THE PSYCHOLOGY AND EPISTEMOLOGY OF (MOSTLY MORAL) INTUITIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mark William Fedyk

August 2009

THE PSYCHOLOGY AND EPISTEMOLOGY OF (MOSTLY MORAL) INTUITIONS

Mark William Fedyk, Ph.D.

Cornell University 2009

This dissertation is composed of four stand-alone papers, organized here as four chapters. The first chapter gives a philosophical account of the nature of intuitive judgments. It proposes a conceptual framework that captures what are hopefully the essential properties of intuitions, and offers a description of the conditions under which intuitions will be reliable. The second chapter considers and rejects a recently popular theory in moral psychology, the linguistic analogy. According to this theory, human moral cognition is importantly similar to linguistic cognition, just so long as the later is understood using the theory of universal generative grammar that is currently fashionable in contemporary linguistics. The third chapter considers and rejects another recently popular theory in moral psychology. This theory, called the social intuitionist model of moral judgment, holds that moral reasoning does not function to promote moral truth. Rather, the proper function of moral reasoning is to create patterns of agreement in both people's moral intuitions and any attendant moral sentiments. Finally, the last chapter of this dissertation argues against the currently established view that moral intuitions ought to occupy an epistemically privileged role in moral inquiry. It uses Frank Jackson's moral epistemology as a stalking horse, and in contrast to some elements of his epistemology, the chapter outlines a view of reflective equilibrium that explains how more sources of moral insight than just moral intuitions can play an evidential role in moral inquiry.

BIOGRAPHICAL SKETCH

Mark William Fedyk received a B.A. (Hons) in Philosophy from Queen's University, Kingston, Ontario, in 2004. He considered a career as a computer systems analyst for a while, but turned to philosophy after experiencing the dreariness of the day-to-day cubicle lifestyle. He received a Ph.D. in Philosophy in 2009 from Cornell, and is teaching for a short stint there while he figures out what to do next.

ACKNOWLEDGEMENTS

Many people have helped me with this project. Deserving of the most thanks is the chair of my special committee, Richard Boyd. He has been extremely generous with his time, support, and expertise over the years. A student could not ask for a better mentor and supervisor. Thanks, Dick.

I'm also grateful for the patient help and encouragement I have received from the other members of my special committee, Andrew Chignell and Nicholas Sturgeon, and to my parents, Frank and Sue, for their much appreciated assistance. My colleagues at Cornell (most of all Elina Nurmi and Nate Jezzi) also deserve a word of thanks for all the stimulating conversations over the years.

Additionally, I'd like to credit the faculty of the Sage School of Philosophy for their instruction and guidance; the students who have taken my classes at both Cornell and the Elmira Correctional Facility, for helping me become both a better teacher and a better writer; the Cornell graduate students who put together the LaTeX dissertation template; the Institute of Humane Studies for their financial support; and, finally, Stephen Leighton, for his encouragement, his introduction to Aristotle, and, frankly, his patience with my early philosophical stubbornness.

My good friend Michel Shamy deserves my gratitude for years of excellent conversation about matters philosophical. He also deserves credit for (perhaps) unintentionally steering my research towards ethics, and for showing me new ways of thinking about old philosophical issues.

Lastly, I want to say a special word of thanks to Stephanie Bedolla for her devotion, intelligence, patience, and support. Steph, I'm sorry I had to work so hard. But I couldn't have completed this project without you.

TABLE OF CONTENTS

CHAPTER 1

PHILOSOPHICAL INTUITIONS

1.1 Introduction

Many philosophers use intuitions as a source of evidence. But recently, interest in the epistemology and psychology of intuitions themselves has increased dramatically, and some philosophers – most of whom are proponents of a research programme called "experimental philosophy" – have argued that we now have evidence that shows that we should be sceptical of philosophical intuitions. These philosophers contend that we should no longer use just the intuitions of philosophers as evidence in philosophical theorizing (if we need to use intuitions at all), because it can be shown on the basis of experimental data that non-philosophers typically have intuitions that do not agree with the intuitions of philosophers, and that the intuitions of non-philosophers vary according to factors that are seemingly philosophically irrelevant, like the intuitor's socio-economic status.

