

7-18-2008

The Non-moral Basis of Cognitive Biases of Moral Intuitions

Bradley Charles Thomas
bradley-thomas@uiowa.edu

Follow this and additional works at: http://digitalarchive.gsu.edu/philosophy_theses

Recommended Citation

Thomas, Bradley Charles, "The Non-moral Basis of Cognitive Biases of Moral Intuitions" (2008). *Philosophy Theses*. Paper 44.

This Thesis is brought to you for free and open access by the Department of Philosophy at Digital Archive @ GSU. It has been accepted for inclusion in Philosophy Theses by an authorized administrator of Digital Archive @ GSU. For more information, please contact digitalarchive@gsu.edu.

THE NON-MORAL BASIS OF COGNITIVE BIASES OF MORAL INTUITIONS

by

Bradley Charles Thomas

Under the direction of Eddy Nahmias

ABSTRACT

Against moral intuitionism, which holds that moral intuitions can be non-inferentially justified, Walter Sinnott-Armstrong argues that moral intuitions are unreliable and must be confirmed to be justified (i.e. must be justified inferentially) because they are subject to cognitive biases. However, I suggest this is merely a renewed version of the argument from disagreement against intuitionism. As such, I attempt to show that the renewed argument is subject to an analogous objection as the old one; many cognitive biases of moral intuitions result from biases of non-moral judgments. Thus, the unreliability of moral intuitions due to biases (and the reason inferential justification was required) can be removed by clearing up the non-moral biases. Accordingly, biases of moral intuitions do not threaten a slightly qualified version intuitionism which posits non-inferential justification of intuitions when non-moral biases are not present. I also present an empirical study that lends initial support to my argument.

INDEX WORDS: Moral intuitionism, Moral Epistemology, Metaethics, Moral Psychology, Experimental Philosophy

THE NON-MORAL BASIS OF COGNITIVE BIASES OF MORAL INTUITIONS

by

Bradley Charles Thomas

A Thesis Submitted in Partial Fulfillment of Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2008

Copyright by
Bradley Charles Thomas
2008

THE NON-MORAL BASIS OF COGNITIVE BIASES OF MORAL INTUITIONS

by

Bradley Charles Thomas

Committee Chair: Eddy Nahmias

Committee: Andrew Altman
Stephen Jacobson
Andrea Scarantino

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2007

DEDICATION

To my grandparents, Poppy, Lulu, and Grandma and Grandpa Thomas—beautiful people from whom I learned more than they or I will ever know. Also, I have dedicated much of my work on this thesis to the memory of a wonderful mother, neighbor, and friend, Mrs. Loblaw, and to her family.

ACKNOWLEDGMENTS

Thanks to my committee chair, Eddy Nahmias, for his sincere investment and belief in me as a student. Many of the ideas in this thesis, including its central point of objecting to the argument from cognitive biases, were developed in conversations and emails with him, and I am grateful for his guidance and attentive ear. His advisement, encouragement, and support, even in the midst of his own pursuits and commitments, helped me get the most out of my time at Georgia State. Thanks also to my committee for working through the thesis on very short notice. Andrew Altman's input on the history of moral intuitionism and the potential strengths and weaknesses of my arguments brought another level of insight and clarity to the thesis. The knowledge of Stephen Jacobson on the epistemological points was extremely helpful. Andrea Scarantino's ability to discern the weaknesses in my exposition, as always, made this a much better piece of philosophical work. Many thanks to my committee for generously offering their time to work with me.

Thanks also to my love, Sarah Taber, for waiting for me while I worked on this thesis. I owe a massive debt of gratitude to my mom, dad, and sister for teaching me to shoot for the moon, and for keeping me honest. The rest of my family and my many great friends deserve much credit as well, especially Todd Hester, who inspires my intellectual curiosities more than anyone.

In the end, everything of value in this thesis is owed to these people and the many others from whom I have drawn knowledge, inspiration, and conversation, and all of the shortcomings are owed to me.

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGMENTS	v
Introduction	1
1 Skepticism and intuitionism	5
2 The argument from cognitive biases.....	11
2.1 <i>Confirmation of moral intuitions justifies inferentially</i>	14
2.2 <i>Cognitive biases create the need for confirmation</i>	15
2.3 <i>The empirical picture so far</i>	19
2.3.1 Tversky and Kahneman.....	19
2.3.2 Petrinovich and O’Neill	21
2.3.3 Haidt and Baron	23
2.3.4 Nadelhoffer and Feltz.....	24
2.3.5 The status of premise C4.....	25
2.4 <i>Rejecting the internalist bias</i>	30
3 The argument from disagreement.....	31
3.1 <i>Non-moral basis of moral disagreement</i>	33
4 Non-moral basis of cognitive biases of moral intuitions	40
4.1 <i>Revised argument from cognitive biases</i>	41
4.2 <i>Non-moral basis of biases</i>	42
4.3 <i>An initial objection and a response</i>	48
5 Empirical support for the intuitionist’s reply.....	50
5.1 <i>Methods</i>	51
5.2 <i>Results and analyses</i>	53
5.3 <i>Discussion</i>	54
6 Conclusions	59
References	61
Appendices	63
<i>Appendix A: Regression Statistics</i>	63
<i>Appendix B: Moral Dilemmas</i>	64

Introduction

Torture. Incest. Abortion. Honesty. The list of things that evoke our moral intuitions seems endless. Many, like incest, seem obviously wrong to all of us. And surely honesty, in general, is good. But people have diverging intuitions about torture and abortion. Our various moral intuitions are an important part of our everyday experience. We form moral intuitions when we read the news, watch our favorite television program or movie, and interact with, or even just observe those around us. These moral intuitions seem to occur without warning, without reflection, and without our control; yet we generally trust their accuracy without question. We have the strong impression that our moral intuitions track the moral truth, even though they often conflict with the intuitions of other people.

But what is the proper epistemic role of such moral intuitions? Are they justified? Ought we accept our intuitions so easily, or at all? One answer is given by the thesis of moral intuitionism (intuitionism, for short), which is a version of foundationalism in moral epistemology. Intuitionism holds that some moral intuitions, qua moral beliefs, are justified non-inferentially; that is, “some believers are justified in holding some moral beliefs independently of whether the believer is able to infer those moral beliefs from any other beliefs” (Sinnott-Armstrong 2006a, 185). The intuitionist believes that moral intuitions provide the foundation for moral knowledge, the bedrock from which other moral beliefs can be justified.

This thesis has recently come under fire by what I will call “the argument from cognitive biases” put forth by Walter Sinnott-Armstrong, who wields evidence of cognitive biases of moral intuitions against the intuitionist (2006a; 2006b; 2008a; 2008b). A cognitive bias occurs when a cognitive process (like making a decision or forming a judgment, belief, or intuition) is affected by purportedly irrelevant properties of the situation being judged. For example, framing effects

are one type of cognitive bias whereby “decision makers respond differently to different but objectively equivalent descriptions of the same problem”—just changing the wording of a question, without changing the content of the question itself, can make a person form different beliefs, and respond differently (Levin et al. 1998, 150). When a cognitive process is biased in this way, it is not reliable. As Sinnott-Armstrong notes, this “follows from the very idea of cognitive biases” (2008a, 52). Sinnott-Armstrong argues that since many moral intuitions are highly susceptible to cognitive biases, moral intuitions are unreliable. Some moral intuitions or types of moral intuitions may be reliable, even though moral intuitions in general are unreliable (just as some Fords may start reliably, even though Fords in general may be unreliable if many do not start reliably). But for Sinnott-Armstrong, if a class of intuitions is unreliable, then one is justified in accepting a given intuition from that class just in case she can confirm the intuition. Imagine that our intuitions about Fords starting are unreliable—perhaps we are often convinced by Ford’s advertisements that their cars will start, when in fact they often do not. If we ought to know that these intuitions are unreliable, Sinnott-Armstrong would suggest that one is justified in her intuition that a given Ford will start just in case she can confirm that intuition. That is, in case she can confirm that the Ford will start, perhaps by providing evidence that she has started it many times before without incident. And confirmation of an intuition justifies the intuition inferentially; her intuition is justified because she can infer the belief that the Ford will start from the evidence that she has started it many times before. Since confirmation of an intuition justifies the intuition by making a person able to *infer* the intuition from the confirming evidence, confirmation provides inferential justification (see Section 2.1). Thus, the argument from cognitive biases goes, since moral intuitions are unreliable and accordingly require moral confirmation (i.e. confirmation of the moral intuitions) to be justified, they cannot be non-

inferentially justified. Therefore, intuitionism, which holds that some moral intuitions are justified non-inferentially, is false.

The argument from cognitive biases is a recent line of argument against intuitionism, and moral intuitionists have yet to offer an adequate reply. In this thesis I will not aim to defend moral intuitionism against rival moral epistemological theories; rather, I aim only to point out that the argument from cognitive biases involves a hidden premise, which is false. There may be good reasons to reject intuitionism, but the argument from cognitive biases is not one of them, and I will attempt to show why here. I will resurrect an intuitionist response to an old argument against intuitionism, the argument from disagreement, and adapt that response to the argument from cognitive biases (Mackie 1977; Loeb 1998; Sinnott-Armstrong 2002, 2006a, 2006b). I will show that the argument from cognitive biases requires the plausibly false premise which states that the cognitive biases (and unreliability) of moral intuitions are not grounded in cognitive biases (and unreliability) of non-moral judgments. By non-moral judgments, I mean people's relevant perceptions or representations of the non-moral facts of the situation—which can be conscious or represented in subconscious processes. They are relevant in that they are about the non-moral facts that affect the formation of moral intuitions about the situation (like judgments about future consequences). Biases of these non-moral judgments may be caused by subtle changes in the salience of certain non-moral facts, which the subject is not even aware of. Also, biases of non-moral judgments may be short lived and easily changed (e.g. when the subject focuses her attention on different information).

I will suggest that cognitive biases, like framing effects, influence people's moral intuitions largely because they influence such non-moral judgments; that is, framing effects influence people's judgments about the non-moral facts of the case. The idea is that most

differences in a subject's moral intuitions about different, but objectively equivalent presentations of the same situation are due to differences in her non-moral judgments about the different presentations of the situation. And as Sinnott-Armstrong notes in the context of the argument from disagreement, "when disagreements about moral issues arise solely from non-moral disagreements and conceptual confusions that could be removed by further inquiry and reflection, informed and rational people would not disagree" (2006a, 199). This same principle applies to the argument from cognitive biases, in that when differences in a person's moral intuitions are grounded in differences in her non-moral judgments, resolving the difference in non-moral judgments would resolve the difference in moral intuitions (i.e. would remove the cognitive biases of moral intuitions). In this case, moral confirmation is not required for one's moral intuitions to be justified. Since non-moral confirmation removes the biases of moral intuitions, such non-moral confirmation can allow one's moral intuitions to be justified. And, unlike moral confirmation, non-moral confirmation does not threaten intuitionism because one cannot infer her moral intuitions from non-moral confirmation (on pain of inferring an ought from an is). Thus, if the cognitive biases and unreliability of moral intuitions are grounded in cognitive biases and unreliability of non-moral judgments, then the biases of moral intuitions can be eliminated and the original reason for thinking those intuitions were unreliable is gone. Accordingly, by confirming the relevant non-moral judgments, our moral intuitions could be justified *without making those intuitions inferable*. Thus, our moral intuitions can be justified non-inferentially (barring some reason other than cognitive biases to question their reliability). The argument from cognitive biases does not threaten the moral intuitionist who claims that moral intuitions can be justified non-inferentially when biases of non-moral judgments do not occur.

The empirical research cited by the argument from cognitive biases to show that moral intuitions are unreliable leaves open the possibility that the biases of moral intuitions are grounded in biases of subjects' non-moral judgments. And I will argue that, to a significant extent, this is the case—many biases of moral intuitions are grounded in biases of non-moral judgments. I will also present a study I conducted in collaboration with Eddy Nahmias that supports my claim.

In Section 1, I will motivate the topic by introducing the thesis of intuitionism. Then, in Section 2, I will review the argument from cognitive biases. In Section 3, I will show how it parallels the argument from disagreement against intuitionism, and highlight the intuitionist's response from the non-moral basis of disagreement, which holds that many moral disagreements are grounded in non-moral disagreements. Then, in Section 4, I adapt the intuitionist's response from the non-moral basis of disagreement to respond to the argument from cognitive biases—offering a response from the non-moral basis of cognitive biases of moral intuitions. In Section 5, I will discuss an empirical study Eddy Nahmias and I conducted that weighs in favor of my response. I offer some concluding remarks in Section 6.

1 Skepticism and intuitionism

Moral epistemology is the branch of metaethics that studies moral knowledge, beliefs, and justification. The central problems facing the moral epistemologist are whether and how moral claims can be known and moral beliefs can be justified. For our discussion of intuitionism, issues about the justification of moral beliefs will be central, so we will forgo a discussion of moral knowledge. In a recent paper, Walter Sinnott-Armstrong explained justification of moral beliefs quite simply: a moral belief is justified when “the believer ought to hold that belief as opposed to suspending belief, because the believer has adequate epistemic grounds for believing

that it is true” (2008a, 48).¹ Much of the difficulty comes in specifying what epistemic grounds count as “adequate”. But intuitionism is a thesis designed to respond to a more fundamental problem with the justification of moral beliefs raised by the skeptic: perhaps no moral belief is ever justified because of a vicious skeptical regress. The skeptical regress gets started by assuming that the justification of moral beliefs must be inferential:

Inferential justification—S is inferentially justified in believing B if, in order to be justified in believing B, S must be able to infer B from another justified belief, B'.²

The problem here is that if justification must be inferential, then S can only be justified in believing B', the belief which justifies her belief in B, if S has the ability to infer B' from another justified belief, B''. And S must be able to infer B'' from another justified belief B''', and so on. In this way, on the assumption that moral beliefs can only be justified inferentially, the justification of moral beliefs falls into a vicious regress.

The moral intuitionist’s answer to this regress is, as Sinnott-Armstrong aptly puts it, “simply to stop” (2006a, 184). Intuitionism flatly rejects the assumption that moral beliefs can only be justified inferentially, and thus holds that at least some moral beliefs can be justified non-inferentially:

Non-inferential justification—S is non-inferentially justified in believing B if, in order to be justified in believing B, S does not need to be able to infer B from any other belief.

Although S might have the ability to infer B from B', S is non-inferentially justified in believing B when she is justified independently of her ability to infer B from other beliefs. As Brad Hooker

¹ There certainly are alternative ways to understand justification, and one might object to Sinnott-Armstrong’s argument from cognitive biases by disagreeing with him at this early stage. I think, however, that Sinnott-Armstrong’s argument fails on a more fundamental level, in that it requires a false hidden premise, and thus I will largely follow Sinnott-Armstrong’s understanding of the relevant concepts.

