

BIAS AND REASONING: HAIDT'S THEORY OF MORAL JUDGMENT

Forthcoming in *New Waves in Ethics*, ed. Thom Brooks. Palgrave.

© S. MATTHEW LIAO

New York University, 285 Mercer Street, Room 1005, New York, NY 10003, USA; e-mail: matthew.liao@nyu.edu; www.smatthewliao.com

August 12, 2010

Bias and Reasoning: Haidt's Theory of Moral Judgment

Abstract

According to Haidt's Social Intuitionist Model (SIM) of moral judgment, most moral judgments are generated by the intuitive process and the purpose of reasoning is to provide a post hoc and biased basis for justification. The SIM is of great importance for moral philosophers because if the SIM were an accurate description of how we arrive at our moral judgments, the evidential weight of most of our moral judgments may be undercut. In this paper, I question Haidt's claim that reasoning provides a biased basis for justification by challenging his claim that reasoning is biased. After presenting the tendencies that, according to Haidt, make reasoning biased, I draw on the literature on epistemic justification to show that these tendencies are not always biases. If I am right, it is premature to claim that our reasoning is biased, and that the purpose of reasoning is to provide a biased basis for justification.

Keywords: moral judgment; moral intuitions; Social Intuitionist Model; Jonathan Haidt; cognitive biases

Bias and Reasoning: Haidt's Theory of Moral Judgment¹

1. Haidt's Social Intuitionist Model

A topic of significant interest among social psychologists today is the extent to which intuitions, as opposed to reasoning, play a role in determining moral judgments (Haidt, 2001; Greene and Haidt, 2002; Pizarro and Bloom, 2003). Labelling the automatic, effortless, rapid process of intuitions as System 1, and the controlled, effortful, slow process of reasoning as System 2, a dominant perspective in developmental psychology – following the works of Piaget and Kohlberg – has been that our moral judgments are the products of System 2 (Kohlberg 1969; Piaget 1932). This is the so-called Rationalist Model of moral judgment. Recently, however, some social psychologists have proposed that at least some of our moral judgments are the product of System 1. In fact, some have even argued that moral judgments arise predominantly as a result of the intuitive process, and the purpose of reasoning appears not to generate moral judgments, but to provide a post hoc and biased basis for justification. In particular, in a series of work, Jonathan Haidt and his collaborators have defended the ‘Social Intuitionist Model’ (SIM) of moral judgment (Haidt, 2001; Haidt 2007; Haidt and Bjorklund 2007a; Haidt and Bjorklund 2007b). According to the SIM, moral judgments are initially the product of non-conscious automatic intuitive processing. Conscious reasoning then takes place and is typically occupied by the task of justifying whatever intuitions that happen to be presented to the consciousness in a biased, non-truth seeking way.

¹ I would like to thank Jonathan Haidt, Walter Sinnott-Armstrong, Cordelia Fine, Neil Levy, Steve Clarke, Guy Kahane, David Wasserman, Jill Craigie, and Wibke Gruetjen for their helpful comments on earlier versions of this paper.

To support the SIM, Haidt, among other things, appeals to the phenomenon of ‘moral dumbfounding’ (Haidt and Hersh, 2001). Moral dumbfounding occurs when people have moral convictions that they are incapable of justifying, but which they nevertheless continue to hold. For instance, Haidt presents the following case in a survey:

Julie and Mark are brother and sister. They are travelling 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, 2001, p. 814)

Haidt found that most people regarded the action of Julie and Mark as immoral. When asked to justify their judgment, people typically referred to the possibility of inbreeding and the possible psychological harm to the siblings. However, when it was pointed out to them that in this case, it is stipulated that birth control had been used and that there was no psychological harm to the siblings, Haidt found that many would nevertheless continue to insist that the activity of the siblings was morally wrong. Haidt discovered in particular that some people would make up ‘reasons, sometimes bad ones, to justify their condemnation,’ while others would concede that they cannot provide any further

justification for their judgments (Haidt, 2006, p. 64; Haidt, Bjorklund and Murphy, unpublished). According to Haidt, since these individuals were offering bad reasons or were not able to provide any further justification in support of their judgments, it does not seem plausible to hold that their judgments were generated by reasoning. Instead, it seems more likely that these individuals were using reasoning to justify their intuitions in a post hoc and biased way (Haidt, 2001, p. 822).

Haidt claims that in fact most moral judgments arise in this manner. The SIM does include the possibility of using reasoning to influence our judgments and using private reflection to influence our initial intuitions, which can then directly generate new future judgments and intuitions. However, Haidt argues that in reality these are rare occurrences (2001, p. 815). Instead, according to Haidt, the revision of our moral judgment occurs primarily through social interaction, in which other people’s post hoc reasoning evoke new moral intuitions in us.