The present paper enters into the debate initiated by the arguments of the experimental philosophers. The paper's main goal is to clarify the notion of an intuition as it is used in contemporary philosophy and also in the burgeoning experimental literature treating the psychology of intuitions.[1] In so doing, the paper is most concerned with offering a coherent and explanatorily useful

[1]This is, roughly, the use of 'intuition' that is found in the works of Chomsky, Gettier, and Kripke. I believe that this is a logically distinct notion of intuition (see section 4.5 below), compared to the notion of intuitions most closely associated with Ethical Intuitionism; i.e. the notion of intuition familiar in the works of Moore and Ross. However, Audi 2004 and Huemer 2005 both offer views that link the contemporary notion of an intuition with the older notion of an intuition.

account of intuitions that tries collect together many of the properties of intuitions that philosophers have remarked on. The aim, in short, is to provide a report on what might be called the "standard view" of intuitions.[2] And so, in service of this goal, I will provide an analysis of the structure of intuitions, offer an assessment of the conditions under which intuitions will be reliable sources of evidence, and explain why one of the central lines of experimental philosophy's critique of the use of intuitions in philosophy fails. My hope is that the results of this paper will be helpful to both traditional philosophers and those researchers (including experimental philosophers) who are studying intuitions experimentally.

The paper is organized in the following way. I start with a discussion of the structure of intuitions in section 1.2, in the course of which I establish a terminological framework that allows us to talk with an appropriate degree of precision about the properties of intuitions. Sections 1.3 and 1.4 offer some further terminological refinements to this framework. I then turn to the epistemology of intuitions in section 1.5, where I identify some of the conditions that intuitions must meet in order for them to be properly treated as a source of evidence. Section 1.6 then uses the results of the preceding sections to show that, for technical concepts, this paper's view of the epistemology of intuitions implies that we should defer to the intuitions of experts. Section 1.7 shows that a similar conclusion holds for many of the concepts that interest contemporary philosophers, and then draws from this some implications for the sceptical argument from experimental philosophy concerning the use of intuitions as evidence in

[2]It is unsurprising that, given their ubiquity in philosophical practice, nearly every philosopher has at some point in their career offered some insightful remarks concerning the properties of intuitions. Clearly, it isn't possible to catalogue here all these remarks. But for a different attempt to capture the consensus view in philosophy, one which places more emphasis on demonstrating how philosophical intuitions can be a source of a priori knowledge than the present account, see Bealer 1996 and Bealer 1998.

philosophical theorizing. Brief concluding comments are offered in section 1.8.

1.2 Paradigm Examples

We begin with an often recited platitude about the function of intuitions within philosophy. It is often claimed that intuitions involve or somehow reveal what our concepts are. Hilary Kornblith, for example, writes "appeals to intuition are designed to illuminate the contours of our concepts."[3] One way of cashing out this platitude might be to understand intuitions as a kind of judgment about the meaning of a concept. But this cannot be right. For, as we will see in following sections, intuitions are not literally *about* our concepts or their meanings, despite the fact that there are conditions under which intuitions can be a good source of evidence about the meanings of our concepts.

In fact, in order to capture some of the more subtle aspects of intuitions, I believe that we need a better understanding of the structure of intuitions. So, here is, to a good first approximation, what I take the structure of an intuition to be. Intuitions are *about* the *salient feature(s)* of a *case*. The *salient feature(s)* of a *case* are the *object* of an intuition. The *propositional content* of an intuition follows from the *implicated concept*. Thus, the basic idea is that an intuition is about the salient features of a case, it has propositional content, and the propositional content of an intuition is obtained in some way from the implicated concept.

The argument for this terminology is straightforward. I will take four examples of paradigmatic appeals to intuition and show that the distinctions just

[3]Kornblith 2006 p.11

mentioned can be used to make sense of each example.[4] My examples will be the first of Edmund Gettier's two appeals to intuition, Judith Jarvis-Thompson's famous violinist case, an example from Gilbert Harman, and about half of Hilary Putnam's twin earth scenario.