² It is important to note that, according to Sinnott-Armstrong, to be justified in believing B inferentially, S does not actually need to make the inference of B from some other belief; S needs only to be *able* to draw this inference. The idea is that “the requisite information [must] be encoded somehow” in S’s brain at the right time (Sinnott-Armstrong 2006a; Hooker 2002).

puts it, “a moral belief can be justified by some feature other than its inferential relations with other beliefs” (2002, 163).³ Intuitionists differ in their theories of what that “other”, justifying feature of moral intuitions is (reliabilism, Shafer-Landau 2003; moral experientialism, Tolhurst 1990; reflectionism, Audi 2004). But all intuitionists share the thesis that certain moral beliefs can be justified (in some way) non-inferentially. This general thesis will be the topic of the present discussion. Such intuitionism is a version of foundationalism in moral epistemology; the moral beliefs that are justified non-inferentially provide a foundation for the inferential justification of other moral beliefs. In this way, epistemological moral intuitionism provides a stopping point for the skeptical regress—it ends at the non-inferentially justified moral beliefs. The belief B might be justified by being inferred from the justified belief B', but if B' is justified non-inferentially, then the regress has lost its viciousness.

As Sinnott-Armstrong (2006a) notes, intuitionists also disagree about the content of the beliefs that can be justified non-inferentially: are they about what is good (Moore 1903), or what is right (Ross 1930)? And they disagree about the generality of the beliefs: are they beliefs about abstract principles (Sidgwick 1907), generalizations (Ross 1930), or particular actions (Prichard 1968)? Still, they agree that the non-inferentially justified beliefs are, unsurprisingly, moral intuitions. To be precise, I will follow Sinnott-Armstrong in defining a moral intuition as “a strong, immediate moral belief” (2008a). By ‘strong’, Sinnott-Armstrong means that one would not easily give up the belief; by ‘immediate’, he means that the belief is formed non-inferentially. That is, “a belief’s status as an intuition consists in its being arrived at not by

³ I may sometimes refer to a *belief* being justified, but this is only shorthand for saying that the *believer* is justified in holding the belief (Sinnott-Armstrong 2006a).

inference from other moral beliefs” (Hooker 2002, 162).⁴ The idea is that a moral belief is an intuition when it is not formed based on any actually performed inference, regardless of whether the believer is able to infer the moral belief from other (moral or non-moral) beliefs. So a given intuition may be “accessible via inference”—that is, one might, in fact, be able to infer it from other beliefs—but moral intuitions in general need not bear inferential relations to other beliefs (Hooker 2002, 162). They are intuitions just in virtue of their not being formed inferentially.

A point worth noting that seems to have been largely ignored in the contemporary discussion of epistemological moral intuitionism is that this definition of ‘moral intuition’ is quite different from the one offered by some classic intuitionists, like G.E. Moore.⁵ So we are right to ask how these notions of intuitions differ, and how they might be related. Moore thought of a moral intuition more generally as a non-inferentially *justified* (not *formed*) moral belief (I will call such moral intuitions “Moorean moral intuitions”; 1903). In *Principia Ethica*, he is quite explicit in this, saying “when I call such propositions ‘Intuitions,’ I mean *merely* to assert that they are incapable of proof; I imply nothing whatever as to the manner or origin of our cognition of them” (1903). As you can see, in contrast to the definition of intuition above, Moore clearly meant to imply nothing about the way the beliefs or propositions are formed in calling them intuitions. A Moorean moral intuition could be formed via inference, or not—but it certainly was not *provable* or *justifiable* via inference. But Moore then needs to say just which (if any) moral beliefs can be non-inferentially justified: Which moral beliefs are or can be Moorean moral intuitions? The contemporary definition of moral intuition differs from Moore’s in that it builds an answer to this question into the definition. The moral beliefs that can be Moorean moral

⁴ It is not particularly relevant to our discussion, but Hooker suggests this immediacy (being formed or “arrived at” non-inferentially) is all there is to being an intuition. Since we will be discussing Sinnott-Armstrong’s argument on his own terms, we will follow him in adding that the belief must be ‘strong’.

⁵ Andrew Altman pointed out this oversight to me.

intuitions are the ones that are formed non-inferentially, and those are what are currently being called moral intuitions (Sinnott-Armstrong 2008a; Hooker 2002). Moore certainly objected to this answer to the question of which moral beliefs can be Moorean moral intuitions. He thought it was not just any moral belief that is formed non-inferentially that could count as a Moorean moral intuition (i.e. a moral intuition that can be justified non-inferentially), but rather only those moral beliefs that are formed after due reflection and with sufficient expertise. Specifically, Moore wanted to talk about the moral beliefs or propositions offered by the “ethical philosopher”—that is, the ‘moral expert’ (1903; see specifically the Preface).

But for epistemological moral intuitionism, there is a serious flaw in this attempt to identify which moral beliefs can be Moorean moral intuitions, or can be justified non-inferentially. The problem is that expert moral beliefs, such as the belief that M (some expert moral belief), can be justified inferentially based on the fact that they are expert. That is, the expert belief that M can be justified by inferring it from some reason that expert beliefs are epistemically superior (like the fact that they are more reliable than other beliefs), and that an expert believes that M. And thus, expert beliefs as a class of moral beliefs defined by their being ‘expert’ cannot form the class of Moorean moral intuitions. The expert class of moral beliefs is distinguished by its being superior epistemically, and this fact makes the beliefs it contains inferentially justifiable (from the fact that they are part of an epistemically superior class).

One will always need to offer some reason that justifies distinguishing expert beliefs from other beliefs as the beliefs that can be Moorean moral intuitions (i.e. the beliefs that can be justified non-inferentially), like the fact that they are reflected upon more closely, that they are more reliable, or that they are more rational. But whatever the reason, it will make one able to inferentially justify the expert beliefs. Indeed, the reason *needs* to make one able to inferentially

justify the expert beliefs, or else it would not be a good reason to distinguish those beliefs as potential Moorean moral intuitions. Imagine if the above reason that expert beliefs are superior was not that they are more reliable, but that they sound cooler; this would not be a good reason to think that expert beliefs are Moorean intuitions, and it also would not provide inferential justification of the expert beliefs. One could not inferentially justify an expert's belief that M by reasoning from the fact that it sounds cooler. But if sounding cool *was* a good reason to think expert beliefs are Moorean intuitions, then, indeed, sounding cool could provide inferential justification of expert beliefs by plugging the reason (sounding cool) into the inferential equation above; the expert belief that M can be justified by inferring it from the fact that expert beliefs are epistemically superior because they sound cooler, and that an expert believes that M. The point is that any good epistemic reason to distinguish a subset of moral beliefs as candidates for Moorean moral intuitions also provides inferential justification of that subset of beliefs, and thus precludes the subset from being the class of Moorean moral intuitions. A given expert moral belief may be a Moorean moral intuition—but no subset of moral beliefs can be justifiably identified on epistemic grounds, such as their being expert, as being the class of Moorean moral intuitions. To put it another way, some expert beliefs may be Moorean moral intuitions, but this cannot be in virtue of their being “expert,” as such an epistemic virtue makes them inferentially justifiable (and thus *not* Moorean intuitions).

It seems to me that it is for this reason that contemporary intuitionists in moral epistemology define moral intuitions as moral beliefs that are formed non-inferentially—and hold that these are the candidates for Moorean moral intuitions. By defining moral intuitions in this way, they can avoid the folly of offering an epistemic reason for identifying a certain class of moral beliefs as Moorean moral intuitions.

2 The argument from cognitive biases

In a recent paper, Walter Sinnott-Armstrong offers what I have called “the argument from cognitive biases” against intuitionism, which I will suggest is merely a renewed version of the argument from disagreement (2008a). Sinnott-Armstrong wields evidence from empirical psychology suggesting that moral intuitions are subject to cognitive biases to argue that moral intuitions are unreliable, must be confirmed to be justified, and can only be inferentially justified—and thus that intuitionism is false (2008a). I will discuss the parallel with the argument from disagreement in more detail later, but for now, notice that whereas the argument from disagreement argues that moral intuitions are unreliable on the basis of *between*-subject differences in moral intuitions (i.e. moral disagreements between different people or groups), the argument from cognitive biases argues that moral intuitions are unreliable on the basis of *within*-subject differences in moral intuitions. That is, it argues on the basis of cognitive biases which make it likely that a single person will have inconsistent intuitions about different, but objectively equivalent or relevantly similar moral situations. (I will refer to between-subject disagreement simply as moral disagreement, and within-subject disagreement as cognitive biases.)

To be precise, I define a cognitive bias as a tendency for a cognitive process, like the formation of moral intuitions, to be affected by purportedly irrelevant properties of a stimuli; and this susceptibility to modification by irrelevant factors reveals the cognitive process to be unreliable. For example, a cognitive bias might make it likely that a person will have different moral intuitions about effectively equivalent moral scenarios or dilemmas presented using

different words.⁶ Take Sinnott-Armstrong's example of Joseph, who would believe that Marion is fast when told that she runs one hundred meters in ten seconds, but would believe that she is *not* fast (and is slow) when told that she runs one hundred meters in ten thousand milliseconds (2008a). Since Joseph would form such drastically different beliefs about running speed in objectively equivalent situations, his beliefs about (Marion's) running speed, in general, are not reliable. Perhaps his beliefs about 'running speed in seconds' are reliable, and only those about 'running speed in milliseconds' are unreliable (maybe because he thinks that one millisecond is one thousand seconds, rather than one *thousandth* of a second). Nonetheless, his beliefs about running speed, in general, are unreliable (since beliefs about 'running speed in milliseconds' are inaccurate and are part of the class of his beliefs about running speed). To show the unreliability of moral intuitions, Sinnott-Armstrong cites a variety of framing effects, cognitive biases like the one just described in which a subject's moral intuitions are affected by the wording used to describe a moral dilemma or by the order in which dilemmas are presented (2008a). Extending Sinnott-Armstrong's argument, Thomas Nadelhoffer and Adam Feltz argue that moral intuitions are unreliable not only because they are subject to framing effects, but also because they are subject to another well-known cognitive bias, the self-other asymmetry, whereby people's intuitions about a scenario are affected by whether the scenario is presented in the first or third person context (in prep.; Malle 2006).

When these sorts of cognitive biases of moral intuitions occur, there is essentially a within-subject disagreement in moral intuitions. The same subject has inconsistent intuitions

⁶ It is important to note that biases make it *likely* that a given person's intuitions will be affected by irrelevant factors. This means that any given person's intuitions might be completely unaffected by irrelevant factors; the bias merely makes it *likely* that her intuitions will be so affected.

about different presentations of objectively equivalent scenarios. Assuming non-relativism,⁷ at least one of the inconsistent intuitions must be wrong, and thus at least some moral intuitions are fallible. If such cognitive biases, or within-subject moral disagreements, are sufficiently prevalent, then the reliability of moral intuitions as a class of beliefs would be called into question, since that class would include all of the inconsistent pairs of beliefs (half of which are mistaken); one could not rely on her moral intuitions (i.e. could not say that her moral intuitions reliably track the moral truth), since so many of them would be mistaken.

Sinnott-Armstrong argues that if moral intuitions are unreliable due to cognitive biases in this way, then moral intuitions would require confirmation to be justified—since without confirmation there would be no way to know if a given intuition is one of the accurate or inaccurate ones in the class of moral intuitions (2006a; 2008a; 2008b). The problem is that confirmation of a moral intuition can justify that intuition only on inferential grounds. This is because the confirmation justifies the intuition by making the person able to infer her intuition from the confirming evidence (recall the Ford example discussed in the Introduction; see Section 2.1). Thus, if cognitive biases of moral intuitions (i.e. within-subject moral disagreement) are sufficiently prevalent to bring the reliability of moral intuitions into question, moral intuitionism is false; moral intuitions would require confirmation to be justified, and thus cannot be justified non-inferentially. To put it explicitly, the argument from cognitive biases looks like this:⁸

C1) If moral intuitions are sufficiently affected by cognitive biases to show that moral intuitions are, in general, unreliable, then moral intuitions must be confirmed in order to be justified

⁷ I will be objecting to Sinnott-Armstrong's argument from cognitive biases while granting his non-relativist moral realism, which holds that there are objective (non-relativist) moral facts. He says, "I believe that many acts are morally wrong. I think that my positive moral beliefs are true and correspond to moral facts" (2006a, 58). I am aware the one could object to his arguments by rejecting these assumptions; but I believe his argument can be objected to even granting his assumptions.

⁸ This is my formulation of the argument, reconstructed from Sinnott-Armstrong 2008a; 2008b; 2006a. I believe it accurately represents Sinnott-Armstrong's main argument provided in 2008a in light of the expanded version of the argument presented in 2008b and his views on confirmation made explicit in 2006a (see especially Chapter 9).

(because without confirmation, a believer is not justified in believing her intuitions were formed reliably rather than unreliably, and thus she is not justified in believing those intuitions)

- C2) If moral intuitions require confirmation in order to be justified, then they can only be justified inferentially.
- C3) If moral intuitions can only be justified inferentially, then moral intuitionism is false.
- C4) Moral intuitions are sufficiently affected by cognitive biases to show that moral intuitions are, in general, unreliable.⁹
- C5) Moral intuitionism is false.

Most of the action occurs in premises C1, C2, and C4. Premise C3 is straightforward, since intuitionism just claims that some moral intuitions can be justified non-inferentially. In Section 2.1 I will discuss why, as premise C2 suggests, confirmation of moral intuitions provides only inferential justification of those intuitions. In section 2.2 I will look at premises C1 and C4 in more detail. In Section 2.3 I will examine exactly why Sinnott-Armstrong thinks that premise C4 is true, reviewing the evidence of cognitive biases of moral intuitions, and I will assess the status of C4 in Section 2.3.5. In 2.4 I discuss a potential initial objection which rejects the argument's internalist bias. In Section 3, I will discuss the intuitionist's objection to the argument from disagreement—that most moral disagreement is grounded in non-moral disagreement—which I adapt as a response to the argument from cognitive biases in Section 4.

2.1 Confirmation of moral intuitions justifies inferentially

Premise C2 above says that if moral intuitions must be confirmed in order to be justified, then they can only be justified inferentially. This is because, on Sinnott-Armstrong's understanding, confirmation of a belief provides only inferential justification of that belief (2006a). Confirmation makes a person justified in holding a belief because she can *infer* her belief from the evidence that confirmed it. For example, say that Mary believes that Andy

⁹ The intuitions are not just unreliable, but *we ought to know that they are unreliable*. This just means that we ought to be aware of the truth of C4—i.e. of the prevalence of cognitive biases.

Pettitte, who plays baseball for the New York Yankees, took steroids (call this belief S), and she believes S because a rival Red Sox fan told her so. Her belief came from a biased, unreliable source and she knows it, so she is not justified in believing S without confirmation of S. When Mary sees Pettitte on ESPN later that evening confirming his steroid use, she is now justified in believing S, that Pettitte took steroids. And the key is that this confirmation justifies Mary's belief that S *inferentially*; she is justified because after receiving confirmation of S, she now has the ability to infer S from that confirmation. That is, Mary can infer S, that Pettitte took steroids, from the fact that he admitted taking steroids. In this way, confirmation of a belief provides justification of that belief just because a believer can infer the belief from the confirming evidence.