2. Post Hoc Reasons and Motivated Reasoning

To appreciate the novelty and the radicalness of Haidt’s theory of moral judgment, it is useful to contrast the SIM with two other models of moral judgment.² The first is the

² On the dual process model of moral judgment, if one were to take into consideration the relative causal contributions of the intuitive and reasoning processes, as well as the goodness/badness of these processes understood in terms of truth-seeking/biases, one can identify twelve different models of moral judgment:

$S1_B = S2_B$	$S1_G = S2_G$	$S1_G = S2_B$	$S1_B = S2_G$
$S1_B > S2_B$	$S1_G > S2_G$	$S1_G > S2_B$	$S1_B > S2_G$
$S1_B < S2_B$	$S1_G < S2_G$	$S1_G < S2_B$	$S1_B < S2_G$

Reflective Equilibrium Model, according to which we use both System 1 and System 2 to develop and form our moral judgments. This model is akin to (if not the same as) the standard philosophical practice of using intuitions to test against theories and using theories to test against intuitions in order to reach some sort of coherence (Rawls, 1971). On the Reflective Equilibrium Model, we often use intuitions to test against reasoning and reasoning to test against intuitions in order to reach some sort of coherent moral judgment. Unlike the SIM, the Reflective Equilibrium Model does not claim that moral judgments arise predominantly as a result of the intuitive process or that the purpose of reasoning is to provide a post hoc and biased basis for justification. Another model that is also different from the SIM is what might be called the Benign Social Intuitionist Model, according to which moral judgments arise predominantly as a result of the intuitive process but reasoning can sometimes help solve problems that the intuitive process cannot solve or help improve our moral judgments by providing further justifications for our initial intuitive judgments.³ As we shall see, Haidt explicitly rejects the Benign Social Intuitionist Model. Haidt's novel claims are that we use predominantly

Key: S1 = System 1; S2 = System 2; '=' means roughly equal causal contribution; '>' means greater causal contribution; '<' means lesser causal contribution; B = bad, understood in terms of biases; G = good, understood in terms of truth-seeking. The SIM seems to be $S1_G > S2_B$, though sometimes Haidt could be read as holding $S1_B > S2_B$. The Benign Social Intuitionist Model is $S1_G > S2_G$, which Haidt rejects. He certainly does not hold $S1_B > S2_G$. The Reflective Equilibrium Model could be something like $S1_G = S2_G$. And the Rationalist Model could be $S1_G < S2_G$ or more radically, $S1_B < S2_G$.

³ One might get the Benign Social Intuitionist Model if one applies Gigerenzer's view on dual processing to moral judgments (Gigerenzer and Todd, 1999). Steve Clarke (2008)'s SIM_S could support a Benign Social Intuitionist Model, though SIM_S is concerned only with the structure of the SIM and not the goodness/badness understood in terms of truth-seeking/biases of the intuitive and the reasoning processes.

System 1 to form our moral judgments, *and* that System 2 is primarily used for providing a post hoc and biased basis for justification.

The SIM is of great importance for moral philosophers because if it were an accurate description of how we arrive at our moral judgments, the evidential weight of most of our moral judgments may be undercut. Indeed, as Gibbard (2008, p. 14) observes in his Berkeley Tanner Lectures, ‘if intuitions are the sorts of states that figure in Haidt’s picture, [that is, if they are ones that reach moral conclusions all by themselves,] why place any stock in them?’ Or, Kennett and Fine (2009) argue that the SIM challenges our normative conception of ourselves as agents capable of grasping and responding to reasons, and that there can be no ‘real’ moral judgments in the absence of a capacity for reflective shaping and endorsement of moral judgments.

In the literature critiquing Haidt’s work, Fine (2006, p. 95) has helpfully observed that Haidt makes two types of claims regarding our reasoning:

the motivated nature of our reasoning (‘the reasoning process is more like a lawyer defending a client than a judge or scientist seeking truth’ (2001, 820); and *the post hoc nature of reasons* (‘the reasoning process readily constructs justifications of intuitive judgments, causing the illusion of objective reasoning’ (2001, 822)) (my italics).

Most of the SIM’s critics have focused on the post hoc nature of reasons. For instance, a number of writers have pointed out that intuitive judgments may reflect the ‘automatization’ of judgments based on prior moral reasoning (Saltzstein and

Kasachkoff, 2004), and that moral reasoning can disrupt the automatic process of judgment formation described by the SIM, either by a slow, intentional, deliberative and effortful ‘after-the-fact’ correction, or by an ‘up-front’ preconscious control (Fine, 2006).

In response, Haidt has accepted these possibilities, but he claims that these occurrences are rare (Haidt and Bjorklund, 2007a). However, it might be asked, what evidence does Haidt have that the occurrences of reasoning prior to intuitive judgments are in fact rare? After all, it is not implausible to think that the phenomenon of moral dumbfounding occurs only in a handful of special cases. For example, arguably, Haidt’s case of Julie and Mark discussed above is a special case of incest constructed by Haidt. In more mundane cases of incest, the possibility of inbreeding and the possible psychological harm to the sibling would seem to support and justify the judgment that incest is wrong. In discussing how common private reflection takes place, Haidt concedes that ‘there is not at present evidence that would allow an answer to this question’ (Haidt, 2003). This said, to be fair to Haidt, Haidt’s critics also have not shown that the occurrences of reasoning prior to intuitive judgments are not rare. So the debate on this point appears to be at a stand-still at least for now.

