive distrust of others, and so on. Without this uncoupling, my broadly Kantian public morality would indeed

---

[14] On the difference between Kantian and civic-republican ways of grounding and articulating deliberative democracy, see DeMoor (2014).

[15] Cf. Kant (1964, p. 71): "In studying the moral knowledge of ordinary human reason we have now arrived at its first principle [the categorical imperative]. This principle it admittedly does not conceive thus abstractly in its universal form; but it does always have it actually before its eyes and does use it as a norm of judgment."

[16] I put "faculty" in scare quotes here because I am troubled by any attempt to impose a faculty psychology on complex practices like public moral reasoning. Haidt (and Kant and Hume) unfortunately still tends to retain this metaphor.

[17] For an extensive discussion of rationality in emotion, cf. Martha Nussbaum (2001).

depend upon a Platonic moral psychology, which would fall afoul of the Haidtian objection from the causal primacy of intuition.

Reason in History and Human Nature

What I need to show now is that these three conditions do in fact obtain. That is, I need to show that it is possible for reason to be prior in the situation of the judger, if not in the act of judgment itself, and that it can ground and condition the intuition that is the proximate cause in the act of moral judgment. A secondary aim must be to show that this grounding and conditioning makes the judger's intuitive responses normatively answerable and responsive to reasons. We can make sense of these possibilities by showing how the act of judgment is situated in both the historical or cultural development of social practices and institutions and in an abiding human nature. I will show how such a case might be made by offering a broadly Hegelian interpretation of history and culture and a broadly Aristotelian interpretation of human nature.[18]

It might be a truism, but one easily overlooked, that moral judgment is situated ("embedded" if you will) in concrete social practices and institutions. These practices and institutions are the products of historical development which shape and condition those judgments in various ways. More specifically, these social situations train our intuitive responses to events that are themselves shaped by these social situations. For example, the experience of formal schooling conditions us to have certain expectations regarding order and precedence in group situations of certain kinds and thus, to have a negative intuitive reaction when these expectations are confounded. Though we may be able to give some account of the rules of proper conduct in these situations, that is not always the case. Conscious consideration of the rules is not necessary to respond appropriately. However, even when this conscious awareness is not present, there is a "rationality" to our judgments in that they implicitly appeal to norms of appropriateness which are themselves determined by the ends of the social practice or institution within which our intuitions are shaped. Thus, the norms of precedence and order in the classroom are shaped by the ends of providing equal learning opportunities to all students and the necessity to stay on task in order to accomplish the learning outcomes desired.

The "fit" between the norms and the ends of the practice is a question of rationality, specifically of coherence: if adherence to the norms does not advance or cohere with the ends, then there can rightly be said to be something "irrational" about the institutional setting within which our intuitive responses are conditioned and exercised. This irrationality, or lack of fit, is as often intuitively felt—perhaps as frustration and boredom—as consciously recognized. We have, I am assuming, an intuitive, felt orientation toward the satisfaction of the human needs or goods that are served by our practices and institutions,[19] the frustration of which is thus intuitively recognized even when we cannot articulate it. Perduring irrationality of this sort will be corrected by revising either the ends or the norms, and usually both simultaneously,[20] and this process too is rarely simply the

---

[18] I do not wish to get bogged down in details of historical interpretation. I will thus sketch a "broadly Hegelian" take on history and culture and a "broadly Aristotelian" take on (human) nature in the same spirit as I offered a "broadly Kantian" account of public morality.

[19] Why this is and how it can be unpacked rationally is the point of turning to Aristotle and human nature.

[20] This "dialectical" process of contradiction ("lack of fit") and revision is why I call this process "broadly Hegelian." The logic of this process is laid out the dialectic of Concept and Object in Hegel (1977). For a discussion

result of conscious discursive reflection but of "feeling out" new arrangements. Even when it is primarily a matter of "feeling things out," the process of reconciling norms and ends in social institutions is a rational one in so far as it: (a) responds to perceived, even if only intuitively perceived, irrationality (incoherence), and (b) results in a more rational (coherent) order. In other words, the embeddedness of our intuitive moral judgment in social practices and institutions—where those intuitions are both shaped by and respond to a network of norms and ends (as all practices and institutions do)—shows them to be both shaped by an order of reason and responsive to it. Though this process is often greatly facilitated by conscious discursive reflection, it does not depend upon it. Moreover, even when it is so facilitated, it is never unaffected by our intuitive responses. Haidt does recognize something like this social embeddedness (Story 3), but his development of it is truncated, first, by his tendency to offload explanatory weight to his "consilient" version of Story 1, and, second, by his reliance on a "mechanical" model of human agency that makes it difficult to see how "feeling out" our social embeddedness and intuitively "working through" the question of the rational coherence of our social practices can be a process that we *undertake* in response to reasons rather than a set of reactions that we *undergo*.