2.2 Cognitive biases create the need for confirmation

If moral intuitions require confirmation to be justified, then since confirmation provides only inferential justification, moral intuitions could be justified only inferentially. Premise C1 states that sufficient cognitive biases of moral intuitions would create a need for such confirmation of moral intuitions. Sinnott-Armstrong notes that “confirmation is needed for a believer to be justified in holding a belief that other people deny or doubt, when the believer has no reason to prefer one believer to the other” (2006a, p. 193). His point here about *between*-subject disagreement creating the need for confirmation applies analogously to within-subject disagreement (like cognitive biases) creating the need for confirmation. A believer cannot be justified in believing one intuition over an incompatible one without some reason or confirmation that supports one or the other intuition, because there would be no way to know which intuition ought to be relied on (Smythe and Evans 2007). And this is the case whether the incompatible intuitions are from two different people, or the same person under different

conditions as in the case of cognitive biases. For example, imagine a between-subject disagreement between Mary and Patty. While Mary believes S (that Andy Pettitte took steroids) Patty believes $\sim S$ (that Andy Pettitte did not take steroids). Both women are equally informed, having the same evidence, but they have based their differing beliefs on different character assessments of Brian McNamee (the trainer who accused Pettitte of taking steroids). Mary thinks McNamee is trustworthy, while Patty thinks he is a liar. Without any confirmation of one of their beliefs to suggest one belief over the other, neither can be justified. They would need to know who is a more reliable judge of character, or have some other confirmation of one of their beliefs, in order for that belief to be justified. In this way, inconsistency in beliefs creates a need for confirmation of those beliefs.¹⁰

In premise C4, however, inconsistency in moral beliefs due to cognitive biases creates the need for the whole class of moral intuitions to be confirmed, not just some particular moral intuitions. For biases to create the need to confirm a whole class of beliefs, the biases must be sufficiently prevalent to bring the reliability of that class of beliefs into question. To illustrate the point, consider Sinnott-Armstrong's thermometer analogy (2002). Imagine you have a hundred very old thermometers, and you know that some of them are inaccurate. If a sufficient amount of them are inaccurate, then you are not justified in trusting the temperature readings of any one thermometer until you confirm whether it is accurate or inaccurate—this group of thermometers would be unreliable. The same goes for classes of beliefs; if a sufficient amount of a certain class of beliefs is mistaken or unreliable, then nobody is justified in holding a belief from that class

¹⁰ It should be clear that not every disagreement is one that creates a need for *new* or *further* confirmation. A disagreement that persists between a well-informed believer and an ill-informed believer has a need for confirmation or a reason to prefer one believer or the other (since there is a disagreement), but that need has been met by the fact that one believer is better informed than the other. The disagreement still creates the need for confirmation, but that need has been readily met. A reason to prefer one belief over another can be thought of as a type of confirmation of the belief. So confirmation is not necessary when one has a reason to prefer one belief over another because confirmation has already been provided by the reason.

without confirming that the belief is one of the accurate ones in the class—the class is unreliable. The amount that must be mistaken in order to demonstrate unreliability will vary depending on what is being shown to be unreliable, and what the risks are of being mistaken. For the thermometers in this case, there does not seem to be any great risk if we are mistaken about the temperature, so the amount of thermometers that must be inaccurate to show that the thermometers, as a group, are unreliable and require confirmation is probably pretty high—say around 15 percent of the thermometers. In the case of moral beliefs, the stakes are higher, so the percent of moral intuitions that must be mistaken to show that moral intuitions, as a class of beliefs, are unreliable and require confirmation is probably rather low—Sinnott-Armstrong suggests around only 5 percent (2008b). If 5 percent are inaccurate, then we need to confirm a given intuition for it to be justified.

Consider a world in which half of the population believes that using water boarding to torture terrorism suspects is generally morally acceptable, while the other half believes that it is not. Half of this world is right, and half wrong about the moral status of water boarding.¹¹ Considering the whole class of moral beliefs about water boarding in the population, half of the beliefs in that class are accurate while the other half is not. If a class of beliefs only tracks the truth 50% of the time, it seems clear that one should not rely on such beliefs. In order to be justified in relying on a moral belief about water boarding in this world, one would need to confirm that her belief is in the 50% of beliefs about water boarding that are accurate, rather than the other half. Since it is difficult to imagine what sort of confirming evidence one could find about her moral beliefs on water boarding, you can see why Mackie favored the skeptical conclusion; confirmation of one's moral beliefs is needed, which means they cannot be justified non-inferentially, but no confirmation seems imminent, which suggests that they cannot be

¹¹ Again, this is on the assumption of non-relativistic moral realism.

justified inferentially either (1977). The point against intuitionism, though, is just that some confirmation of one's moral intuitions about water boarding in the imagined world is needed, whether it is imminent or not, in order for a believer to be justified in her belief about water boarding. This is because without confirmation there is a 50/50 chance that her belief is inaccurate versus accurate. Thus her moral belief can only be justified with confirmation, and can accordingly only be justified inferentially (since confirmation justifies inferentially, as discussed in Section 2.1).

Thus, for cognitive biases (or moral disagreement) to create the need to confirm moral intuitions in order for them to be justified, the biases (or disagreements) would have to be sufficiently prevalent such that the reliability of moral intuitions, as a class of intuitions, was called into question. And, according to the argument from cognitive biases (and from disagreement), if the biases (or disagreements) are sufficiently prevalent to show that moral intuitions are unreliable, then moral intuitions would require confirmation to be justified and could only be justified inferentially; that is, moral intuitionism would be false, moral intuitions cannot be justified non-inferentially. For Mackie (1977), Sinnott-Armstrong (2006a; 2008a; 2008b), Stich, and many others, this is exactly what they see in the case of moral disagreement; disagreement is sufficiently prevalent to show that moral intuitions are unreliable, and thus require confirmation to be justified and cannot be non-inferentially justified. Machery, Kelley and Stich note that "for almost any moral issue, it is possible to find people who hold opposing moral views" (2005). Sinnott-Armstrong says simply that "the range of disagreements among strongly-held non-inferable moral beliefs...shows that many moral believers are unreliable" (2002). But what about cognitive biases of moral intuitions? Sinnott-Armstrong argues that, indeed, they are sufficiently prevalent to show that moral intuitions are unreliable.

2.3 *The empirical picture so far*

So for Sinnott-Armstrong, if moral intuitions are unreliable due to cognitive biases, then they must be confirmed in order to be justified, they are justified only inferentially, and intuitionism is false. All that appears left for this argument to go through is the truth of premise C4. That is, what remains to be shown is that cognitive biases of moral intuitions are prevalent enough to show that a sufficient amount of moral intuitions are inaccurate, and accordingly that moral intuitions are unreliable; cognitive biases will have to be shown to affect a variety of moral intuitions in a variety of circumstances. To this end, Sinnott-Armstrong cites several studies of framing effects, which I review below in Sections 2.3.1 through 2.3.3 (2008a). Then in Section 2.3.4, I will review evidence of another kind of cognitive bias, the self-other asymmetry, affecting moral intuitions (Nadelhoffer and Feltz in prep.). In 2.3.5, I sum up the status of C4, which looks to be on solid ground. Then I review a potential objection to the argument in Section 2.4. In Section 4 I will offer a different objection to the argument from cognitive biases analogous to the classic objection to the argument from disagreement from the non-moral basis of moral disagreement (which is reviewed in Section 3); just as the argument from disagreement requires that the moral disagreements are not grounded in non-moral disagreements, I will suggest that Sinnott-Armstrong's argument requires that the cognitive biases of moral intuitions must not be grounded in biases of non-moral judgments—and that it appears the biases are so grounded in biases of non-moral judgments. But first, I will discuss the evidence of cognitive biases of moral intuitions.

2.3.1 Tversky and Kahneman

In a seminal paper, Tversky and Kahneman were the first to investigate framing effects. They presented subjects in group 1 (N=152) with the following scenario:

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. (Tversky and Kahneman 1981)

Which of the two programs would you favor?

Subjects in group 2 (N=155) read the same “cover story”, but the programs had different consequences:

If Program C is adopted 400 people will die.

If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Which of the two programs would you favor?

Subjects then had to choose a Program between A and B (for subjects in group 1) or between C and D (for subjects in group 2). What is important to note is that Programs A and C are objectively identical, resulting in the exact same numbers of lives saved (200) and lives lost (400). The same goes for Programs B and D; both have a 1/3 probability that 600 lives will be saved, and a 2/3 probability that 600 lives will be lost. Thus, subjects in groups 1 and 2 were choosing between the exact same programs, just framed in different ways: the first group’s programs are framed in terms of the lives that will be saved, whereas the second group’s programs are framed in terms of the lives that will be lost.

Nonetheless, 72% of subjects in group 1 chose A (and 28% chose B) while only 22% of subjects in group 2 chose C (and 78% chose D). Thus, the wording of the scenarios has a huge influence on people’s moral judgments. When the options are framed in terms of the lives saved people are much more likely to be *risk averse*; most people (72%) chose to save 200 lives and lose 400 lives, rather than choosing to take a chance at saving all 600 lives at the risk of all 600

lives being lost. In contrast, when the options are framed in terms of the lives lost people are much more likely to be *risk taking*; most people (78%) chose to take a chance at saving all 600 lives at the risk of all 600 lives being lost, rather than choosing to save 200 lives and lose 400 lives. While subjects' judgments were about choosing which program should be favored, it seems reasonable to follow Sinnott-Armstrong in interpreting their judgments as reflecting their moral intuitions about the scenarios (2008a).

But as Sinnott-Armstrong notes, there should be little doubt that “descriptions cannot affect what is really morally right or wrong in this situation” (2008a); that is, wording (in terms of lives saved versus lives lost) is a morally irrelevant property of the stimuli. People's moral beliefs should not depend on such factors.¹² Accordingly, the Tversky and Kahneman study shows a clear example of a cognitive bias, specifically a framing effect, on moral beliefs. It's clear, Sinnott-Armstrong says, that “such moral beliefs are unreliable” (2008a). But this is only one cognitive bias affecting moral intuitions in one set of circumstances; is there more evidence of cognitive biases of moral intuitions?

2.3.2 Petrinovich and O'Neill

In a study by Petrinovich and O'Neill, the authors again found that “there were framing effects [on people's moral intuitions] produced by differences in wording” (1996). They presented subjects with the standard trolley dilemma:

A trolley is hurtling down the tracks. There are five innocent people on the track ahead of the trolley, and they will be killed if the trolley continues going straight ahead. There is a spur of track leading off to the side. There is one innocent person on that spur of track. The brakes of the trolley have failed and there is

¹² I, along with Sinnott-Armstrong, Tversky and Kahneman, and many others, take it that this is an apparently irrelevant property of the situation; the burden would be on someone who disagrees to show why it is in fact relevant.

a switch that can be activated to cause the trolley to go to the side track.

You are an innocent bystander (that is, not an employee of the railroad, etc.). You can throw the switch, saving the five innocent people, which will result in the death of the one innocent person on the side track. What do you do? (Petrinovich and O'Neill 1996)

Subjects were then asked to rate on an odd number scale (no zero) from -5 (strongly disagree) to +5 (strongly agree) how much they disagreed or agreed with each of the two horns of the dilemma—throwing the switch and not throwing the switch. For half of the subjects, the dilemma was presented in a Kill wording: “throw the switch which will result in the death of the one innocent person on the side track” or “do nothing, which will result in the death of the five innocent people.” For the other half of the subjects, the dilemma was presented in a Save wording: “throw the switch, which will result in the five innocent people on the main track being saved” or “do nothing, which will result in the one innocent person being saved” (Petrinovich and O'Neill 1996). Notice that the Kill and Save wording are objectively identical, and differ only in terms of the way the outcomes are framed.

As you've probably guessed, subjects who saw the Save wording were likely to agree more strongly with throwing the switch (mean agreement of +0.65 on the -5 to +5 scale) than subjects who saw the Kill wording (mean agreement of -0.78). People had different moral intuitions, depending on how the dilemma was framed—using Kill wording or Save wording. Since, as before, wording is a rather obvious example of a morally irrelevant property of a stimuli (a property that should not affect one's moral beliefs about the stimuli), Sinnott-Armstrong concludes that the Petrinovich and O'Neill study presents another circumstance in which moral intuitions are unreliable due to the effects of a cognitive bias.

2.3.3 Haidt and Baron

Sections 3.2.1 and 3.2.2 highlighted examples of word framing effects on moral intuitions. Haidt and Baron (to their surprise) found that moral intuitions can also be affected by order framing effects (1996). In one experiment, they presented all subjects with two versions of a scenario in which Nick is trying to sell his car, 1984 Mazda MPG, to Kathy, and subjects were asked to rate Nick's "goodness" in each version (from extremely bad/reprehensible to extremely good). In the Act version, Nick actively lies to Kathy about the year that many Mazda MPGs had a manufacturer's defect so that she does not know that it happened in 1984, the year Nick's car was made. In the Omission version, Kathy mentions to Nick that she thinks she heard somewhere that Mazda MPGs had a defect in 1983, and Nick fails to correct her (he withholds, or omits, the information about the correct year of the defect). Half of the subjects read and responded to the Omission version first and the Act version second, and the other half read and responded in the opposite order—Act first and Omission second.

Haidt and Baron found a large order effect, as "eighty percent of subjects in the omission-first condition rated the act worse than the omission, while only 50 percent of subjects in the act-first condition made such a distinction" (1996). So, when subjects read Omission before Act, they are very likely to say that Nick is worse in Act (when he actively lies) than Omission (when he merely withholds information). In contrast, when subjects read Act before Omission, they are much less likely to say that Nick is worse in Act than Omission. Haidt and Baron suggest that much of this effect is due to subjects who read Act first being unwilling to "'pull back' and partially excuse the omission" after rating Nick on the 'extremely bad' end of the scale in Act (1996). Thus, subjects who read Act first rate Nick in Act and Omission as quite a bad person. But subjects who read Omission first have no trouble excusing Nick's omission, rating him as

not so bad, since they haven't antecedently judged him to be bad in Act—and then when they read Act, they rate Nick as bad because it is much easier to 'add-on' blame than it is to 'pull it back'.

There is probably an evolutionary story that could be told to explain why adding-on blame is easier than pulling it back; perhaps about how risking over-blaming others (by easily adding blame, and not easily taking it back) is a safe, advantageous practice. But whatever the best explanation of the data turns out to be, the fact is that moral intuitions about the moral status of Nick's character in Act and Omission are affected by the order in which people read the two versions. Thus, on the assumption that the order in which moral intuitions are formed is morally irrelevant—i.e. should not affect the intuitions—it seems that this is another circumstance in which moral intuitions are unreliable due to the effects of a cognitive bias.