However, Haidt’s claim is not just that reasoning is post-hoc, but that it is also *biased*.⁴ So, even if we cannot yet settle the issue of whether moral judgments arise predominantly as a result of the intuitive process, Haidt could still maintain the other

⁴ Some might think that the motivated nature of reasoning is not essential to the SIM. However, in personal communication, Haidt has explicitly said that it is essential to his view that reasoning is biased. Also, recall Fine’s point that Haidt makes two types of claims regarding our reasoning: the post-hoc nature of reasons and the motivated nature of reasoning. In addition, Haidt has said that reasoning is biased in numerous places in his writings (see, e.g., Haidt, 2006, Chapter 4). Furthermore, if Haidt did not regard reasoning as being biased, the SIM would not be all that different from, e.g., the Benign Social Intuitionist Model, which Haidt explicitly rejects.

aspect of the SIM, according to which the purpose of reasoning is to provide a biased basis for justification.

In response to this point, Fine has argued that we can also be motivated to be truth-seeking (2006, p. 94), and Neil Levy could be read as proposing that the biases Haidt has in mind can be cancelled out as long as reasoning is put in the open and subjected to a ‘community-wide’ scrutiny ‘led by moral experts’ (2006). However, to support the idea that reasoning is biased, Haidt has appealed to the vast literature on motivated reasoning and biases. If that literature is to be believed, it is not obvious that biases can be so easily cancelled out, as, e.g., Levy seems to suppose. For example, there are studies that show that people become more biased in groups (Kerr, MacCoun, and Kramer, 1996). Also, there are studies that show that peer review may be subject to the kinds of motivated reasoning that are seen in individuals (Mahoney, 1977). Indeed, some people have also speculated that theoretical physics in the last 20 years has been dominated by advocates of string theories at the expense of other theories (Smolin, 2006). If biases can take place at the community-wide level just as they can at the individual level, community-wide scrutiny may only have limited effects in cancelling out biases.

In the following, I shall examine in greater detail the literature on motivated reasoning and biases, which forms the basis for Haidt’s claim that reasoning is biased. I begin by presenting the tendencies that, according to Haidt, make reasoning biased. After noting that intuitions are equally vulnerable to these tendencies, and that there might be debiasing strategies that could be employed to counter these tendencies on behalf of reasoning, I then point out that Haidt and others simply assume that these tendencies are biases. Drawing on the literature in epistemology on epistemic justification, I argue that

these tendencies are not *always* biases. If this is correct, it is premature to claim that our reasoning is biased, and that the purpose of reasoning is to provide a biased basis for justification.

3. Haidt's Case for the Biased Nature of Reasoning

On the Benign Social Intuitionist Model, moral judgments arise predominantly as a result of the intuitive process, but reasoning is the smarter, but cognitively more expensive, process that is employed whenever the intuitive process is unable to solve a problem cheaply. Haidt argues though that empirical evidence does not support this model (2001, p. 820). Referring to the literature on motivated reasoning, Haidt points to two factors that, according to him, make reasoning biased.

The first is the relatedness motives (Haidt, 2001, p. 821). According to these motives, we tend to be motivated to agree with our friends, and, as a result, we tend actually to agree with our friends. Haidt cites, for example, studies by Chaiken, Shechter and Chaiken (1996) that found that people who expected to discuss an issue with a partner whose attitudes were known tended to express attitudes that were evaluatively consistent with the partner's opinion. As Chaiken et al. explain, sometimes we are motivated to agree with our friends just because we do not want to come across as being disagreeable. For example, when the views of our friends are unknown, we might employ a 'moderate opinions minimize disagreement' heuristic in order to have a smooth interaction with such a friend (Chen, Shechter, and Chaiken, 1996, p. 263). Or, when the views of our friends are known, we might employ a 'go along to get along' heuristic to serve the same goal (Chen, Shechter, and Chaiken, 1996, p. 263). As Haidt says, the

existence of such a motive means that the ‘mere fact that your friend expresses a moral judgment against X is often sufficient to cause in you a critical attitude toward X,’ which, according to Haidt, is a kind of bias (2001, p. 821).

The second kind of factor that makes reasoning biased is what Haidt calls the coherence motives (2001, p. 821). According to these motives, we tend to try to keep our attitudes and beliefs congruent with the beliefs and attitudes that are central to our identity, and we tend to dismiss evidence that threatens attitudes and beliefs that are constitutive of our identity. Haidt points, for example, to studies that found that we tend to accept evidence supporting our prior beliefs with *less scrutiny* and we tend to subject opposing evidence to *greater scrutiny*. For example, in the classic study by Lord, Ross, and Lepper (1977) in which two groups of people were selected, one strongly in favor of capital punishment and the other strongly opposed to capital punishment, the researchers gave people from both groups some ‘mixed evidence,’ namely, a piece of new research that suggested that capital punishment was an effective deterrent and a piece of new research that suggested that capital punishment was not an effective deterrent. The researchers found that people tended to regard the research that was consistent with their original views to be better conducted and more convincing than the research that conflicted with their original views. Moreover, the researchers found that people tended to hold on to their initial views even more strongly after having been presented with the mixed evidence. According to Lord, Ross and Lepper, this belief polarization does not seem to be a rational response, because learning about research that conflicted with our views should cause us to have reduced, rather than increased, confidence in our views.

Haidt goes on to cite other studies that, according to him, revealed how the relatedness and coherence motives make people act like ‘intuitive lawyers’ rather than ‘intuitive scientists’ (2001, p. 821). As Haidt explains, he is not trying to demonstrate that ‘people are stupid or irrational’ (2001, p. 821). Instead, Haidt believes that he is only trying to show that ‘the roots of human intelligence, rationality, and ethical sophistication should not be sought in our ability to search for and evaluate evidence *in an open and unbiased way*’ (my italics) (2001, p. 821).