For example, Gettier asked us to consider the following familiar case in his famous 1963 article:

> Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:
>
> > (d) Jones is the man who will get the job, and Jones has ten coins in his pocket.
>
> Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones' pocket ten minutes ago. Proposition (d) entails:
>
> > (e) The man who will get the job has ten coins in his pocket.
>
> Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true. But imagine, further, that unknown to Smith, he himself, not Jones, will

[4] A point of clarification: I am not offering an analysis of thought experiments. Of course, some of the most famous thought-experiments in 20^{th} and 21^{st} century analytic philosophy consist of, basically, a single appeal to an intuition about a hypothetical case. But it nevertheless seems to me that thought-experiments both within and outside of philosophy usually consist of more than just an appeal to an intuition. So perhaps in order to understand many thought-experiments we need to first understand intuitions, but I do not think that a single explanation will be able to account for both appeals to intuitions and thought-experiments.

> get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e) is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not know that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones' pocket, whom he falsely believes to be the man who will get the job.[5]

First, a small interpretative claim. When Gettier claims that "it is equally clear that Smith does not know that (e) is true", he is reporting an intuition. I believe that this is the standard interpretation of this passage.

Now, it makes sense to ask what the intuition is about. Following the platitude mentioned above, a natural answer is this: the intuition is *about* our concept of knowledge. But this is not exactly right. For on further examination, it looks as though the intuition is really about the imaginary scenario described by Gettier - i.e., the *case*. In fact, it is more precise to say that the intuition is about some of the properties of the case, namely the fact that Smith's belief is true and that it has been derived from things that Smith already knows, but that Smith's reasoning relied on (unbeknownst to him) false lemmas. We can say that these properties constitute Smith's doxastic state, and that these properties are the *salient features* of the *case*, and they are therefore the *object* of the intuition.

It is obvious that intuitions have *propositional content*. This is what we 'get'

[5]Gettier 1963 p.122

when we have an intuition. And in this example, it is clear that the propositional content of the intuition is what is expressed when Gettier reports his intuition: that Smith doesn't know that (e) is true.

Now, it should seem that the concept of knowledge is in some way or another *implicated* in the intuition. In the next section I'll clarify this claim further, but for now it will do to simply establish our terminology. The idea, in rough, is that the *propositional content* of an intuition is obtained by the intuitor applying her concept of knowledge to the case, and we say that her concept of knowledge is the *implicated concept*.

Putting the terminology together now, we have the following. The *object* of this particular intuition is the *salient feature* of the *case*, namely Smith's doxastic state. The concept of knowledge is *implicated* in this intuition. And what the intuitor of the intuition 'gets' when she has the intuition is the *propositional content* of the intuition, which, if you share Gettier's intuition, is: Smith does not know that (e) is true.

If by now you think that this account of the structure of an intuition is basically correct, or at least close enough to work with for the remainder of this paper, please feel free to skip ahead to the next section. Otherwise, let's move through the next three examples a little more quickly.

So, once again, we can see that the same distinctions can be used to make sense of Thompson's famous violinist case. Here is the relevant passage:

> You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of

> Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, "Look we're sorry the Society of Music Lovers did this to you - we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug you would be to kill him. But never mind, its only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you." Is it morally incumbent on you to accede to this situation? ... do you *have* to accede to it?[6]

The intuition that people usually report is that there is no moral obligation to agree to remain plugged into the famous violinist. Whether or not this case is appropriately analogous to the situation of an unwanted pregnancy is not our concern here.

Now, at first it may seem plausible that this intuition is about what morality requires of us, or what our moral duty would be in such circumstances. But for the same reason as above, this is not exactly right. For we can see that, literally, the intuition is about the imaginary choice to remain hooked up to the famous violinist or not. The conditions of this choice are the *salient features* of the *case*, and so we say that these features are the *object* of the intuition. Additionally, it looks as though our concept of moral duty is *implicated* in this intuition, while the *propositional content* of the intuition is, as we have just observed, that one

[6]Thompson 1971 p.48-49

need not agree to remain plugged into the famous violinist.

Let us turn to our two final examples. First, from Harman, and then from Putnam.