2.3.4 Nadelhoffer and Feltz

So far we have looked at framing effects on moral intuitions, which are a specific type of cognitive bias whereby a cognitive process is biased by the presentation of the dilemma (e.g. wording or order effects). Nadelhoffer and Feltz have recently found that moral intuitions can also be affected by another well-known cognitive bias, the self-other asymmetry (in prep.; Malle 2006). This is a bias which causes people to make different judgments about, or perceive differently, stimuli presented in the first versus third-person point of view. An example is the famous attribution bias in social psychology: people are likely to attribute other people's behavior as deriving from their character, but likely to attribute the exact same behavior as deriving from external circumstances when it is their own behavior (Malle 2006). The authors in the present study presented subjects with the traditional trolley dilemma, almost identical to the one used by Petrinovich and O'Neill (1996). Half of the subjects were placed in the Self

condition, and saw the original dilemma, where the subject herself is depicted as performing the action in the dilemma (of throwing or not throwing the switch). The dilemmatic question then asks if it is morally permissible for “you” (the subject) to throw the switch. The other half of the subjects were placed in the Other condition and saw a modified version of the trolley dilemma in which it is another person (John) who is depicted as performing the action in the dilemma. The dilemmatic question then asks if it is morally permissible for “John” to throw the switch.

Rather unexpectedly, 65% of subjects in the Self condition judged throwing the switch permissible, whereas 90% of subjects in the Other condition judged throwing the switch permissible (Nadelhoffer and Feltz in prep.). That is, subjects are significantly more likely to judge throwing the switch to be morally permissible when it is another person who throws the switch than when it is the subjects themselves throwing the switch. According to Nadelhoffer and Feltz, “it is clear that it can’t be less morally permissible for *me* to hit the switch than it is for *someone else* to hit the switch—all things being equal” (in prep.). Presumably Sinnott-Armstrong would agree; all else being equal, the specific person who performs an action is a morally irrelevant property of the scenario. Thus, we have found another instance of moral intuitions being affected by a cognitive bias.¹³

2.3.5 The status of premise C4

Now that we have reviewed all of the evidence offered by Sinnott-Armstrong, and some additional support from Nadelhoffer and Feltz, we are prepared to evaluate the status of premise C4, which states that moral intuitions are sufficiently affected by cognitive biases to show that moral intuitions are, in general, unreliable. On Sinnott-Armstrong’s account, “together these

¹³ It may be worth noting that Nadelhoffer and Feltz explain this bias by arguing that the Self condition is more emotionally salient than the Other condition because the subject is actually described as being present and as performing the action (in prep.). They argue that this higher salience leads subjects to be more likely to reject the flipping the switch in Self than they are in Other.

studies show that moral intuitions are subject to framing effects in many circumstances” (2008a). And we could add that it seems that moral intuitions are subject not just to framing effects, but at least one other cognitive bias as well (the self-other asymmetry). Thus, Sinnott-Armstrong concludes that premise C4 is on solid ground.

There are a variety of responses that intuitionists can offer against premise C4. First of all, it seems clear that we have not seen evidence of anything close to “many circumstances” in which moral intuitions are subject to cognitive biases. A handful of studies about hypothetical and often unrealistic dilemmas (e.g. trolley dilemmas) shows very little about whether our real world moral intuitions are affected by cognitive biases at all, let alone whether they are affected by biases in many circumstances. In a brief response to Sinnott-Armstrong’s argument from cognitive biases, William Tolhurst noted that “the studies [cited by Sinnott-Armstrong] were designed to elicit framing effects in the subjects”; whereas “the situations in which we generally form our spontaneous moral beliefs are not” (2008). Moreover, our ordinary intuitions are not formed in response to verbally formulated moral questions—they are in response to perceptually presented stimuli (like seeing a person steal a candy bar or tell a lie) (Tolhurst 2008). The point is that the evidence reviewed above shows very little about our ordinary moral intuitions, whether or not they are affected by cognitive biases, and if they are so affected, to what extent they are affected. As Shafer-Landau suggests, it is fair to say “that, as yet, we simply do not have a sufficient number of relevant experiments to give us much indication of how many of our moral beliefs are subject to framing effects” (2008). Furthermore, consider the fact that while Petrinovich and O’Neill did report framing effects on people’s moral intuitions, the authors also reported that “the effects were not always large nor did they always appear” (1996). They also note that “the research evidence for reflection and framing effects is mixed rather than uniformly

positive, being dependent on differences in problem characteristics that are only partly understood” (1996). Considering the countless moral intuitions we form in our ordinary, day-to-day lives—like when we read the newspaper, watch television, or gossip with friends—“mixed” evidence for framing effects in some laboratory experiments does not seem like evidence of our intuitions being biased in “many circumstances”.

Sinnott-Armstrong attempts to rebut this objection about insufficient evidence. He suggests that for moral intuitions to be justified they must have a very high degree of reliability, probably close to what scientists require for the justification of scientific beliefs: “if moral beliefs are to be justified in anything like the way scientific beliefs are justified, then it has to be reasonable to assign them a probability [of truth] of at least .95” (2008b). While he offers this general degree of reliability, he tries to avoid committing to a specific degree, or to a specific prevalence of cognitive biases of moral intuitions that would be required to show unreliability of the whole class of moral intuitions. He says simply that “if someone denies that [the empirical results discussed above] are large enough, then my only recourse is to recite the details of the studies, to invoke the high costs of mistaken moral intuitions, and to remind critics that only a minimal kind of [moral] confirmation is needed” (2008b, 101).¹⁴ His point is that a very high percentage of moral intuitions in the class of moral intuitions must be true in order for the class to be reliable, because our moral intuitions are too important to risk them being inaccurate. Thus, he suggests, the results above demonstrate that cognitive biases of moral intuitions are sufficiently prevalent to show that moral intuitions are unreliable.

¹⁴ By minimal confirmation, Sinnott-Armstrong simply means that the confirmation needed to justify the moral intuition does not actually need to be used to infer the intuition; for Sinnott-Armstrong, a believer’s confirmed moral intuition is justified so long as the believer is able to infer the intuition, even if she never actually performs the inference—that is, so long as she can offer the confirmation if asked, and demonstrate an understanding of the inferential link between the confirmation and the intuition.

But Sinnott-Armstrong's response here assumes that the moral intuitions probed in the studies of cognitive biases are ordinary moral intuitions. It is our ordinary moral intuitions that are at issue in the argument from cognitive biases, and it is not clear that the empirical studies cited above show *any* evidence that our *ordinary* moral intuitions are subject to cognitive biases at all, let alone to a sufficient degree to raise questions about their reliability. Clearly moral intuitions can be biased, but professional scientist's scientific intuitions could probably be biased as well if they are presented with appropriate stimuli. I presume that this would not lead us to judge their ordinary scientific intuitions (i.e. the intuitions that they have in the lab every day) to be unreliable, since the stimuli used to generate the bias are likely to be far from the *ordinary* one's that the scientists would encounter in their lab. If ordinary stimuli do generate significant biases of people's intuitions, then we can draw the conclusion that ordinary intuitions are subject to cognitive biases, and thus, depending on the prevalence of the biases, we can assess the reliability of those ordinary intuitions. The evidence offered by Sinnott-Armstrong does not use ordinary stimuli to generate biases. And therefore, the conclusions we can draw about ordinary moral intuitions are limited, at best.

Moreover, the experiments cited have subjects making a variety of judgments, like choosing which program to implement or agreeing with a course of action. It takes some work to show that these judgments are reflective of people's moral intuitions. Subjects' responses likely reflect a variety of considerations, not just moral ones. But even if we grant the subjects' responses are reflective of a moral judgment, as I mentioned briefly in Section 2.3.1, it is not clear that these judgments reflect *moral intuitions* (as we've defined them here) or moral beliefs. In order to show that the effects probed in the cited studies are biases of moral *intuitions*, there should also be evidence that the moral beliefs subjects form are moral beliefs that have not been

formed inferentially and that subjects are unwilling to give them up easily. Again, the empirical picture simply isn't complete enough to make a determination on this issue. Despite these concerns, I am inclined to grant Sinnott-Armstrong the benefit of the doubt that sufficient evidence of cognitive biases of ordinary moral intuitions will emerge. Thinking about my own everyday moral reasoning, it seems certain that my moral intuitions are affected by morally irrelevant factors from my own mood to perceptual properties of the stimuli.

There are a variety of examples from the empirical literature other than the ones cited by Sinnott-Armstrong that could be interpreted as relevant here. For example, consider the startling evidence of implicit prejudice or implicit gender biases (see Brauer, Wasel & Niedenthal 2000 for a review of implicit prejudice; see Banjai and Greenwald 1995 for implicit gender bias). People who explicitly reject prejudice are likely to show evidence of implicit racial biases. For example, people who explicitly reject prejudice are likely to have different psychophysiological responses (e.g. heart rate and skin conductance) to stimuli depicting black people versus those depicting white people. While this evidence of implicit prejudice does not show a cognitive bias of moral intuitions, it seems reasonable to infer that implicit prejudice does bias moral intuitions; that is, it seems reasonable to think that implicit prejudice leads the same person to form inconsistent moral intuitions about people from different races performing objectively identical actions. The race of the person performing an action is quite clearly a morally irrelevant aspect of the situation being judged, but given the evidence of implicit prejudice, this is exactly what we should expect. Imagine Jane, who is walking down a city street late at night and notices a man approaching her from behind. Upon looking back, it seems that implicit prejudice could result Jane being likely to form different moral intuitions about the man depending on his race. We like to think that we are unaffected by such implicit biases, but the evidence of implicit prejudice is

extensive, and such evidence leads me to accept the fact that our moral intuitions are likely affected by many cognitive biases in many circumstances. Thus, I think it is fair to assess the argument from cognitive biases on the assumption that premise C4 will be revealed to be true. And if it is true, then the intuitionist will need a more nuanced response to Sinnott-Armstrong.

2.4 Rejecting the internalist bias

A first attempt to respond to Sinnott-Armstrong might involve rejecting the internalist bias in his argument (Shafer-Landau 2008). The argument from cognitive biases assumes that under conditions in which a believer has reason to believe that her moral intuitions are unreliable, she must confirm those intuitions to be justified. The intuitions might indeed have been formed reliably, and reliably track the moral truth; but if one believes that this is not the case, then, for Sinnott-Armstrong, an internalist clause must be applied and she must confirm the intuition to be justified. The uncompromising externalist can simply reject this internalist clause; even when a person has reason to doubt the reliability of her intuitions, if they are, in fact, formed reliably, then they are justified—without any need for confirmation. This seems to me to be a perfectly sound way to reject the argument from cognitive biases. But it requires that one adopt an uncompromising externalist moral epistemology. For those who are externalists of a more compromising sort, or who are not externalists at all, rejecting the internalist bias will not work. I will attempt to offer a response to the argument from cognitive biases that does not require one to adopt any particular moral epistemology. The argument from cognitive biases fails, I believe, on a much more fundamental level than its epistemological assumptions—it fails even granting Sinnott-Armstrong’s internalist clause.

3 The argument from disagreement

The argument from disagreement is one of the most common arguments against intuitionism (Mackie 1977; Loeb 1998; Sinnott-Armstrong 2002, 2006a, 2006b), and in this section I will suggest that Sinnott-Armstrong's argument from cognitive biases is simply a fresh take on this old argument (2008a); the argument from cognitive biases attacks intuitionism based on inconsistent intuitions within a single subject across different circumstances, while the argument from disagreement attacks intuitionism based on inconsistent intuitions between different subjects. With this in mind, I will spend some time in this section reviewing the argument from disagreement and the intuitionist's response from the non-moral basis of moral disagreement, which I will adapt as a response to the argument from cognitive biases in Section 4.

Recall that the argument from cognitive biases cites the prevalence of cognitive biases, or *within*-subject moral disagreements, to argue that moral intuitions are unreliable, must be confirmed to be justified, and can only be inferentially justified—and thus that moral intuitionism is false. Similarly, the argument from disagreement cites the prevalence of moral disagreements (or inconsistencies) *between*-subjects to argue that moral intuitions are unreliable, must be confirmed to be justified, and can only be inferentially justified—and thus that intuitionism is false (2008a). The only difference between the argument from disagreement and the argument from cognitive biases is that the unreliability of moral intuitions is due to moral disagreement, rather than cognitive biases. To make clear the parallel with the argument from cognitive biases, the argument from disagreement looks like this:

D1) If moral disagreement is sufficiently prevalent to show that moral intuitions are, in general, unreliable, then moral intuitions must be confirmed in order to be justified (because without confirmation, a believer is not justified in believing her

- intuitions were formed reliably rather than unreliably, and thus she is not justified in believing those intuitions).
- D2) If moral intuitions must be confirmed in order to be justified, then they can only be justified inferentially.
- D3) If moral intuitions can only be justified inferentially, then moral intuitionism is false.
- D4) Moral disagreement is sufficiently prevalent to show that moral intuitions are, in general, unreliable.¹⁵
- D5) Moral intuitionism is false.¹⁶

Each of these premises was discussed in detail above, in relation to the argument from cognitive biases. Again, the only difference is that cognitive biases have now been replaced by moral disagreement; within-subject disagreement has been replaced by between-subject disagreement. The point of the argument remains the same. Moral intuitions are unreliable because of the prevalence of some sort of disagreement, in this case between-subject moral disagreement. Thus, intuitions must be confirmed to be justified—and confirmation justifies only inferentially. So moral intuitions must be justified inferentially if they are to be justified at all, and moral intuitionism is false; moral intuitions cannot be justified non-inferentially. The prevalence of moral disagreement is supposed to show that moral intuitionism is false in the same way that the prevalence of cognitive biases of moral intuitions was supposed to show that intuitionism is false.

¹⁵ One might argue that moral disagreement, by itself, does not need to be sufficiently prevalent to make moral intuitions unreliable, but rather disagreement might be just one of several sources of unreliability that, together, make moral intuitions reliable (Sinnott-Armstrong 2006a). Thus, it may be that for moral disagreement to present a problem for moral intuitionism it need not be so prevalent as to bring moral intuitions' reliability into question; it need only be prevalent enough to do so in combination with other sources of unreliability. I actually think that other sources of unreliability can be reduced to forms of disagreement, and thus that this way of arguing would reduce to the argument from disagreement.

¹⁶ This is my formulation of the argument, which has been reconstructed from Mackie 1977, Sinnott-Armstrong 2006, and Machery et al. 2005. I have formulated it in this way to make clear the parallel with the argument from cognitive biases.

3.1 *Non-moral basis of moral disagreement*

The intuitionist's classic response to the argument from disagreement is to argue that most moral disputes are not genuine (based solely on disagreement about the moral facts), they are derived—they are grounded in disagreement about non-moral facts. David Brink says that most “moral disputes are *in principle* resolvable” because they are grounded in “resolvable disagreements over the non-moral facts” (Brink 1984). That is, moral disputes can be resolved without moral confirmation; they can be resolved by resolving the non-moral disagreements that the moral disputes are based on. In this vein, Richard Boyd suggests, perhaps overly optimistically, that “careful examination will reveal...that agreement on nonmoral issues would eliminate *almost all* disagreement about the sorts of moral issues which arise in ordinary moral practice” (1988, p. 123). People disagree about non-moral facts like reincarnation, the ability of animals to feel pain, and the long-term consequences of actions—and disagreements in moral intuitions often are based on disagreements in such non-moral issues. And this is important because moral disagreement grounded in non-moral disagreement would not threaten moral intuitionism, since such disagreement could be resolved without making the moral intuitions inferable. Once the non-moral disagreement is resolved, the reliability of moral intuitions would not be called into question; at least not on the basis of moral disagreement—since there would be no moral disagreement once its grounding non-moral disagreement is cleared up. Thus, the moral intuitions could (potentially) be justified non-inferentially (so long as there was no other reason to question their reliability); the argument from disagreement would not have force against intuitionism if moral disagreement is derived. There may be other arguments that do have force, but if moral disagreements are derived, and not genuine, then the argument from disagreement would not threaten intuitionism.