4. Some Initial Defense of Reasoning

Haidt is surely correct that reasoning can sometimes be biased. In addition to the possible biasing factors that Haidt had mentioned, we can add to the list the fact that we are not very reliable at detecting correlations; we tend to be overconfident about the power of our reasoning; and so on (Nisbett and Ross, 1980; Trope and Liberman, 1996). Moreover, our cognitive limitations, including limits on our memory and attention, can further contribute to the relative unreliability of our reasoning.

However, these concerns should be put in perspective. I shall present four considerations, the last of which I shall discuss in greater detail in the remaining parts of the paper.

First, according to those social psychologists whose studies form the basis for Haidt’s claim that reasoning is biased, there are three main kinds of ‘motivational underpinnings of information processing’ (Chaiken, Giner-Sorolla, and Chen, 1996). Haidt has mentioned the relatedness and the coherence motives, but there are also the ‘accuracy motives,’ which are characterized by ‘a relatively impartial, open-minded

treatment of information' (Chen, Shechter, Chaiken, 1996, p. 263). While social psychologists believe that we are sometimes or often motivated by the relatedness and the coherence motives, they also believe that we are also often motivated by the accuracy motives (Fine, 2006, p. 94). Among other things, these social psychologists have found that even when the relatedness and the coherence motives are activated, the accuracy motives can also remain activated. To give one example, in a study in which the participants are told that introverts are more likely to be successful in order to motivate the participants to rate themselves as being more introverted, it was found that subjects, who were predominantly extraverted to begin with, viewed themselves as less extraverted when they believed introversion to be more desirable, but they still viewed themselves as extraverted (Kunda and Sanitioso, 1989). *Pace* Haidt, such research suggests that reasoning is receptive to unbiased evidence even when the other motives are activated.

Secondly, while our reasoning can be biased, so can our intuitions. As Chaiken et al. (1996, p. 263) have noted, the relatedness and the coherence motives can just as easily influence our intuitions (what they call heuristic processing) as they can our reasoning (what they call systematic processing). If so, even if Haidt were correct that reasoning is biased, intuitions may not fare much better. Note that Haidt may not be too troubled by the idea that our intuitions are also distorted. However, at least in some of his writings, he seems to think that intuitions can be fairly accurate. For example, he says that 'Rather than following the ancient Greeks in worshiping reason, we should instead look for the roots of human intelligence, rationality, and virtue in what the mind does best: perception, intuition, and other mental operations that are quick, effortless, and *generally quite accurate*' (my italics) (Haidt, 2001, p. 822).

Thirdly, psychologists and philosophers have explored the possibility of debiasing strategies for our reasoning (and for our intuitions). For example, Wilson and Brekke (1994) offer the following general strategies: First, we must be aware of the biases. Secondly, we must be motivated to correct these biases. Thirdly, even if we were motivated to correct the biases, we must also be aware of the direction and the magnitude of the bias. Finally, we must be aware of whether we have sufficient control over our responses to be able to correct these biases. Or, Bishop and Trout (2005) propose that one should consider explanations for propositions that one does not believe; make statistical judgments in terms of frequencies rather than probabilities; and so on.⁵ Levy's idea of 'distributed cognition' may also be considered as a debiasing strategy. These strategies may give us some ideas of the conditions under which we can counter biases.

Fourth, Haidt and others simply assume that the relatedness and the coherence motives are biases. But it is worth asking whether they always are, since if they are not, this would undercut Haidt's basis for believing that reasoning is biased. In the remaining part of the paper, I shall examine whether these motives are always biases. To do this, it is important to say more about the concept of bias.

5. Bias and Epistemic Justification

According to Wilson and Brekke, who have examined the extensive social psychology literature on bias (or what they call 'mental contamination'), there have been two general ways of defining bias in this literature (1994, p. 120). According to the first, people who do not use a rule that experts agree is correct are biased (Nisbett and Ross, 1980, p. 13).

⁵ See also (Sinnott-Armstrong, 2006).

According to the second, which is Wilson's and Brekke's preferred approach, 'a judgment, emotion, or behavior is said to be contaminated if it was influenced in an 'unwanted' way (i.e., unwanted by the person whose judgment, emotion, or behavior it is)' (1994, p. 120).

For our purposes, I believe that we should not use either definition. To see why, consider the first definition. It seems that something can be a bias even if the experts say otherwise (e.g. the alleged bias towards string theory). Or, suppose that being motivated to agree with our friends is a bias. It seems that this would remain a bias even if experts were to say that this is not a bias. The first definition implies though that if experts say that this is *not* a bias, then it is no longer a bias.

Consider the second definition. It seems that something can remain a bias even if it is a 'wanted' influence. For example, suppose again that being motivated to agree with our friends is a bias. However, suppose that I want to be motivated to agree with my friends. Since my judgment is influenced in a 'wanted' way, the second definition implies that I am no longer biased.

To be fair, the social psychologists most likely intended for these two definitions to be 'operational' definitions, that is, definitions that can be used in empirical research. Still, neither appears to be adequate as a conceptual definition of bias.