Another avenue by which to articulate the way in which even intuitive moral judgments can be rational is to show that our intuitions are justificatorily grounded in and conditioned by a human nature that is rational. If the nature that leads to our having the intuitions we do is rational, then those intuitions can be thought to be prima facie rational in so far as they would accord with an order of reason when everything is functioning "naturally." I will show how this might go by sketching a broadly Aristotelian, teleological, account of human nature that accounts for the basic features of Haidt's intuitionist moral psychology but shows them to be "rational" in the relevant sense.

The foundation of this account is the notion that human beings are naturally oriented toward their good, in fact, toward a coherent variety of goods.[21] Some of these goods are necessities for survival (i.e., food, water, reproduction) while others are perfections of our natural capacities (i.e., rationality, beauty, moral virtue), and most of them, at least the latter sort, are irreducibly social goods (Taylor, 1995) that cannot be achieved or enjoyed by individuals in their isolation. This orientation toward the good can be seen, indeed experienced, at all levels of human awareness or responsiveness from bodily sensation (e.g., hunger, thirst) to intuitive or emotional response (e.g., anger at injustice, pleasure in friendship), to self-conscious discursive thought. In fact, these different levels tend to work simultaneously and usually in concert.[22] Even when this last level is largely absent in the subject himself, the bodily and intuitive responses are rational when they are

---

of how it unfolds in Hegel's political thought focusing on the concept of "realization" [*verwirklichung*], see Robert Pippin (1981).

[21] Thomas Aquinas (1988, pp. 35-36) calls this "synderesis… a natural disposition… described as impelling us to good and opposing evil…"

[22] For example, "confirmation bias"—which Haidt adduces as a strong indication of captivity of our reason to our intuitive elephant (Haidt, 2012, pp. 79-81)—is a coordination of intuitive and discursive levels that, though it may cause you to make irrational judgments in any particular instance, is a rational disposition to have, since it prevents us from an overhasty abandonment of our commitments in the face of recalcitrant experience. Without it, we would be at sea rationally speaking, buffeted about by the waves of whatever theory or argument has the most recent word. In other words, people with confirmation bias are more likely to achieve the proper ends of human cognition than people without it and are hence more rational. This points to a level of rationality that Haidt simply does not account for because of his resolutely nonteleological take on Story 1.

functioning properly, that is, when they imply correct judgments[23] about the relationship between the situation at hand and the genuine good(s) of the human beings involved.[24]

Although this orientation to the good is natural, it nevertheless needs to be cultivated (Story 3) at both the intuitive and the discursive levels. John McDowell (1996, p. 84) describes this process as one of: "having one's eyes opened to reasons at large by acquiring a second nature… it is what figures in German philosophy as *Bildung*." Our moral intuitions (moral sentiments, perhaps), though natural, need to be shaped, constrained, directed, and articulated in social practices and institutions. This "cultural" aspect of human nature—this "second nature"—is not an add-on or a supplement, but implied directly in the goods to which we are naturally oriented. In virtue of this fact, we are not only naturally oriented toward goods but are also naturally *receptive* to cultural influences in shaping that orientation. As discussed above, these social processes can be seen to be rational. Thus, the receptivity to cultural influence even at the intuitive level that is built into our nature is a kind of responsiveness to reason(s). A rationalist take on human nature (Story 1) is both compatible with and re-enforced by a rationalist take on history and culture (Story 3)[25] and both are compatible with an intuitionist account of the psychological process of moral judgment.