More specifically, if non-moral disagreements (disagreements about non-moral facts) are the basis for most moral disagreements, then premise D1 would be false; the unreliability of moral intuitions due to moral disagreement would not show that moral intuitions must be confirmed for them to be justified. The unreliability of moral intuitions due to moral disagreement *which is grounded in non-moral disagreement* shows that *either* people's moral intuitions *or* their relevant non-moral beliefs must be confirmed to allow their moral intuitions to be justified.¹⁷ This is because moral disputes based on non-moral disputes can be resolved by confirming the moral intuitions—which would justify those moral intuitions inferentially—*or* the non-moral ones—which would not justify the moral intuitions, but would remove the need to inferentially justify them that was created by the moral disagreement (since the moral disagreement would dissolve once its grounding non-moral disagreement is resolved).

As an example, recall the world in which half of the population believes that water boarding is morally wrong and half believes it is right. Now imagine that this moral disagreement is grounded in a non-moral disagreement, like a disagreement about the average effectiveness of water boarding in getting accurate information out of terrorism suspects. Suppose that one half of people believe that water boarding is wrong only because it is *ineffective* in getting information, while the other half believes it is right only because it is *effective* in getting information. Now imagine that it is determined that water boarding is, in fact, completely ineffective in getting accurate information out of enemies. In this case, everyone in the world will agree that water boarding is wrong, since half of the population thought it was right only because they (wrongly) believed it to be effective in getting information. The key is that since the moral disagreement is grounded in non-moral disagreement, the moral disagreement and unreliability of moral intuitions could be resolved by resolving the non-moral

¹⁷ The relevant non-moral beliefs are the ones responsible for the moral disagreement.

disagreement and unreliability of non-moral beliefs; only the non-moral beliefs (about the effectiveness of torture) were confirmed, yet the moral disagreement is resolved. After the non-moral disagreement was resolved, there remained no moral disagreement that could bring the reliability of people's moral beliefs into question and could demand confirmation of them. Thus, while the moral disagreement is sufficiently widespread to bring the reliability of people's moral intuitions about water boarding into doubt, their *moral intuitions* did not need to be confirmed for them to be justified; confirmation of their *non-moral beliefs* cleared up the moral disagreement and, accordingly, cleared up questions of the reliability of their moral intuitions. The non-moral confirmation does not itself justify the moral beliefs, but it removes the need for inferential justification that was created by the moral disagreement. (There may very well be other reasons inferential justification is required, but moral disagreement would not be one of them—since no moral disagreement would exist after the non-moral confirmation is provided.) Moral confirmation could have been provided to clear up the moral disagreement—if they somehow discovered that, in fact, water boarding is wrong—but none was *required*. Thus, premise D1 as stated is false. While, as the antecedent of premise D1 suggests, moral disagreement in our imagined world was sufficiently prevalent to show that moral intuitions (about water boarding) are, in general, unreliable, contrary to the consequent of premise D1, this entailed that either the moral *or non-moral* beliefs needed confirmation in order for the moral intuitions to be justified.

To be clear, if most moral disagreement turns out to be derived, or based on non-moral disagreement, this would not show that only non-moral beliefs are unreliable instead of moral intuitions.¹⁸ That is, it would not show that premise D4 and the antecedent of premise D1 are

¹⁸ By “most moral disagreements”, I mean that enough moral disagreements are grounded in non-moral disagreements so that the reliability of moral intuitions, in general, cannot be called into question by the amount of

false. Derived moral disagreement still shows that moral intuitions are unreliable, because the same amount of the intuitions are accurate and the same amount remain inaccurate—there is still moral disagreement, after all. All it shows is that moral intuitions depend on unreliable non-moral beliefs for their formation. In this way, derived moral disagreement would suggest that premise D1 is false; sufficient disagreement in moral intuitions such that the intuitions are unreliable would not entail that the moral intuitions must be confirmed to be justified, but only that either the moral intuitions or the non-moral beliefs must be confirmed.

It may seem that premise D1 could simply be amended to premise D1', which states that if moral disagreement is sufficiently prevalent to show that moral intuitions are, in general, unreliable, then moral intuitions *or the relevant non-moral beliefs* require confirmation in order for moral intuitions to be justified. By adopting this move, however, we lose the connection between premises D1' and D2; the consequent of premise D1' states that moral intuitions or non-moral beliefs require confirmation for moral intuitions to be justified, while the antecedent of premise D2 refers to only moral intuitions requiring confirmation for them to be justified. And premise D2 is false when its connection with premise D1' is restored by amending it to D2', which states that if moral intuitions *or non-moral beliefs* require confirmation for moral intuitions to be justified, then moral intuitions are justified only inferentially. This is because *non-moral* confirmation (i.e. confirmation of one's non-moral beliefs) does not justify moral intuitions inferentially—it does not allow one to infer her moral intuitions from the non-moral confirming evidence of her non-moral beliefs.¹⁹ As it relates to the example above, people in our

moral disagreement that is not grounded in non-moral disagreement—that is, enough moral disagreements are grounded in non-moral disagreements so that the solely moral disagreement that remains is not sufficiently prevalent to make premise 4 true.

¹⁹ I may sometimes say that the non-moral confirmation 'justifies the moral intuitions'—but by this I mean that the confirmation removes any reason to question the intuitions reliability, and thus (barring some other reason to reject the intuitions), moral intuitions are justified.

imagined world cannot infer from the non-moral confirmation of the ineffectiveness of torture to the normative claim that torture is wrong, on pain of inferring an ought from an is.²⁰ Still, the non-moral confirmation cleared up the source of the unreliability of the moral disagreement, and thus allows the moral intuitions to be justified. While the non-moral confirmation does not *make* the intuitions justified, it removes the requirement of moral confirmation for the intuitions' justification—and it does so without making them *inferable* from the confirming evidence. There may be other reasons that moral intuitions are unreliable and require confirmation to be justified, but moral disagreement would no longer be one of them if that disagreement is derived.

To make D2' true, it would have to read that if moral intuitions or non-moral beliefs must be confirmed for moral intuitions to be justified, then, moral intuitions could be justified either inferentially *or non-inferentially*. However, such a conclusion (stated in the consequent), clearly poses no threat to intuitionism. Moral intuitionism holds that some moral intuitions can be justified non-inferentially, and the conclusion above allows for such non-inferential justification.

In this way, the availability of non-moral confirmation as a way of resolving moral disagreements would allow the intuitionist to say that moral disagreement does not threaten the non-inferential justification of moral intuitions—under conditions in which the non-moral disagreement has been resolved, or under 'ideal conditions'. This is the thesis of what I will call a moderate moral intuitionism, which holds that moral intuitions can be non-inferentially justified under ideal conditions. So, *if* most moral disagreements are grounded in non-moral disagreements, then the argument from disagreement against a moderate moral intuitionism fails because premise D1 is false, and making premise D1 true by amending it to D1' yields a conclusion to the argument from disagreement that has no force against intuitionism. Moral disagreement may be widespread, but if most of it is grounded in non-moral disagreement then it

²⁰ This is the problem that Hume made famous in *A Treatise of Human Nature* (Book III, Part I, Section I).

does not show that most moral intuitions must be confirmed to be justified, and that intuitionism is false. Rather, it would show that either moral intuitions or the relevant non-moral intuitions require confirmation, and thus leave the door open for moderate intuitionism (a thesis which I think would be perfectly acceptable to most intuitionists, and quite objectionable to their critics, like Sinnott-Armstrong).

So the unreliability of moral intuitions due to moral disagreement creates the need for confirmation of one's moral intuitions for them to be justified *only if* the moral disagreement is not grounded in non-moral disagreement. That is, premise D1 is true only when the cited moral disagreement is genuine—i.e. is not grounded in non-moral disagreement. Thus, for the argument from disagreement to go through, it requires an additional premise:

The relevant moral disagreement is genuine—that is, not grounded in non-moral disagreement (disagreement about non-moral facts).

So we can now formulate a revised argument from disagreement:

- D1*) If moral disagreement is sufficiently prevalent to show that moral intuitions are, in general, unreliable, *and moral disagreement is not grounded in non-moral disagreement*, then moral intuitions require confirmation in order to be justified (because without such confirmation, a believer is not justified in believing her intuitions were formed reliably rather than unreliably, and thus she is not justified in believing those intuitions).
- D2*) If moral intuitions require confirmation in order to be justified, then they can only be justified inferentially.
- D3*) If moral intuitions can only be justified inferentially, then moral intuitionism is false.
- D4a*) Moral disagreement is sufficiently prevalent to show that moral intuitions are, in general, unreliable.²¹

²¹ One might argue that moral disagreement, by itself, does not need to be sufficiently prevalent to make moral intuitions unreliable, but rather disagreement might be just one of several sources of unreliability that, together, make moral intuitions reliable (Sinnott-Armstrong 2006). Thus, it may be that for moral disagreement to present a problem for moral intuitionism it need not be so prevalent as to bring moral intuitions' reliability into question; it need only be prevalent enough to do so in combination with other sources of unreliability. I actually think that other sources of unreliability can be reduced to forms of disagreement, and thus that this way of arguing would reduce to the argument from disagreement.

D4b*) The moral disagreement cited in 4a'' is *genuine*—that is, not grounded in non-moral disagreement (disagreement about non-moral facts).

D5*) Moral intuitionism is false.

Note that D4a* is the same as D4, and that D1* is the same as D1, with the addition of D4b* to its antecedent. By genuine moral disagreement, I just mean moral disagreement that is not grounded in non-moral disagreement (which I will call *derived* moral disagreement). After the argument is properly amended in this way, we see how the intuitionist's objection from the non-moral basis of moral disagreement works—by rejecting premise D4b*. For example, David Brink argues that “a great many moral disagreements depend upon disagreements over the non-moral facts”—that is, moral intuitions would converge under ideal conditions, in which the non-moral facts are agreed upon (1984). In contrast, adherents to the argument from disagreement suggest that non-moral disagreement is not sufficient to explain the overwhelming prevalence of moral disagreement (Machery, Kelly, and Stich 2005; Doris and Stich 2005; Sinnott-Armstrong 2006a and 2008a; Brandt 1959). Doris and Stich argue for “pessimistic conclusions regarding the possibility of convergence in moral judgment”; few, if any moral intuitions would converge in ideal conditions (when non-moral facts are agreed upon), and thus disagreement in moral intuitions is not typically grounded in non-moral disagreement. But Loeb conservatively advises more serious empirical investigation into whether or not and how many moral disagreements are grounded non-moral disagreements before assessing the argument from disagreement and the objection from the non-moral basis of moral disagreement (Loeb 1998). These perspectives will be important to keep in mind in what follows, as I adapt the intuitionist's objection from the non-moral basis of disagreement to object to the argument from cognitive biases on the grounds of the non-moral basis of cognitive biases.

4 Non-moral basis of cognitive biases of moral intuitions

In what follows, I will attempt to adapt the response from the non-moral basis of moral disagreement that the intuitionist offered against the argument from disagreement. This response held that most moral disagreements are derived from non-moral disagreements, and thus can be resolved by either moral or non-moral confirmation. Since moral intuitions were deemed unreliable in premise D4a* because of the prevalence of moral disagreement, once the moral disagreement is cleared up by clearing up the non-moral disagreement, there is no reason to think that moral intuitions are unreliable. In this way, non-moral confirmation (i.e. confirmation of non-moral judgments) can allow moral intuitions to be justified by removing the moral disagreement (which was grounded in the non-moral disagreement). It does not make them justified, but it rebuts the argument from disagreement because it prevents moral intuitions' reliability from being called into question *on the grounds of moral disagreement*. And the key is that non-moral confirmation does this without making subjects able infer their moral intuitions from the confirming evidence that cleared up the moral disagreement, since that confirmation was non-moral confirmation of their non-moral judgments, and one cannot infer a moral intuition from non-moral evidence. For this reason, the argument from disagreement needed to be amended to include the premise that the moral disagreement is genuine—i.e. not grounded in non-moral disagreement. The intuitionist then rejects the amended argument from disagreement by rejecting this additional premise and holding that most moral disagreement is derived from disagreement in non-moral judgments.

In Section 4.1 I will reformulate the argument from cognitive biases to include the insight that the disagreements (in this case within-subject, cognitive biases) must be *genuine*. In 4.2 I adapt the response from the non-moral basis of moral disagreement to respond to the argument

from cognitive biases, arguing for the non-moral basis of cognitive biases. Then in Section 4.3 I will discuss some replies to this objection suggested by Sinnott-Armstrong. In the end, I believe that more and better designed experiments can resolve this dispute, and in Section 5 I will present an initial experiment I have conducted in collaboration with Eddy Nahmias that weighs in favor of the intuitionist's response from the non-moral basis of cognitive biases.

4.1 Revised argument from cognitive biases

Recall that the only difference between the argument from cognitive biases and the argument from disagreement is that the unreliability of moral intuitions is due to cognitive biases, rather than moral disagreement. So we should expect unreliability due to cognitive biases to have the same limitation as unreliability due to moral disagreement; namely, that if cognitive biases of moral intuitions are grounded in biases of non-moral judgments (i.e. if the cognitive biases of moral intuitions are *derived*), then those biases would not threaten intuitionism since they can be resolved by non-moral confirmation. Thus, we can formulate a revised argument from cognitive biases:

- C1*) If moral intuitions are sufficiently affected by cognitive biases to show that moral intuitions are, in general, unreliable, *and those cognitive biases are not grounded in biases of non-moral beliefs*, then moral intuitions must be confirmed in order to be justified (because without confirmation, a believer is not justified in believing her intuitions were formed reliably rather than unreliably, and thus she is not justified in believing those intuitions)
- C2*) If moral intuitions require confirmation in order to be justified, then they can only be justified inferentially.
- C3*) If moral intuitions can only be justified inferentially, then moral intuitionism is false.
- C4a*) Moral intuitions are sufficiently affected by cognitive biases to show that moral intuitions are, in general, unreliable.
- C4b*) The cognitive biases cited in C4a are *genuine*—that is, not grounded in biases of non-moral judgments (biases of people's beliefs or intuitions about non-moral facts).