So what is an adequate conceptual account of bias? It is beyond the scope of this paper to give a full account of this concept. To offer something that is I think relatively uncontroversial and sufficient for our purpose, I propose that we understand the concept of bias in terms of the vast literature in epistemology on epistemic justification. In

particular, if someone, X, is *epistemically justified* in believing P, then X is not biased about P. And, if X is biased about P, then X is not epistemically justified in believing P.

On a standard account of epistemic justification, what makes a judgment, P, epistemically justified depends in part on whether one has adequate evidence for P and whether one bases one's judgment that P on that evidence (Alston, 1985). If one has adequate evidence for P and one bases one's judgment for P on that evidence, then one's judgment that P is epistemically justified. If one lacks adequate evidence for P or if one does not base one's judgment that P on some adequate evidence, then one's judgment that P is not epistemically justified. For our purpose, this means that someone who is not biased about P is someone who judges that P with adequate evidence for P and on the basis of that evidence, and someone who is biased about P is someone who judges that P without adequate evidence for P or without basing the judgment that P on some adequate evidence.

Here it is worth remembering that epistemic justification does not require true beliefs. That is, one can have epistemically justified false beliefs. This means that two people can be epistemically justified in disagreeing with one another, *without being biased*, even if one of them is holding a false belief.

Let me accept that in some, more general, sense of the term 'bias,' one could be epistemically justified in believing P and still be biased about P. But the kinds of biases at issue here are not these kinds of biases. Consider, for example, why the tendency to be motivated to agree with our friends could be a bias. Arguably, an explanation is that the fact that someone is our friend is typically not the right kind of epistemic evidence that one should be using to base a judgment. If so, there could be a bias in this case precisely

because one might not have met the appropriate epistemic standard of justification in basing a judgment on the fact that someone is our friend. Or, consider why the tendency to subject one's opponent's views to much greater scrutiny could be a bias. A plausible explanation is that one might be ignoring certain valid evidence against one's own judgment. If so, there could also be a bias in this case precisely because one has not met the appropriate epistemic standard of justification in basing a judgment on selectively chosen evidence. No doubt, more can be said about this matter. But I shall assume that these are the kinds of biases that are at issue here. Note that since the concept of bias in this paper is related to epistemic justification, to vary my choice of words, I shall sometimes refer to someone who is not biased as being 'epistemically rational' and someone who is biased as being 'epistemically irrational.'

I shall now consider whether the relatedness and the coherence motives are always biases using this, more adequate, conception of bias. Since the empirical literature on these two motives is vast, I cannot hope to discuss the entire literature here. What I shall aim to do is to offer proof of concept. In particular, I shall consider some of the strongest cases for why these motives appear to be biases and I shall show that even by these lights, these motives are not always biases.

6. Are Relatedness Motives Always Biases?

Recall that according to the relatedness motives, we tend to be motivated to agree with our friends, and as a result, we tend actually to agree with our friends. Are these tendencies always biases?

Let me mention that I am aware that some of the literature on the relatedness motives suggests that these motives may work for both friends and non-friends alike. However, it seems that if the relatedness motives are pernicious at all, they will be particularly pernicious in cases of friends. Moreover, as we have seen, Haidt also focuses on our tendency to agree with our friends. As he says,

The existence of motivations to agree with our friends and allies means that we can be directly affected by their judgments (the social persuasion link). The mere fact that your friend expresses a moral judgment against X is often sufficient to cause in you a critical attitude toward X (Haidt, 2001, p. 821).

So I shall also focus my discussion on this tendency to agree with our friends.

Let us consider first whether the tendency to be motivated to agree with our friends is always a bias. Here is a reason why it is not.

A friend is typically someone whose judgments you have reasons to trust in general. That is, you have reasons to trust that when your friend disagrees with you, she will be honest with you, try to familiarize herself with the relevant available evidence that bear on the issue, exercise epistemic virtues such as intelligence and thoughtfulness, and so on. Suppose that you find yourself in a disagreement with your friend about a particular matter. Suppose further that you and she are both fallible and neither has any special epistemic advantage regarding the issue at stake. Can it be epistemically rational for you to reason that because she is your friend and because she disagrees with you, you should suspend your judgment or at least move your judgment towards her judgment?

Arguably, it can be epistemically rational to do so. Why? Given that you trust her judgments in general, it can be epistemically rational for you to believe that there is a

chance you might be mistaken and that she might be correct on this particular occasion. Given this, it can be epistemically rational for you at least to move your judgment towards her judgment. In other words, it can be epistemically rational for you to be motivated to agree with your friend in certain circumstances, because typically a friend is someone whose judgments you trust in general.

In contrast, consider someone who is not a friend. Suppose that he is someone whose judgments you neither trust nor distrust in general. Suppose that you find yourself in a disagreement with this person about a particular matter. Suppose further that you and he are both fallible and neither has any special epistemic advantage regarding the issue at stake. Can it be epistemically rational for you to reason that because he is not your friend, you need not suspend your judgment or move your judgment towards his judgment?

Again, arguably, it can be epistemically rational to do so. Why? Given that you neither trust nor distrust his judgments in general, it can be epistemically rational for you to believe that there is a chance that he might be mistaken and that you might be correct on this particular occasion. Given this, it can be epistemically rational for you not to suspend your judgment or move your judgment towards his judgment. In other words, it can be epistemically rational for you not to be motivated to agree with someone who is not your friend on certain occasions, because that person might be someone whose judgment you neither trust nor distrust.