Furthermore, the diversity of the goods to which we are naturally oriented can be the basis on which this broadly Aristotelian account of human nature can give an account of the various "moral modules" that Haidt identifies in his "moral foundations theory." Haidt characterizes the various flavors of (intuitive) moral concern as having evolved in response to various adaptive challenges for early humans that have been (culturally) developed beyond their "original triggers" to include responsiveness to "current triggers" that have become intuitively associated with the original triggers (Haidt, 2012, p. 125). Obviously, Haidt's characterization of these moral foundations is offered in the conceptual categories of his "consilient" take on Story 1. The account of human nature that I am currently developing could—without loss to the experimental empirical basis of the theory—offer a related story. The moral foundations are responses to different human goods or modes of human flourishing, rather than merely to "adaptive challenges." For example, the "care/harm" foundation develops because relationships of mutual concern and intimacy (characteristic of the family, primarily, but beyond it as well) are crucial and irreducible to our being fully human. Similarly, the "authority/subversion" foundation points to our natural need for *Bildung* and governance in order to flourish, which in turn requires relationships of authority and obedience (e.g., between teacher and student, parent and child). The "sanctity/degradation" module, then, may make reference to the fact that human beings are ordered not only to temporal goods but to ultimate ones as well, and that enjoying these ends require a sense of the "holy" or the "sacred," and perhaps the renunciation of certain kinds of temporal goods as conflicting with the achievement of these higher goods.

---

[23] It is implied here that intuitions are or contain (implicit) judgments. For a discussion see Nussbaum (2001, p. 19ff). As far as I can see, nothing in Haidt's experimental results imply noncognitivism about emotion, unless cognition is reduced to conscious discursive reasoning.

[24] Because this relationship need not be explicitly recognized—lest this picture fall afoul of Haidt's intuitionism—and yet confers rational justification on the moral judgments, this view implies a kind of externalism about epistemic (moral) warrant, perhaps by analogy to the "proper function" account of warrant in Alvin Plantinga (1993).

[25] In other words, I take my "Hegelian" picture of the process of cultural development to be compatible with my "Aristotelian" picture of human nature. Though I do not wish my argument to depend on matters of historical interpretation, I will say that this view is re-enforced by some very real and important connections between Hegel's and Aristotle's thought (cf. Ferrarin, 2007).

In light of this "teleological" interpretation of the "natural" basis of the moral foundations, the development from "original triggers" to "current triggers" can be interpreted not merely as an aimless cultural embroidering on a biological cloth, but as a rational, teleologically guided development, unfolding perhaps in a cultural process like the one I described above. Thus, the movement from having the emotions associated with the care/harm module being "triggered" only by the needs of one's own child to being triggered also by the neediness of children in other groups or by animals can be evaluated in terms of whether these changes are fitting or appropriate given the kind of human flourishing associated with that module. This openness to normative critique is grounded in the fact that these changes happen *for* reasons and *in response to* reasons.

Finally, the irreducible sociality of these goods provides a rational basis for the evolution of what Haidt calls our "groupishness" activated by our "hive switch." If many of the proper forms of human flourishing are goods that cannot be enjoyed alone, but only as a member of a group, then it is fitting and appropriate—not only for our survival, but for our comprehensive well-being—that we are naturally disposed to identify with various groups and, when necessary, to subordinate our individual interests to the common good of those communities. Of course, this orientation can sometimes drive us to do awful or stupid things but it is one that is rational both in the sense that, first, it is right and fitting that we have it, given what our good consists in; and, second, its deployment can be rationally evaluated and revised.[26] This capacity for revision (responsiveness to reasons, if you will) is grounded in the fact that our groupishness is not only natural but is also a kind of "second nature" cultivated in processes of *Bildung* through which, not only our ideas but our intuitions are shaped in (potentially) rational processes. Thus, this "Aristotelian" take on Story 1 accounts for all of the aspects of our moral psychology that Haidt experimentally justifies in his version of Story 2, but does so in a way that does not conflict with, but actually underwrites a "rationalist" version of Story 4, to which deliberative democrats are committed.

Conclusion

I have argued that a broadly Hegelian picture of cultural formation (Story 3) and a broadly Aristotelian picture of human nature (Story 1) can be employed to show that a rationalist public morality (Story 4)—to which I take it deliberative democrats are committed—can be compatible with an intuitionist account of the causal process of individual moral judgment (Story 2), granted only that we assume that: (a) the rationality of our moral judgment can depend upon a fit between the good/rational and the contents of our judgment that we as individuals may not be able to articulate ("externalism"); and (b) the development and revision of our intuitive dispositions to make moral judgments can be rational even when their responsiveness to reasons does not take the form of explicit, conscious discursive attitudes ("embedded" rationality).[27] As a matter of fact, I

---

[26] For a parallel, see my discussion of confirmation bias in footnote 22.