C5*) Moral intuitionism is false

This revised argument has added premise C4b*, which holds that the cognitive biases cited in the argument must not be grounded in biases of non-moral judgments, and has also added that proposition to the antecedent of C1. The reason is the same as it was for adding premise D4b* to the revised argument from disagreement. Moral inconsistency is what brings the reliability of moral intuitions into question in the first place, and what accordingly drives the arguments against intuitionism from disagreement and cognitive biases. If the inconsistencies in moral intuitions cited in these arguments are grounded in inconsistencies in non-moral beliefs or intuitions, then clearing up the non-moral difference would dissolve the moral inconsistency. And the key is that it would dissolve the moral inconsistency without providing inferential justification of the moral intuitions (since one could not infer her moral intuitions from the non-moral confirmation; see Section 3.1 for more on this).

4.2 Non-moral basis of biases

Recall that the argument from disagreement is based on between-subject differences in moral intuitions, while the argument from cognitive biases is based on within-subject differences in moral intuitions (e.g. the *same* subject would have different intuitions about different presentations of objectively equivalent moral dilemmas or scenarios). The objection from the non-moral basis of moral disagreement rejects premise D4b* in the argument from disagreement by holding that much moral disagreement is derived, or based on non-moral disagreement; that is, many differences in moral intuitions between-subjects are grounded in differences non-moral judgments between-subjects. Analogously, this response can be adapted to respond to the argument from cognitive biases by rejecting premise C4b*, and holding that the cognitive biases of moral intuitions are derived; that is, many differences in a single subject's moral intuitions

(about objectively equivalent dilemmas presented in different ways) are grounded in differences in that subject's non-moral judgments²² (about the objectively equivalent dilemmas presented in different ways). This response from the non-moral basis of cognitive biases suggests that a given subject makes different *non-moral* judgments about a scenario when the scenario is framed in different ways, and it is largely on the basis of this difference in non-moral judgments that the subject forms different moral intuitions about the scenario framed in different ways. Accordingly, I contend that many of the cognitive biases of subjects' moral intuitions are grounded in biases of their non-moral judgments in this way. By "many", I mean that enough cognitive biases of moral intuitions are derived (i.e. grounded in non-moral biases) such that any remaining genuine biases of moral intuitions are not sufficient to call into question the reliability of the class of moral intuitions under ideal conditions, or when non-moral biases are not present.

As an example, consider how non-moral biases might ground the moral biases found in Petrinovich and O'Neill's trolley dilemma. Subjects were more likely to say that flipping the switch is appropriate in the Save wording than in the Kill wording. The response from the non-moral basis of cognitive biases suggests that a significant part of the reason that subjects respond differently to the Kill and Save wordings is because they make different non-moral judgments about the scenario depending on the wording used. When framed in terms of the people that will be killed, a subject's options are:

"Throw the switch which will result in the death of the one innocent person on the side track." (Petrinovich and O'Neill 1996)

One possibility is that when presented with this Kill wording, the emphasis in that wording on the killing of the one person on the side track makes it likely that subjects will form the non-moral judgment that throwing the switch will be very likely to kill the one person, but

²² Recall that these differences in non-moral judgments are not necessarily consciously formed—they may be subconscious and may be easily changed. For example, they might fluctuate when the context changes, when new information becomes available, or even when the subject's attention focuses on new aspects of the scenario.

might *fail* to save the five people on the main track (i.e. will be a futile attempt to save the five people). This is because the Kill wording explicitly highlights the fact that the one will be killed, but does not mention what will happen to the five. And if throwing the switch, killing one person, is not going to save the five people, then it makes perfect sense not to throw the switch. Similarly, subjects may be less likely to make these judgments when presented with the Save wording:

“Throw the switch, which will result in the five innocent people on the main track being saved.” (Petrinovich and O’Neill 1996)

When subjects read this Save wording, they may be more likely to make the non-moral judgment that flipping the switch will, in fact, save the five people on the main track (after all, that is exactly what it says). Such a non-moral judgment might at least partially explain why subjects who read the Save wording were more likely to endorse flipping the switch—because they were more likely to judge that flipping the switch would save the five people. Furthermore, the Save wording may have made it less likely that subjects would believe that flipping the switch would kill the one person (after all, the question asking if it is appropriate to flip the switch says that the five will be saved, but makes no mention of the one person being killed); in contrast, the Kill wording might have the opposite effect, making it more likely that subjects would believe that flipping the switch would kill the one person (after all, the question says that the one will be killed, but makes no mention of the five being saved).

Also, Petrinovich and O’Neil reported that subjects were more likely to disagree with *not* flipping the switch in Kill than in Save. This again may be explained by the fact that not flipping the switch is described as resulting in the five people on the main track dying in Kill, while resulting in the one person on the side track being saved in Save. The non-moral facts are the same in both Kill and Save—if the switch is not flipped, the one person on the side track will live

in Kill, and the five people on the main track will die in Save. But these different frames highlight different non-moral facts of the dilemma, and thus are likely to result in subjects making different non-moral judgments about the dilemma in the two different presentations (Kill versus Save). And the key is that their different non-moral judgments may explain why they formed different moral intuitions. They saw the non-moral facts of the case differently in the two conditions, and moral intuitions are formed based on the non-moral facts *as one sees them*.

Or perhaps, more simply, non-moral judgments may be biased because subjects' attention in Save is subconsciously focused on the fact that flipping the switch would save five people. They may not even be aware they that are paying more attention to certain facts of the case. But nonetheless, their non-moral judgments about the case may be subconsciously influenced by their attention being focused on saving the five people rather than killing the one person—and the differences in their moral intuitions could be explained by this difference in non-moral intuitions. It is true that if asked about whether the trolley would kill the one person, they could probably easily report that it would, because the question would refocus their attention from the fact that flipping the switch would save the five people to a different aspect of the case (*viz.* whether or not flipping the switch would kill the one person). And when attending to this other fact people will probably be able to report it accurately. But it is subjects' non-moral judgments about the non-moral facts of the case *when they form their moral intuition* that are relevant—and the different frames likely modify their non-moral judgments about the facts of the case (subconsciously, and perhaps only for a short time) by focusing their attention on certain facts of the case rather than others.

We should note that subjects are *supposed* to base their responses on exactly what the scenario says. Since it says in the Save scenario (even if not in the description of the options) that

flipping the switch will result in the death of the one person on the side track, then that fact should be what subjects base their judgments on. But what is important in determining the basis of the cognitive biases effects on moral intuitions is not what subjects are supposed to base their responses on, but what they, in fact, do base their responses on. Thus, again, what is relevant is the non-moral judgments that subjects actually have in mind *when they offer their moral judgments*. And it is likely that those judgments are not true to the facts of the scenario, but are affected by the way the scenario is presented.

The main problem with the response to *the argument from disagreement* from the non-moral basis of disagreement is that it suggests that most moral beliefs are virtually universally agreed upon, and that for most moral disagreements people actually agree about the moral principles and what is morally at stake, they just disagree about the facts of the case at hand. This seems like a rather unlikely consequence of the response from the non-moral basis of disagreement; it seems reasonable (to many philosophers, myself included) that many moral disagreements are genuine moral disagreements, rather than mere derived moral disagreements actually based on disagreement about the non-moral facts of the case (Machery et al. 2005; Doris and Stich 2005; Sinnott-Armstrong 2006a and 2008a; Brandt 1959)

The response from the non-moral basis of *cognitive biases* does not seem as susceptible to an analogous problem. It seems very likely that part of ordinary moral psychology is a form of moral supervenience, where moral facts are taken to supervene on non-moral facts. That is, it seems likely that most people hold that if two cases do not differ in their non-moral properties, then they do not differ in their moral properties. Surely most people subscribe to such a thesis.²³ Surely people are surprised to find out about the cognitive biases of moral intuitions discussed above; if the dilemmas are objectively equivalent, obviously they should be judged the same.

²³ This statement is ripe for experimental investigation.

Such supervenience seems as obvious to me as the wrongness of killing children for fun. At the very least, it seems far more likely that people do ascribe to the thesis of moral supervenience than reject it. And if one agrees that people probably ascribe to moral supervenience, then it seems strained to think that within-subject differences in moral intuitions are grounded in anything other than within-subject differences in non-moral judgments; since people subscribe to the thesis of moral supervenience, the same person should have different moral intuitions about the same scenario only if she made different non-moral judgments about that scenario. Admittedly, this is a statement about how people's intuitions *should* behave—but by *should* here, I mean not only that rationality demands intuitions to behave in this way, but also that we should expect that, upon empirical investigation, people's intuitions actually do behave in this way. If most people do ascribe to the thesis of moral supervenience, then we should expect that empirical probing will show that cognitive biases of moral intuitions occur only (or mostly) when biases of non-moral judgments also occur.

Once it is established that many cognitive biases of moral intuitions are grounded in biases of non-moral intuitions, the response from the non-moral basis of cognitive biases follows the same route as the response from the non-moral basis of moral disagreement. Since cognitive biases of moral intuitions are due to biases of non-moral intuitions, they can be resolved by either moral or non-moral confirmation. Thus, premise C1 is false because the unreliability of moral intuitions does not imply that moral confirmation is required for a moral intuition to be justified—only that either moral or non-moral confirmation is required. The non-moral confirmation would not itself justify the moral intuitions, but it would remove the cognitive biases and unreliability of moral intuitions since those biases were based on biases and unreliability of non-moral judgments, and the non-moral judgments are not unreliable once they

are confirmed. In this way, non-moral confirmation of non-moral beliefs allows moral intuitions to be justified, and it does so without making people able to infer their moral intuitions from the non-moral confirmation (again, refer to Section 3.1 for more on why this is so).

4.3 An initial objection and a response

Before moving on to more empirical discussions, consider Sinnott-Armstrong's initial response to the idea that within-subject differences in moral intuitions are grounded in non-moral differences. He says that the "changes in [a subject's] moral belief [described in the research on framing effects] cannot be due to changes in the facts of the case, because consequences, knowledge, intention, and other facts are held constant" (2008a, 67). Sinnott-Armstrong is *almost* anticipating the response from the non-moral basis of cognitive biases; however, it is not non-moral differences in the stimuli that are relevant (since, as he notes, there are not any relevant non-moral differences in the stimuli). Rather, it is differences in subjects' non-moral judgments *about* the stimuli that are relevant.

Sinnott-Armstrong does, however, anticipate a sort of objection from the non-moral basis of cognitive biases of moral intuitions, but it seems that he fails to appreciate its force. He says that in the relevant empirical studies, the "descriptions of the cases were admittedly incomplete, so subjects might have filled in gaps in different ways" (2008a, 67). That is, the biases of moral intuitions might be due to the fact that subjects fill in the unspecified non-moral facts of the case in different ways. But he responds to this point by saying that "moral intuitions would still be unreliable. Wording and context would still lead to conflict moral judgments about a single description of a scenario" (2008a, 69). Since they are unreliable, Sinnott-Armstrong argues, they still demand confirmation to be confirmed.

However, he fails to note that the confirmation the unreliable moral intuitions would demand can be non-moral confirmation of the non-moral judgments—which would remove the biases of moral intuitions without making the intuitions inferable from the confirmation. This is why I have gone out of my way to show the parallels between the argument from cognitive biases and the argument from moral disagreement; because it is widely agreed that if moral disagreement is based on non-moral disagreement, then the argument from moral disagreement would have no force against the intuitionist. According to Sinnott-Armstrong himself, “when disagreements about moral issues arise solely from non-moral disagreements and conceptual confusions that could be removed by further inquiry and reflection, informed and rational people would not disagree” (2002). Such derived moral disagreements do not present a serious threat to intuitionism because all the intuitionist needs to claim is that moral intuitions can be non-inferentially justified *once the non-moral disagreement is cleared up*. And the argument from cognitive biases is just the argument from disagreement redone, with the between-subject disagreement being replaced with within-subject disagreement. Thus, we should expect that it faces a similar restriction; if cognitive biases of moral intuitions are grounded in biases of non-moral intuitions, then the argument from cognitive biases does not threaten (a moderate form of) moral intuitionism. And I have argued that it fails to meet this restriction. Cognitive biases of moral intuitions are likely grounded in biases of non-moral intuitions.

At the very least, it should be clear by now that the response from the non-moral basis of cognitive biases cannot be easily dismissed. What is needed to evaluate the response is an empirical exploration of people’s moral intuitions and how they are related to their non-moral intuitions—are biases of moral intuitions based on biases of non-moral intuitions? In this vein, in Section 5 I will discuss an initial study that Eddy Nahmias and I conducted which suggests that

people's non-moral intuitions about the believability of certain outcomes stipulated in a moral dilemma have a significant influence on their moral intuitions about the dilemma.

5 Empirical support for the intuitionist's reply

The studies cited in Sections 2.3.1 through 2.3.4 fail to determine whether or not the cognitive biases of moral intuitions that they reveal are genuine or derived from biases of non-moral intuitions. To initially examine this point, I conducted a survey study in collaboration with Eddy Nahmias in which we probed people's moral intuitions about a novel battery of moral dilemmas, and also examined their non-moral judgments about the dilemmas. There were two primary questions this study was designed to answer: (i) is there a self-other bias in moral intuitions about *personal* moral dilemmas? And (ii) are differences in people's moral intuitions correlated with differences in their non-moral judgments about a dilemma. While Nadelhoffer and Feltz found a self-other bias using impersonal dilemmas, we found no self-other bias using personal dilemmas, suggesting that the self-other biasing of moral intuitions should be further probed to determine the reliability of these findings and the nature of the self-other bias of moral intuitions before this bias is used to argue against intuitionism. Also, we found that non-moral judgments predict subjects' moral intuitions across a variety of moral dilemmas, which suggests that differences in moral intuitions about moral dilemmas might be grounded in biases of subjects' non-moral judgments—differences in subjects' moral intuitions about different, but objectively equivalent presentations of a moral dilemma might be due to differences in their non-moral judgments about the different presentations of the dilemma. Much more work will need to be done to continue to evaluate this claim.

5.1 Methods

Two different surveys, Survey 1 (N=210) and Survey 2 (N=187), each consisting of 8 different personal moral dilemmas²⁴ (making a total of 16 different dilemmas),²⁵ were presented to Georgia State undergraduates enrolled in psychology and philosophy courses using Question Pro (an online survey program). Here is one of the dilemmas, nicknamed Underwater Hatch:²⁶

You are the director of an underwater research facility. There has been an explosion and water is pouring into the facility, causing all of the doors to begin closing automatically. You and six other researchers are in a room that is being sealed off by a closing door. At the last minute, someone in the hall tries to make it to safety in the room by sliding under the closing door, but he is too late and the door closes on his arm, trapping him outside and making it impossible for him to get into the room. This also prevents the door from closing and allows water to rush into the room.

If nothing is done, the water will flood the room and drown you and everyone else inside, including the person trapped under the door. The only way to save you and the other researchers is for you (the facility director) to cut the trapped person's arm off with an axe, leaving him out in the hallway where he will drown, allowing the door to close and preventing the room from flooding.

After reading the dilemma, subjects were asked the moral question:

M) How appropriate is it for you to chop the trapped person's arm off, leaving him out in the hallway where he will drown in order to save yourself and the six other researchers in the room?