> You are a doctor in a hospital's emergency room when six accident victims are brought in. All six are in danger of dying but one is much worse off than the others. You can just barely save that person if you devote all of your resources to him and let the others die. Alternatively, you can save the other five if you are willing to ignore the most seriously injured person.
>
> It would seem that in this case you, the doctor, would be right to save the five and let the other person die.[7]

Once more we can see that this appeal to an intuition has the suggested structure. For it is clear that the object of this intuition is a salient feature of a case - that is, the intuition is about the imaginary choice concerning which patient to save as presented in the hypothetical E.R. scenario. The implicated concept appears once again to be the concept of moral duty, and the propositional content of the intuition is that it would be right to save the five and let the other person die.

Our final example will consist of only a part of Putnam's famous twin-earth argument, since I assume that readers are familiar with the details of the case

[7]Harman 1977 p. 3. Harman here thinks that philosophers are "reporting our feelings about an imagined example" (p.4), and so it may be that Harman once held an emotivist view of the psychology of moral intuitions. Recently, however, he has written in support of the idea that moral intuitions are derived from a developmentally-endogenous moral grammar. Put roughly, the idea is that moral intuitions are caused by a cognitive capacity that encodes moral knowledge in something like the structure of a generative grammar. C.f. Harman 2008.

and therefore nothing of importance is lost by omitting Putnam's full description. Here is the relevant passage.

> Suppose that somewhere in the galaxy there is a planet we shall call Twin Earth. Twin Earth is very much like Earth (...) One of the peculiarities of Twin Earth is that the liquid called "water" is not H_20 but a different liquid whose chemical formula is very long and complicated. I shall abbreviate this chemical formula simply as *XYZ*. (...) If a spaceship from Earth ever visits Twin-Earth, then the supposition at first will be that "water" has the same meaning on Earth and on Twin Earth. This supposition will be corrected when it is discovered that "water" on Twin Earth is *XYZ*, and the Earthian spaceship will report somewhat as follows: "On Twin Earth the word 'water' means *XYZ*"[8]

The case here is the scenario involving the various uses of "water" plus the details of the nature of Twin Earth, and the salient features of it are the "peculiarities" of Twin Earth including their use of "water". The intuition is about these salient features, and its propositional content is, of course, what the Earthian spaceship reports: that on Twin Earth the word "water" means *XYZ*. In this example, it is the concept of meaning that is implicated in the intuition.

So, we have now examined four examples of appeals to intuitions, and we've seen that each case the intuition can be sensibly interpreted using the following structure: the *object* of an intuition is the *salient features* of a *case*, that a particular concept is *implicated* in the occurrence of an intuition, that intuitions also have

[8]Putnam 1975 p. 223

propositional content, and that there is a sense in which the *propositional content* of an intuition is obtained from the *implicated concept*.

1.3 The Relationship between Concepts and Propositional Content in Intuitions

If the preceding proposal concerning the structure of an intuition is right, then it will be hard to say anything more about what makes some intuition or another reliable without saying something, first, about concepts, and second, about the relationship between a concept and the propositional content of an intuition. Sorting out these two issues is the purpose of this section.

So let me begin with concepts. The consensus amongst philosophers and psychologists is that concepts are mental representations and that they are referring entities. It is also uncontroversial, I believe, that some concepts are more accurate than other concepts,[9] which is to say, some concepts correspond better to the things that fall under them than other concepts, while some concepts correspond to nothing at all. Examples are easy to find. For instance, consider the concept of gravity as understood by Newtonian physicists compared with the concept of gravity as understood by physicists after the introduction of the theory of general relativity, and either concept of gravity compared to the concept of a unicorn or the concept of a square circle. In fact, I think that these three assumptions about the nature of concepts are all that is required to get a plausible epistemology of intuitions off the ground. My case for this contention is found

[9]That is, unless an atomistic conception of concepts is correct; see Fodor 2008. And see also Machery 2009, who, to put it very roughly, argues that the notion of a concept is too heterogeneous to have any explanatory use.

in the remainder of the paper.