[27] I do not have the space to explore this concept of "embedded" rationality at any greater length in this article, except to note that I am inclined to think that it has significant consequences for how deliberative democrats conceive the relationship between the public sphere and civil society. The institutions and social practices characteristic of civil society are the "natural" home for embedded reasoning of a variety of sorts, depending on the ends of those institutions (e.g., religious, aesthetic, cultural). The "intramural" and "embedded" conversations within those institutions will need to be the crucial space within which not only new public possibilities are *discovered*, but in which they need to be *justified* in order to secure genuine subjective legitimacy and hence the public sphere should be as inclusive and even as supportive as possible of voices shaping and shaped by those intramural conversations (e.g., religiously-based schools). On the context of justification and the context of discovery in deliberative democracy see Jürgen Habermas (1996, p. 307). For a critique, see Michael DeMoor (2014).

would be inclined to argue that Haidt's own normative commitments and articulation of the social character of moral judgment make these assumptions not only optional, but necessary, but that is beyond the scope of this article.

References

Aquinas, T. (1988). *St. Thomas Aquinas on politics and ethics* (P. Sigmund, Trans. and Ed.). New York, NY: Norton.

Cohen, J. (1989). Deliberation and democratic legitimacy. In A. Hamlin & P. Petit (Eds.), *The good polity: Normative analysis of the state* (pp. 17-34). Oxford, UK: Basil Blackwell.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York, NY: Putnam.

DeMoor, M. (2014). Legitimacy, public justice, and deliberative democracy. In J. Bowlin (Ed.), *The Kuyper center review, Volume 4: Calvinism and democracy* (pp. 104-133). Grand Rapids, MI: Eerdmans.

Dryzek, J. (2011). *Foundations and frontiers of deliberative governance*. Oxford, UK: Oxford University Press.

Dryzek, J., & Niemeyer, S. (2011). Pluralism and meta-consensus. In J. Dryzek & S. Niemeyer (Eds.), *Foundations and Frontiers of Deliberative Governance* (pp. 85-115). Oxford, UK: Oxford University Press.

Ferrarin, A. (2007). *Hegel and Aristotle*. Cambridge, UK: Cambridge University Press.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnot-Armstrong (Ed.), *Moral psychology, Volume 3: The neuroscience of morality: Emotion, Brain Disorders, and Development* (pp. 35-79). Cambridge, MA: MIT Press.

Gutmann, A., & Thompson, D. (2004). *Why deliberative democracy?* Princeton, NJ: Princeton University Press.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Rehg, Trans.). Cambridge, MA: MIT Press.

Habermas, J. (1998). Actions, speech acts, linguistically mediated interactions and the lifeworld. In M. Cooke (Ed.), *On the pragmatics of communication* (pp. 215-256). Cambridge, MA: MIT Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon.

Hegel, G. W. F. (1977). *Hegel's phenomenology of spirit* (A.V. Miller, Trans.). Oxford, UK: Oxford University Press.

Kant, I. (1964). *Groundwork of the metaphysics of morals* (H. J. Paton, Trans.). New York, NY: Harper & Row.

Krause, S. R. (2008). *Civil passions: Moral sentiment and democratic deliberation*. Princeton, NJ: Princeton University Press.

List, C. (2007). Deliberation and agreement. In S.W. Rosenberg (Ed.), *Deliberation, participation and democracy: Can the people govern?* (pp. 64-81). London, UK: Palgrave MacMillan.

McDowell, J. (1996). *Mind and world: With a new introduction by the author*. Cambridge, MA: Harvard University Press.

Mutz, D. C. (2008). Is deliberative democracy a falsifiable theory? *Annual Review of Political Science, 11*, 521-38.

Nussbaum, M. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge, UK: Cambridge University Press.

Pippin, R. (1981). Hegel's political argument and the problem of *Virwirklichung*. *Political Theory, 9*(4), 509-532.

Plantinga, A. (1993). *Warrant and proper function*. Oxford, UK: Oxford University Press.

Rawls, J. (1996). *Political liberalism: With a new introduction and the 'reply to Habermas'*. New York, NY: Columbia University Press

Smith, A. F. (2014a). Political deliberation and the challenge of bounded rationality. *Politics, philosophy & economics, 13*(3), 269-291.

Smith, A. F. (2014b). Religion in the public sphere: Incentivizing reciprocal deliberative engagement. *Philosophy and social criticism, 40*(6), 535-554.

Taylor, C. (1995). Irreducibly social goods. In C. Taylor (Ed.), *Philosophical arguments* (pp. 127-145). Cambridge, MA: Harvard University Press.

Wundt, W. (1907). *Outlines of psychology*. Leipzig, Germany: Wilhelm Englemann.

Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science, 313*, 1451-1452.