They rated the appropriateness of the action on a 6-point Likert scale with no zero, with -3 being completely inappropriate and +3 being completely appropriate.

²⁴ I follow Green et. al (2001) in defining a personal moral dilemma as a dilemma in which the action in question "(a) could reasonably be expected to lead to serious bodily harm (b) to a particular person or a member or members of a particular group of people (c) where this harm is not the result of deflecting an existing threat onto a different party."

²⁵ See Appendix A for the full text of all of the dilemmas. You will notice that the dilemmas were constructed to vary on a variety of dimensions within the context of a personal moral dilemma. For example, some involve personally killing a family member, killing in various ways (drowning, shooting, etc.) which will not be separately analyzed here. These will be ignored for the purposes of this paper, but will present an interesting area for deeper exploration of the data.

²⁶ None of the surveys that subjects completed included the dilemmas' nicknames or the question labels (e.g. M or L1).

To probe their non-moral judgments about the scenario, subjects were then asked two “likeliness” questions. They rated how likely *they thought* two stipulated outcomes of the scenario were on a 5-point scale, with zero being “not likely at all” and 4 being “definitely will happen”. The judgments subjects express in response to these questions are judgments about probabilities of potential consequences of performing, or not performing, the action described in the scenario, and thus they are judgments about the non-moral facts of the case. The first likeliness question asked what subjects thought about the outcome of *not* performing the proposed action as stipulated in the scenario:

L1) If you do NOT chop the trapped person’s arm off, how likely do *you think* it is that the room will flood and you, the six other researchers in the room, and the trapped person will drown?

The consequent here is what the scenario says *will* happen if “you do not chop the trapped person’s arm off”; so the question is asking how much you believe what the scenario says will happen. The second likeliness question asked what subjects thought about the stipulated outcome of performing the action:

L2) If you DO chop the trapped person’s arm off, how likely do *you think* it is that the door will seal and you and the six other researchers in the room will be safe, while the trapped person will drown outside?

Again, subjects are being asked how much they believe what the scenario says will happen.

Lastly, subjects were asked a Yes/No comprehension question, to detect if they were paying attention to relevant details of the scenario:

In this scenario, is the research facility flooding with water?

To look for a self-other bias within these personal dilemmas, each of the two surveys, 1 and 2, presented subjects with 4 dilemmas in the Self condition (where the subject was the proposed actor, as in Underwater Hatch above) and 4 dilemmas in the Other condition (where another person, e.g. “John”, was the proposed actor). And there were two versions of each

survey, such that the two versions had opposite dilemmas in the Self and Other conditions. So for Survey 1, version A, 4 dilemmas were in the Self condition, and 4 in the Other condition—and for Survey 1, version B, those same dilemmas were placed in the opposite condition (and similarly for the 8 dilemmas of Survey 2). Thus, each of the 16 dilemmas was in the Self condition in one version of a survey, and in the Other condition in another version of that survey.

To control for order effects, each survey version (A and B) consisted of two blocks (X and Y) of 4 dilemmas. The blocks consisted of 2 Self and 2 Other dilemmas quasi-randomized to alternate between Self and Other dilemmas. Within a given survey version, one block began with a Self dilemma, while the other block began with an Other dilemma. So, consider Survey 1, version A. Half of subjects saw the following order of dilemmas, the other half reversed the two blocks: Block X (Dilemma 1 [Self], Dilemma 2 [Other], Dilemma 3 [Self], Dilemma 4 [Other]), Block Y (Dilemma 5 [Other], Dilemma 6 [Self], Dilemma 7 [Other], Dilemma 8 [Self]). Survey 1, version B presented the exact same dilemmas in the same orders, except that each dilemma was flipped between the Self and Other conditions. Survey 2 (versions A and B) followed the same pattern, except with Dilemmas 9-16.

5.2 Results and analyses

Subjects who missed more than two of the eight comprehension questions on their survey were excluded from the analyses (Survey 1, 24 excluded; Survey 2, 19 excluded). Also, subjects who took more than 60 minutes or less than 5 minutes were excluded, because it is likely that they were not taking the survey all at once, or were moving through the survey without actively participating (an additional 8 from Survey 1, and 5 from Survey 2). So for the analyses, there were 168 subjects in Survey 1 and 161 subjects in Survey 2.

To determine if there was an effect of the Self versus Other conditions on moral responses (to question M), independent samples t-tests (two-tailed) were performed on each of the dilemmas, comparing moral judgments across conditions. There were no significant differences in moral responses between the Self and Other conditions for any of the 16 moral dilemmas.

To determine if subjects' non-moral judgments predict their moral intuitions, three separate multiple regression analyses were performed on each dilemma (the data is presented in Appendix A). The following background factors were controlled for in all of the regression analyses (and were entered in the regression equations in this order): gender, whether the subjects have completed a philosophy course, the course they were participating in the survey for (Philosophy or Psychology), the dilemma version (Self or Other), and dilemma order. For all of the dilemmas, L1 was a significant predictor ($p < 0.001$) of moral responses when these factors were controlled for, in addition to the predictive effects of L2 being controlled for. For 10 of the 16 dilemmas, L2 was a significant predictor ($p < 0.05$) of moral responses when the background factors were controlled for, in addition to the predictive effects of E2 being controlled for. L2 was a significant predictor ($p < 0.10$) for 4 of the dilemmas, and non-significant for 2 dilemmas (Secret Agent and Remote Waterfall). Also, for all of the 16 dilemmas, a composite likeliness rating (the sum of a subject's responses to L1 and L2) was a significant predictor ($p < 0.001$) of moral responses when the background factors were controlled.

5.3 Discussion

There was no significant difference between subjects' moral judgments in the Self and Other conditions. The present study did not go far enough to ensure that the moral judgments subjects made were representative of subjects' moral *intuitions*, rather than moral beliefs that

they formed inferentially or that they would easily give up; this is something that researchers of moral psychology should focus on in future studies. However, on the assumption that the judgments do indicate intuitions (an assumption proponents of the argument from cognitive biases would endorse), we can say that there was no significant difference between subjects' moral intuitions in the Self and Other conditions. That is, the present study suggests that there is not a self-other cognitive bias in moral intuitions about personal moral dilemmas. Further investigation will be required to determine if there is a self-other bias of moral intuitions at all, and if so whether it is restricted to a bias of intuitions about impersonal dilemmas. Thus, contrary to Nadelhoffer and Feltz, it is not clear that the self-other bias can offer much support for premise C4a* of the argument from cognitive biases, which holds that moral intuitions are sufficiently subject to cognitive biases to show that they are unreliable. If intuitions about impersonal dilemmas are, in fact, subject to a self-other bias, while intuitions about personal dilemmas are not, this would still create a lot of biased moral intuitions in the 'moral intuition' class of moral beliefs. But given the early nature of the evidence, the fact that impersonal dilemmas have been shown to be affected by a self-other bias, while personal dilemmas have not, suggests that further investigation is needed to assess whether or not, and to what extent moral intuitions are affected by a self-other bias.

Nonetheless, as discussed in Section 2.3, there remain a variety of other cognitive biases of moral intuitions about other dilemmas that are quite well established, and that can be used to support C4a*. So I take it that the more important aspect of the present study is the high degree of correlation between subjects' moral intuitions and their non-moral judgments; non-moral judgments were highly significant predictors of moral intuitions for each of the 16 dilemmas. This finding is not interesting simply because it shows a link between subjects' moral intuitions

and non-moral judgments. Obviously the non-moral judgments we make about a case affect our moral intuitions. When we judge someone to be shooting a gun in target practice, we do not intuit it to be wrong, but when we judge someone to be shooting a gun at an innocent child, we do intuit it to be wrong. Our different judgments of the non-moral facts about what the gun is being shot at helps explain why we form different moral intuitions.

What is interesting here is instead that subjects' non-moral judgments about *how believable they found the scenarios* predict their moral intuitions. It seems that subjects' moral intuitions about the 16 moral dilemmas examined in the present study were not formed simply based on the stipulated facts of the dilemmas—that is, on the stipulated likeliness of the outcomes of performing or not performing the proposed actions (outcomes which were always said to be what *will* happen). Rather, their moral intuitions were formed using their own judgments about the likeliness of the outcomes, which differed across the scenarios. Thus, it is likely that when responding to a moral dilemma like the ones taken to show biasing of moral intuitions (from Section 2.3), subjects do not respond based on the stipulated facts of the case, like the stipulated likeliness of the outcomes of the potential courses of action (e.g. “You can throw the switch, saving the five innocent people, which will result in the death of the one innocent person on the side track”; Petrinovich and O’Neil 1996). Rather, they respond based on what they judge the likeliness of the outcomes to be, even when the scenario specifies what the outcome will be. For example, in Underwater Hatch, when a subject judged that it is ‘not at all likely’ that *not* chopping off the person’s arm will result in the room from flooding, she was likely to judge that chopping off the person’s arm is wrong. It is clearly wrong to chop off the person’s arm if that action is not necessary to prevent the room from flooding (i.e. if the room will not flood—or if you do not think it is likely to flood—if you *do not* chop off his arm). The

scenario stipulated that if you do not chop off his arm, then the room will flood; but subjects varied in how probable they thought this stipulation was, and those who thought it was less probable were more likely to judge chopping off the arm as wrong.

The ability to use non-moral judgments to predict moral intuitions supports the intuitionist's objection from the non-moral basis of cognitive biases of moral intuitions, which rejects premise C4b* by arguing that cognitive biases of moral intuitions are grounded in biases of non-moral intuitions. People's non-moral judgments about the believability of moral dilemmas predicts their moral responses—so it is possible that some of the cognitive biases of moral intuitions used to support premise C4a* (discussed in section 2.3) are grounded in biasing effects on subjects' non-moral judgments, such as biases of judgments about the likeliness of the stipulated outcomes of the dilemmas.

This is just one sort of non-moral judgment that might be affected by a cognitive bias, leading to the biasing of moral intuitions. But there are many other non-moral judgments that might be biased in the relevant dilemmas—like subjects' beliefs about the facts of the case that are not described in the short presentation of the dilemma (like the character traits of the other people described in the dilemma, their health, the likely future consequences of the actions, and so on). For example, Kuhn showed that people's non-moral judgments about the probability of certain events—probabilities which were unspecified in the descriptions of the scenarios—are affected by framing effects (1997).

Admittedly, the present study has demonstrated only a correlation between moral intuitions and non-moral judgments—so it does not directly support the intuitionist's objection from the non-moral basis of cognitive biases. This objection requires that the differences in a person's moral intuitions are *caused by* the differences in her non-moral judgments. But the

correlation between moral and non-moral judgments could be explained by saying that the differences in non-moral judgments are caused by differences in moral intuitions. For example, moral and non-moral judgments might be correlated because subjects adjust their non-moral judgments in a way to rationalize their moral intuitions. In the Underwater Hatch dilemma, a subject might say that chopping off the trapped person's arm is wrong, and then rationalize that judgment by saying that it is 'not at all likely' that (i) the room will flood if she does *not* chop off his arm, and that (ii) the room will not flood if she *does* chop off his arm. That is, after saying it is wrong to chop off the person's arm, she might rationalize that judgment by saying that not chopping of the person's arm wont result in the room flooding, and that chopping the person's arm off wont prevent the room from flooding—making it obviously wrong to chop off the person's arm, since doing so has no beneficial consequences. She rationalizes her judgment by thinking that the room will flood or not flood regardless of whether or not she chops off the person's arm.

However, I think there is good reason to think that the causal arrow goes the other way, with the differences in non-moral judgments causing the differences in moral intuitions. As discussed in Section 4.2, it seems likely that part of ordinary moral psychology is the thesis of moral supervenience, which holds that if two cases do not differ in their non-moral properties, then they do not differ in their non-moral properties. And this would suggest that within-subject differences in moral intuitions (which would include cognitive biases of intuitions) should be grounded in within-subject differences in non-moral judgments; given the thesis of moral supervenience, a given subject would not have different moral intuitions about two cases unless she made different non-moral judgments about them. Thus, the fact that subjects' non-moral, epistemic judgments are significant predictors of their moral intuitions provides is best explained

by the non-moral judgments playing a causal role in the formation of subjects' moral intuitions. This makes it seem very likely that many cognitive biases of subjects' moral intuitions are grounded in biases of their non-moral intuitions; that is, premise C4b* of the argument from cognitive biases is plausibly false, and the argument fails to go through.

There are a variety of limitations to the present study that should make us cautious in interpreting the results. Most importantly, the claim that ordinary moral psychology includes the thesis of moral supervenience is an empirical claim that to my knowledge has not been evaluated. It seems likely to be vindicated, but without such vindication at present my claim that the differences in non-moral, epistemic judgments causally contribute to differences in moral intuitions is largely speculative and should be understood as such. Also, as already mentioned, this research has failed to address the difference between moral beliefs and moral intuitions; the moral judgments subjects make might represent moral intuitions (qua strongly held, non-inferentially formed moral beliefs), or they might represent mere moral beliefs that subjects would easily give up.

6 Conclusions

To conclude, I have attempted to show the parallel between Sinnott-Armstrong's recent argument from cognitive biases and the well-known argument from disagreement. Because of this tight parallel, the argument from cognitive biases is subject to an objection that is often levied against the argument from disagreement; many of the cognitive biases cited to show that moral intuitions are unreliable are not genuine, they are derived. That is, the cognitive biases of moral intuitions are grounded in biases of non-moral intuitions. And derived biases of moral intuitions, like derived moral disagreements, can be resolved by resolving the relevant non-moral biases, which would remove the unreliability of moral intuitions without making them inferable.

Since that unreliability was the reason moral intuitions required inferential justification in the first place, when it is removed without making moral intuitions inferable, those intuitions can be justified non-inferentially. Thus, derived cognitive biases would not threaten a moderate moral intuitionism, which holds that moral intuitions can be justified under ideal conditions. There may be reasons why moral intuitions cannot be justified, inferentially or non-inferentially, but cognitive biases are not among of them, contrary to the claims of proponents of the argument from cognitive biases (Sinnott-Armstrong 2008a, 2008b; Nadelhoffer and Feltz in prep.; Vayrynen forthcoming).