The general point here is that the relatedness motives merely show that we have a tendency to be motivated to agree with our friends. But one should distinguish between

- a) our tendency to be motivated to agree with our friends because we want to be harmonious with our friends, and
- b) our tendency to be motivated to agree with our friends because we tend to trust our friends' judgments in general.

Both a) and b) can explain why we have the tendency to be motivated to agree with our friends. But while a) may be epistemically irrational, b) need not be, since it can be epistemically justified to trust a friend's judgment (e.g. based on how the friend has judged in the past about other matters). Given that a) is Haidt's explanation for why we have the tendency to be motivated to agree with our friends, but b) is a viable possibility and is not epistemically irrational, the fact that we tend to be motivated to agree with our friends need not always be a bias. Note that my claim here is not the strong one that it is never the case that when you are motivated to agree with your friend, you are motivated in this way because you want to agree with her. Instead, my claim is the weaker one that it is not the case that when you are motivated to agree with your friend, you are always motivated in this way because you want to agree with her.

Of course, we may also have friends whose judgments in general we do not trust. In those cases, it would certainly be a bias if we were to be motivated to agree with such friends, even when we do not trust their judgments in general.

Let us next consider whether the fact that we tend actually to agree with our friends entails that we are biased. Again, it is useful to distinguish two cases.

- c) We tend to have similar evidence and we tend to treat the evidence in a similar way, because we are friends.
- d) We are friends, because in part we tend to have similar evidence and we tend to treat the evidence in a similar way.

c) may indeed involve a bias, but d) need not always involve a bias. The reason is that if our friends happen to have similar evidence as we do and tend to treat them in similar ways as we do, it is not epistemically irrational to have similar judgments as our friends, given that we both possess similar evidence.

This is not to say that we cannot and should not disagree with our friends. Nor does it mean that we cannot and should not agree with someone who is not our friend. On a particular occasion, if we have evidence that a friend is wrong, then we are rationally required to disagree with our friend. We would be epistemically irrational on this occasion if we were to suppress the evidence so that we and our friend could agree. Similarly, on a particular occasion, if someone who is not our friend has evidence that we are wrong, then we are rationally required to move towards his judgment. We would be epistemically irrational on this occasion if we were to ignore the evidence just because he is not our friend. However, as long as we are making judgments based on adequate evidence, and as long we and our friends tend to agree because we have similar evidence, our judgments can be epistemically justified even if they tend to be the same. If so, our tendency actually to agree with our friends also need not always be a bias.

7. Are Coherence Motives Always Biases?

Recall that the coherence motives are ones according to which we tend to try to keep our attitudes and beliefs congruent with the beliefs and attitudes that are central to our identity. As we have seen earlier, a purported strong piece of evidence of this tendency is our tendency to accept evidence supporting our prior beliefs with less scrutiny and to subject opposing evidence to much greater scrutiny. Is this tendency always a bias?

Here are a few scenarios in which we might have such a tendency and in which it is not a bias to have such a tendency.

Scenario 1:

Suppose that you believe P because of E_1 , and your opponent believes $\sim P$ because of E_2 . It may be epistemically rational for you to accept E_1 with less scrutiny, while subjecting E_2 to greater scrutiny, for the following reason. You are already aware of E_1 , but you are not aware of E_2 . Since epistemic justification requires that you seek the truth about P, you should subject E_2 to greater scrutiny in case E_2 *does* undermine P. If so, in this scenario, your accepting evidence supporting your prior beliefs with less scrutiny while subjecting opposing evidence to greater scrutiny need not be a bias.

Scenario 2:

Suppose that you believe P, and your opponent believes $\sim P$. There might be objections to P. However, you are only aware of objections to $\sim P$ based on r, s, and t. It may be epistemically rational for you to spend more time thinking about objections to $\sim P$ based on r, s, and t rather than objections to P, because you are actually aware of r, s, and t, but you are only aware that there might be objections to P. In other words, it does not seem

epistemically irrational for someone to spend more time thinking about objections of which she is aware rather than objections of which she might become aware.

Of course, suppose that you are only aware that there might be objections to P and that there might be objections to \sim P. In such a case, it would seem to be epistemically irrational if you were to spend all your time trying to find objections to \sim P and not spend any time trying to find objections to P. Or, suppose that you are aware of objections to P based on x, y, and z, as well as objections to \sim P based on r, s, and t. In such a case, it would also seem to be epistemically irrational if you were to spend all your time only thinking about objections to \sim P based on r, s, and t, and not objections to P based on x, y, and z.

In other words, in cases in which you are *symmetrically* aware of how much evidence there is against P and against \sim P, it would seem to be epistemically irrational to be uneven-handed in your treatment of the evidence. But in cases in which you are *asymmetrically* aware of how much evidence there is against P and against \sim P, it need not be epistemically irrational for you to spend more time thinking about objections of which you are actually aware rather than objections of which you might become aware. If this is right, asymmetric cases would be ones in which your accepting evidence supporting your prior beliefs with less scrutiny, while subjecting opposing evidence to greater scrutiny need not be a bias.