References

- Banjai, M., Greenwald, A. (1995). Implicit Gender Stereotyping in Judgments of Fame. *Journal of Personality and Social Psychology* 68: 181-198
- Brandt, R.B. (1961). *Ethical Theory*. Prentice-Hall: Englewood Cliffs, NY.
- Brauer, Wasel & Niedenthal (2000). Implicit and explicit components of prejudice. *Review of general psychology* 4: 79-101.
- Brink, D. (1984). Moral Realism and the Sceptical Arguments from Disagreement and Queerness. *Australasian Journal of Philosophy* 62: 2: 111-125.
- Doris, J.M. & Stich, S.P. (2005). As a Matter of Fact: Empirical Perspectives on Ethics. In *The Oxford Handbook of Contemporary Philosophy*, Jackson, F. & Smith, M. (eds.). Oxford University Press.
- Greene, J.D., et. al (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment.
- Haidt, J. & Baron, J. (1996). Social Roles and the Moral Judgment of Acts and Omissions. *European Journal of Social Psychology* 26: 201-218.
- Hauser, M.D. (2006): *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: Ecco/Harper Collins.
- Hooker, Brad (2002). Intuitions and Moral Theorizing. In *Ethical Intuitionism*, Stratton-Lake, P. (ed.). Oxford University Press: New York.
- Hume, D. *A Treatise of Human Nature*.
- Kuhn, K. (1997). Communicating Uncertainty: Framing Effects on Responses to Vague Probabilities. *Organizational Behavior and Human Decision Processes* 71: 55-83
- Levin, I.P., Schneider, S.L., & Gaeth, G.J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes* 76: 149-188.
- Loeb, D. (1998). Moral Realism and the Argument from Disagreement. *Philosophical Studies* 90: 281-303.
- Machery, E., Kelly, D., & Stich, S.P. (2005). Moral Realism and Cross-Cultural Normative Diversity. *Behavioral and brain sciences* 28: 830-830.
- Malle, B. F. (2006). The actor-observer asymmetry in causal attribution: A (surprising) meta-analysis. *Psychological Bulletin* 132: 895-919.

- Mikhail, J., Sorrentino, C., and Spelke, E.S. (1998). Toward a universal moral grammar. *Proceedings of the Cognitive Science Society*.
- Moore, G.E. (1903). *Principia Ethica*. Cambridge University Press: Cambridge.
- Nadelhoffer, T. & Feltz, A. (in prep.). The Actor-Observer Bias and Moral Intuitions: Adding Fuel to Sinnott-Armstrong's Fire.
- Petrinovich, L. & O'Neill, P. (1996). Influence of Wording and Framing Effects on Moral Intuitions. *Ethology and Sociobiology* 17: 145-171.
- Ross, W.D. (1930). *The Right and the Good*. Revised edition (2002), Philip Stratton-Lake (ed.). Oxford University Press: Oxford.
- Shafer-Landau, R. (2008). Defending Ethical Intuitionism. In *Moral Psychology, Volume 2: The Cognitive Science of Morality*, Sinnott-Armstrong, W. (ed.). MIT Press: Cambridge.
- Sinnott-Armstrong, W. (2002). Moral Relativity and Intuitionism. *Philosophical Issues* 12: 305-328.
- Sinnott-Armstrong, W. (2006a). *Moral Skepticisms*. Oxford University Press: New York.
- Sinnott-Armstrong, W. (2006b). Moral Intuitionism Meets Empirical Psychology. In *Metaethics After Moore*, Horgan, T. & Timmons, M. (eds.). Oxford University Press: New York.
- Sinnott-Armstrong, W. (2008a). Framing Moral Intuitions. In *Moral Psychology, Volume 2: The Cognitive Science of Morality*, Sinnott-Armstrong, W. (ed.). MIT Press: Cambridge.
- Sinnott-Armstrong, W. (2008b). Reply to Tolhurst and Shafer-Landau. In *Moral Psychology, Volume 2: The Cognitive Science of Morality*, Sinnott-Armstrong, W. (ed.). MIT Press: Cambridge.
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology* 27: 76-105.
- Tolhurst, W. (2008). Moral Intuitions Framed. In *Moral Psychology, Volume 2: The Cognitive Science of Morality*, Sinnott-Armstrong, W. (ed.). MIT Press: Cambridge.
- Tversky, A. & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science* 211: 453-458.
- Vayrynen, Pekka (forthcoming). Some Good and Bad News for Ethical Intuitionism. *Philosophical Quarterly*.

Appendices

Appendix A: Regression Statistics

The dependent variable was subjects' moral responses. You can read the text of the scenarios in Appendix B. The following were controlled for (and were entered in the regression equation in this order): gender, whether the subjects have completed a philosophy course, the course they were participating in the survey for (Philosophy or Psychology), the dilemma version (Self or Other), and dilemma order. After these factors were partialled out, the predictor variable was entered, and the data below represents the change in the regression model's ability to predict the dependent variable when the predictor was added. E1 is the first non-moral question (asking "If you do NOT X, how likely is it that Y?"). E2 is the second non-moral questions (asking "If you DO X, how likely is it that Y?"). E Composite is the sum of a subject's responses to E1 and E2, and represents an overall non-moral judgment of the believability of the scenario; that is, a non-moral judgment about the believability of the outcomes that are stipulated in the scenario.

*significant at $p = 0.05$ **significant at $p = 0.01$ ***significant at $p = 0.001$

Dilemma	Predictor	ΔR^2	F ratio (for ΔR^2)	p value
Secret Agent	E1	0.036	7.167	0.008**
	E2	0.005	0.961	0.328
	E Composite	0.42	8.391	0.004**
Underwater Hatch	E1	0.149	34.836	0.000***
	E2	0.015	3.582	0.060
	E Composite	.0201	45.378	0.000***
Remote Waterfall	E1	0.175	42.985	0.000***
	E2	0.010	2.412	0.122
	E Composite	0.194	45.599	0.000***
Hospital Worker	E1	0.164	36.557	0.000***
	E2	0.015	3.432	0.066
	E Composite	0.155	33.240	0.000***
Sniper	E1	0.102	23.206	0.000***
	E2	0.038	8.630	0.004**
	E Composite	0.177	40.053	0.000***
Antidote	E1	0.065	13.999	0.000***
	E2	0.018	3.919	0.049*
	E Composite	0.108	23.320	0.000***
Space Shuttle	E1	0.107	25.101	0.000***
	E2	0.023	5.297	0.023*
	E Composite	0.182	42.146	0.000***
Desert Burial	E1	0.144	32.998	0.000***
	E2	0.014	3.296	0.071
	E Composite	0.156	34.497	0.000***
Bunker	E1	0.111	22.095	0.000***
	E2	0.045	9.082	0.003**
	E Composite	0.133	26.451	0.000***

Hot Air Balloon	E1	0.120	29.690	0.000***
	E2	0.044	10.979	0.001**
	E Composite	0.233	57.643	0.000***
Firefighter Bucket	E1	0.066	12.062	0.001***
	E2	0.019	3.393	0.067
	E Composite	0.101	18.370	0.000***
Agonizing Choice	E1	0.067	13.039	0.000***
	E2	0.086	16.872	0.000***
	E Composite	0.156	30.750	0.000***
Seaside Cave	E1	0.140	30.769	0.000***
	E2	0.021	4.593	0.034*
	E Composite	0.154	32.520	0.000***
Crying Baby	E1	0.090	17.920	0.000***
	E2	0.068	13.619	0.000***
	E Composite	0.170	34.114	0.000***
Mountain Climbing	E1	0.106	21.299	0.000***
	E2	0.030	5.974	0.016*
	E Composite	0.156	31.291	0.000***
Mine	E1	0.043	7.866	0.006**
	E2	0.022	4.113	0.044*
	E Composite	0.112	20.766	0.000***

Appendix B: Moral Dilemmas

Survey 1

Secret agent (Self version)

You are a government agent who has been captured by the enemy. They want you to give them the code to access information on the location of a bomb. If you do give them the code, the enemy will gain access to a bomb which they will use to blow up a government building, resulting in hundreds of deaths.

If you do not give them the code, they will torture an unknown four-year-old child to death right before your eyes.

Underwater Hatch (Other version)

You work in an underwater research facility. There has been an explosion and water is pouring into the facility, causing all of the doors to begin closing automatically. You, the facility director, and five other researchers are in a room that is being sealed off by a closing door. At the last minute, someone in the hall tries to make it to safety in the room by sliding under the closing door, but he is too late and the door closes on his arm, trapping him outside and making it impossible for him to get into the room. This also prevents the door from closing and allows water to rush into the room.

If nothing is done, the water will flood the room and drown you and everyone else inside, including the person trapped under the door. The only way to save you and the other researchers is for the facility director to chop the trapped person's arm off with an axe, leaving him out in the hallway where he will drown, allowing the door to close and preventing the room from flooding.

Remote waterfall (Self version)

John, you, and your four children are hiking alongside a tropical river on your way to explore a remote waterfall. You have accidentally walked across a local river tribe's sacred ritual site.

For this offence, tribal law says that all of you must be put to death. However, the leader of the tribe has decided that he will let John, you, and your three oldest children live if you yourself will sacrifice your youngest child by drowning her in a sacred river ceremony.

Hospital Worker (Other version)

Your son is a patient in the hospital who needs an organ transplant. He and four other patients will soon die, each due to a different failing organ because there are no organs available to save them. There is another patient in the hospital who is in a coma on life support and her doctors agree that she will never wake up.

The only way that your son and the four other dying patients will live is if Fred, who works in the hospital, pulls the plug on the comatose patient's life support machine, killing her so that her organs will be available for transplanting. If Fred does this, then the comatose patient will die and your dying son and the four other dying patients will receive their transplants and will be able to live.

Sniper (Other version)

Tom is the sniper in a police SWAT team which is positioned outside of the office building where your daughter works. A terrorist has taken twenty people, including your daughter, hostage inside the building and is about to detonate a bomb, which would blow up the building, killing all of the hostages.

The only way to prevent the terrorist from blowing up the office building is for Tom to shoot him dead before he can detonate the bomb. The terrorist is protecting himself by holding your daughter hostage in front of him so that Tom must shoot and kill her in order to kill the terrorist.

Antidote (Self version)

You are a scientist working in a small mountain village. A deadly substance has contaminated the local water supply and villagers are starting to get sick. Soon hundreds of villagers will be dead, including you. You have just created an antidote for this deadly substance, but have only one vial. You are taking the vial to your lab, where you can replicate the antidote and save yourself and the other villagers.

On your way to the lab you come across a villager who has already become very sick and will die if he does not immediately get the antidote. If you give this man the antidote, he will

survive and you will have no antidote left to replicate. As a result, there would be no way to save yourself and the other villagers.

Space shuttle (Other version)

You and your son are astronauts onboard a space shuttle that is about to return from a visit to the moon. An explosion has caused the shuttle to lose most of its fuel supply. Even after dumping as much weight as possible off of the shuttle, there is still not enough fuel for the shuttle to make it back from the moon. If the shuttle does not get back soon it will run out of fuel, and everyone on board will die, including you, your son, and the other four astronauts.

One of the astronauts has become seriously ill and will not survive no matter what. If the captain of the shuttle dumps this fatally ill astronaut into outer space, there will be just enough fuel and the shuttle will carry the captain, you, your son, and the other two astronauts back from the moon safely.

Desert burial (Self version)

You are a tour guide and are taking your two children to visit ancient ruins in the middle of the desert. You and your children have accidentally wandered into an off-limits tomb belonging to a local clan.

The clan's law says that any children who are found in the off-limits tomb must be put to death. However, the elders of the clan have taken a liking to your children and have decided that they will let them live if you yourself will sacrifice a young child from the clan by burying her alive.

Survey 2

Bunker (Other version)

Joe is a military commander stationed in a sealed bunker deep underground. One of his soldiers contracted a deadly virus before entering the bunker, which is now sealed and cannot be opened. Joe is the only person in the bunker who has been vaccinated against this virus. If the infected soldier continues breathing the virus into the air then it will spread quickly, killing the other twenty soldiers in the bunker, except for Joe.

In order to prevent the deadly virus from spreading, Joe must shoot dead the infected soldier, which will stop him from breathing the virus into the air.

Hot air balloon (Self version)

You are a hot air balloon pilot flying a group of ten sightseers high in the air on a hot air balloon ride. You realize that the balloon is carrying more people than it was designed to carry and is beginning to strain under the weight. The balloon will soon fall sending you and the sightseers plummeting to the ground, killing you all.

If nothing is done, the balloon will crash and you and everyone else on board will fall to your deaths. However, you know that there is an elderly woman on board who will die from a

terminal illness within a week. If you throw that woman off of the hot air balloon, the balloon will not fall and you and the other eight sightseers will make it down to the ground safely.

Firefighter Bucket (Other version)

A firefighter has just rescued you, your child, and your two neighbors from the 15th floor of your burning apartment building. He has deposited the four of you at the top of his fire truck's ladder, which is lowering you down to safety. There is too much weight on the ladder, and it begins to break.

If nothing is done, the ladder will break, and you, your child, and your neighbors will fall 15 stories and be instantly killed. If your one neighbor pushes your other neighbor, who happens to be very large, off of the ladder, this will prevent the ladder from breaking so that you, your child, and your one neighbor can be brought down safely.

Agonizing choice (Self version)

You and your three children are inmates in a concentration camp. Your youngest child has tried to escape from the camp. To make an example of her a guard is going to hang her in the middle of the camp.

He has stood her up on a chair and put a noose around her neck. The guard has told you to hang her by pulling the chair out from underneath her feet, causing her to fall and be strangled to death. If you do not hang her, the guard has decided he will personally hang her and your two other children.

Seaside Cave (Self version)

You are a geologist who is exploring a tiny cave on a secluded coastline with a group of ten of your colleagues. You are the first person to exit the cave, and the person behind you gets stuck in the entrance, trapping the rest of the group inside the cave. In a short time high tide will be upon the cave.

The person cannot be pulled from the cave entrance no matter how hard you try, and unless he is unstuck, he and the rest of the group, except for you, will drown. You have brought with you a stick of dynamite. If you use the dynamite to blow the person from the entrance, the person will be blown to pieces and the other geologists will be able to escape the cave safely. Since you are already outside of the cave, you will be safe no matter what.

Crying baby 2 (Other version)

Your apartment building has just been taken over by a local militia. They have orders to kill anyone they find inside. When the militia took over, you were painting your kitchen while your neighbor Judy watched your baby. Judy (who is holding your baby) and eight other neighbors have taken cover in a dark hidden attic. You made it to the attic just in time to hide, and outside you hear the militia searching the apartment for valuables.

Your baby begins to cry loudly. Judy covers his mouth to block the sound. If she removes her hand from his mouth his crying will get the attention of the militia, who will kill Judy, your

baby, you and the others hiding in the attic. If Judy is to save herself, you, and the other neighbors, she must continue covering your baby's mouth until she smothers him to death.

Mountain Climbing (Self version)

You are leading a climbing expedition, which includes your son, up a frozen mountaintop. A climber attached to the same rope as you, your son, and another person is falling off the side of a cliff, but you grab her hand as she falls.

The weight of the fallen climber begins to drag you, your son, and the other person off of the cliff. If you do nothing, the three of you will be pulled over the cliff by the fallen climber and you will all plummet to your deaths. The only way for you to save yourself, your son, and the other person is to let go of the fallen climber's hand, dropping her to her death.

Mine 2 (Other version)

Your grown-up children work in a small underground mine. A dam near the mine has burst, sending water rushing toward the mine. If nothing is done, the water will flood the section of the mine where your children and eight other miners are working, causing them all to drown.

The mine manager is safe in his office, and the only way for him to avoid the deaths of your children and the other miners working in the mine is to radio a worker who is safe outside the mine and order her to seal off the mine from the outside. If the manager does this, the worker will not be able to get to safety before the water comes and she will be swept away just outside the mine and killed.