Scenario 3:

In the belief polarization studies by Lord, Ross and Lepper, two groups with divergent views were presented with mixed evidence, and their views became even more divergent. Can this be epistemically rational?

To see how it can be, let us begin with a simple case in which two people initially have the same view about a particular matter; they are presented with the same new evidence; and their views diverge rationally. Suppose that A and B both initially believe to the same degree that climate change is a problem. Suppose that A and B then learn that Al Gore has announced in a documentary that climate change is a serious problem. A's and B's views about climate change can rationally diverge in the following way.

Suppose that A is a documentary maker. She has previously received memos from credible sources that Al Gore would make a documentary about climate change only if he is certain that there is sufficient evidence that climate change is a serious problem. Upon learning that Al Gore has announced in a documentary that climate change is a serious problem, A may become more certain in an epistemically rational way that climate change really is a serious problem. A's change of view can be epistemically rational because the memos about Al Gore and her knowledge of them are *prima facie* appropriate epistemic evidence.

In contrast, suppose that B is a political historian. Through her independent research, she has found a significant statistical correlation between political candidates who have lost in an election and their trying to reclaim spotlights in the media by making grandiose, but false claims. Upon learning that Al Gore has announced in a documentary that climate change is a serious problem, B may become more skeptical in an epistemically rational manner that climate change is a serious problem. B's change of

view can be epistemically rational because her independent research is also prima facie appropriate evidence.

Here then is a case in which two people with the same initial view about a particular matter can have rationally divergent views about that matter when given the same new evidence. As one can see from the example, this can occur because people can have different, but epistemically rational, background beliefs, which can lead them to hold rationally divergent views when they are presented with the same new evidence.

Similar things can be said about the more complex case involving two people with two opposing views, who are presented with mixed evidence and whose views diverge even further as a result. Suppose that A believes that consequentialism is the correct moral theory, and B believes that consequentialism is not the correct morally theory. Suppose that A and B are then presented with some mixed evidence, namely, an argument in favor of consequentialism (F_1) and an argument against consequentialism (F_2).

F_1 : Consequentialism is the correct moral theory because a correct moral theory would say that when we are faced with the decision to kill one to save 100, we should kill the one to save the 100, and consequentialism does say this.

F_2 : Consequentialism is not the correct moral theory because a correct moral theory should not be too demanding and consequentialism is too demanding.

Here is how A and B can rationally accept evidence that supports their respective views and rationally dismiss the evidence against their respective views, which would result in their views' diverging even further, without being biased. Suppose that A believes – independently of the debate about the correctness of consequentialism and with appropriate epistemic justification – that 'morality should be demanding' and that 'when we are faced with the decision to kill one to save 100, we should kill the one to save the 100.' When faced with F_1 , A will accept it as confirming his belief that consequentialism is the correct moral theory, since it is consistent with A's independent and epistemically justified belief that 'when we are faced with the decision to kill one to save 100, we should kill the one to save the 100.' When faced with F_2 , A will reject it, since it conflicts with A's independent and epistemically justified belief that 'morality should be demanding.'

In contrast, suppose that B believes – independently of the debate about the correctness of consequentialism and with appropriate epistemic justification – that 'morality should not be demanding' and that 'when we are faced with the decision to kill one to save 100, we should not kill the one to save the 100.' When faced with F_1 , B will reject it, since it conflicts with B's independent and epistemically justified belief that 'when we are faced with the decision to kill one to save 100, we should not kill the one to save the 100.' When faced with F_2 , B will accept it, as it is consistent with B's independent and epistemically justified belief that 'morality should not be demanding.'

This case illustrates then that two people with opposing views can, *without being biased*, rationally accept evidence that supports their respective views and rationally dismiss evidence against their respective views, which can then result in their views'

diverging even further. Similar to the previous case, this can occur because people can have different, but epistemically justified, background beliefs, which can lead them to hold rationally divergent views when they are presented with new but mixed evidence.

Note that I am not arguing that two people with opposing views are never biased when they accept evidence that supports their respective views and dismiss evidence against their respective views, which results in their views' diverging even further. My limited claim is that finding belief polarization does not mean that the individuals involved are therefore biased. If this is right, our tendency to scrutinize other people's views to a greater degree than our own views and the coherence motives generally need also not always be biases.

At this point, Haidt may accept that in theory, the relatedness and the coherence motives are not always biases, but he may say that in practice, these motives always are biases. This is an empirical point. As far as I am aware, existing research on the relatedness and coherence motives has not tried, e.g., to assess whether our tendency to be motivated to agree with our friends is due to the fact that we want to be harmonious with our friends or due to the fact that we tend to trust our friends' judgments in general. Nor has this research asked these people about their background beliefs, which, as we have seen, can be relevant for determining whether they are epistemically justified in their judgments or not. At the very least, it seems that future research in this area should control for these possibilities. Until then, it seems premature to claim that reasoning is biased and that the purpose of reasoning is to provide a biased basis for justification.

8. Conclusion

The relatedness and the coherence motives are Haidt's basis for believing that reasoning is biased. In this paper, I argued that Haidt simply assumes that the relatedness and coherence motives are biases. Drawing on the literature on epistemic justification, I showed that these motives are not always biases. Among other things, the relatedness motives need not be biases because we may have the tendency to be motivated to agree with our friends because we tend to trust our friends' judgments in general, which can be an epistemically rational response. And, among other things, the coherence motives need not be biases because our epistemically justified background beliefs can play a significant role in determining the way we respond to new evidence. If this is correct, Haidt has not established that our reasoning is biased. If so, this also calls into question the aspect of the SIM, according to which the purpose of reasoning is to provide a biased basis for justification. And if all of this is right, it is premature to draw firm conclusions about how the SIM may undercut the evidential weight of our moral judgments.

References

- Alston, W. P. 1985: Concepts of Epistemic Justification. *Monist*, 68, 57-89.
- Bishop, M. A. and Trout, J. D. 2005: *Epistemology and the psychology of human judgment*, New York: Oxford University Press.
- Chaiken, S., Giner-Sorolla, R. and Chen, S. 1996: Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In Gollwitzer, P. M. and Bargh, J. A. (eds) *The psychology of action: Linking cognition and motivation to behavior*. New York: Guilford Press.

- Chen, S., Shechter, D. and Chaiken, S. 1996: Getting at the truth or getting along: Accuracy- versus impression-motivated heuristic and systematic processing. *Journal of Personality and Social Psychology*, 71, 262-275.
- Clarke, S. 2008: SIM and the City: Rationalism in Psychology and Philosophy of Haidt's Account of Moral Judgment. *Philosophical Psychology*, 21, 799-820.
- Fine, C. 2006: Is the emotional dog wagging its rational tail, or chasing it? *Philosophical Explorations*, 9, 83-98.
- Gibbard, A. 2008: *Reconciling our aims: In search of bases for ethics*, New York: Oxford University Press.
- Gigerenzer, G. and Todd, P. M. 1999: *Simple Heuristics that Make Us Smart*, Oxford: Oxford University Press.
- Greene, J. and Haidt, J. 2002: How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6, 517-523.
- Haidt, J. 2001: The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.
- Haidt, J. 2003: The emotional dog does learn new tricks: A reply to Pizarro and Bloom. *Psychological Review*, 110, 197-98.
- Haidt, J. 2006: *The Happiness Hypothesis*, London: Arrow Books.
- Haidt, J. 2007: The new synthesis in moral psychology. *Science*, 316, 998-1002.
- Haidt, J. and Bjorklund, F. 2007a: Social intuitionists answer six questions about moral psychology. In Sinnott-Armstrong, W. (ed) *Moral psychology, Vol. 2: The cognitive science of morality: Intuition and diversity*. Cambridge MA: MIT Press.

- Haidt, J. and Bjorklund, F. 2007b: Social intuitionists reason, in conversation. In Sinnott-Armstrong, W. (ed) *Moral psychology, Vol. 2: The cognitive science of morality: Intuition and diversity*. Cambridge MA: MIT Press.
- Haidt, J., Bjorklund, F. and Murphy, S. 2000: Moral dumbfounding: When intuition finds no reason. *Unpublished Manuscript*.
- Haidt, J. and Hersh, M. A. 2001: Sexual morality: the cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31.
- Kennett, J. and Fine, C. 2009: Will the Real Moral Judgment Please Stand Up? *Ethical Theory and Moral Practice*, 12, 77-96.
- Kerr, N. L., Maccoun, R. J. and Kramer, G. P. 1996: Bias in judgment: Comparing individuals and groups. *Psychological Review*, 103, 687-719.
- Kohlberg, L. 1969: Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, D. A. (ed) *Handbook of socialization theory and research*. Chicago: Rand McNally.
- Kunda, Z. and Sanitioso, R. 1989: Motivated changes in the self-concept. *Journal of Experimental Social Psychology*, 25, 272-85.
- Levy, N. 2006: The wisdom of the pack. *Philosophical explorations*, 9, 99-103.
- Lord, C. G., Ross, L. and Lepper, M. R. 1979: Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098-2109.
- Mahoney, M. 1977: Publication prejudice: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1, 161-75.

- Nisbett, R. and Ross, L. 1980: *Human inference: Strategies and shortcomings of social judgment*, Englewood Cliffs, NJ: Prentice Hall.
- Piaget, J. 1932: *The moral judgement of the child*, New York: Harcourt, Brace Jovanovich.
- Pizarro, D. A. and Bloom, A. 2003: The intelligence of the moral intuitions: Comment on Haidt (2001). *Psychological Review*, 110, 193-196.
- Rawls, J. 1971: *A Theory of Justice*, Oxford: Oxford University Press.
- Saltzstein, H. D. and Kasachkoff, T. 2004: Haidt's moral intuitionist theory: a psychological and philosophical critique. *Review of General Psychology* 8, 273-282.
- Sinnott-Armstrong, W. 2006: Moral Intuitionism Meets Empirical Psychology. In Horgan, T. and Timmons, M. (eds) *Metaethics After Moore*. New York: Oxford University Press.
- Smolin, L. 2006: *The Trouble with Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next*, London: Allen Lane.
- Trope, Y. and Liberman, A. 1996: Social hypothesis testing: Cognitive and motivational mechanisms. In Higgins, E. T. and Krugalski, A. W. (eds) *Social psychology: Handbook of basic principles*. New York: Guilford.
- Wilson, T. and Brekke, N. 1994: Mental Contamination and Mental Correction: Unwanted Influences on Judgements and Evaluations. *Psychological Bulletin*, 116, 117-142.