But that said, I should mention one important caveat and introduce an important distinction. The caveat first. Beyond the assumptions just mentioned, I am not initially proposing to take on board any further commitments about the psychology or semantics of concepts.[10] That is, I want to remain neutral on the issue of the details of the underlying psychology that allows an individual to deploy a concept, and likewise I want to remain neutral on the issue of how to properly specify the meaning or intensional content of a concept – i.e., whether the meaning of a concept is given by a stereotype, an inferential role, an analytic definition, beliefs that encode a representation of a cluster of properties, or whatever else it may be. I do want to register my doubt that being competent with some particular concept always or necessarily involves knowing something like an analytic definition for the concept,[11] but most of the following arguments do not depend on any particular position on these various issues being correct.

So, with these points in mind, in the following I will talk about the salient features of a case 'satisfying' the intensional content of the implicated concept, and what I mean by this, of course, is that the salient features correspond to a sufficient number of the properties encoded in the representational structure that constitutes the intensional content of the intuitor's concept, whatever that structure may be, however knowledge of that structure is realized psychologically in the mind, and whatever number suffices as a sufficient number. The idea here is that, when applying a concept to a case, if the salient features satisfy the intuitor's concept, this will ordinarily cause an intuitor to affirm that the

[10] Nor, for that matter, am I taking on any deep commitments about the correct analysis of propositional content.

[11] C.f. Block and Stalnaker 1999

salient features fall under her concept; and if the salient features do not satisfy the intuitor's concept, this will ordinarily cause her to withhold affirming that her concept is satisfied by the salient features of the case.[12]

As for the promised distinction, in the following sections I will reserve "concept" for a particular individual's mental representation of the property that is picked out by her concept. Using "concept" in this ways allows us to make explicit the distinction between someone's concept of knowledge, for instance, the prevailing conception of knowledge, and/or the correct concept of knowledge. Given this, it is of course clear that someone's concept of knowledge can fail to coincide with the prevailing conception of knowledge, if the intensional content of the individual's concept of knowledge does not map onto the intensional content of the prevailing conception, and where the prevailing conception is (if there is one) the statistically normal way of representing properties that fall under the associated concept(s). An example of this is, arguably, the concept of knowledge possessed by many reliabalists. Although this is a claim that could be confirmed by experimental test (see section 1.4 below), plausibly the prevailing conception of knowledge amongst non-philosophers comports best with internalist analyses of knowledge. But a reliabalist's concept of knowledge will usually not comport with such analyses. So, since someone's concept may disagree with an associated prevailing conception, someone may find that certain properties satisfy her concept, but it may be that these same properties do not fall under the associated prevailing conception. Since this fact will have consequences for the epistemology of intuitions, it makes sense to set up our terminology for talking about intuitions and concepts in a way that does not

[12]It is of course important to keep in mind that satisfying the intuition's concept of, e.g., a plant is not the same as satisfying the correct of concept of a plant. Consider, for instance, intuitions about fungi or coral reefs.

suppress this distinction.

Turning now to the relationship between a concept and the propositional content of an intuition, I want to begin examining this issue with the following observation. We can see that the salient features of any sufficiently detailed case will satisfy more than one concept, and that the salient features of any one case will typically either satisfy or fail to satisfy an even larger number of concepts. Since intuitions involve both cases in which the implicated concept is satisfied and cases in which the implicated concept is not satisfied by the salient features of the case, this means that the overall details of the case will typically be insufficient to determine the identity of the implicated concept. Thus, it looks as though *which* of the intuitor's concepts gets to be the implicated concept must ultimately a pragmatic matter, determined by the way in which a particular intuition is being used in argumentation, for example.

Let me illustrate this point with two examples. First, with apologies to Gettier, consider the following case:

> Smith and Jones have applied for the same job. As both wait to be interviewed outside the office of the president of the company's office, neither Smith nor Jones has any evidence that he will be the one to get the job, and not the other man. Still, on the basis of nothing more than a guess, Smith forms the following belief: he will be the one to get the job. But as a matter of fact, Smith and Jones are the only two candidates being considered for the job and Jones is patently unqualified, so Smith's belief is true.

Clearly a number of concepts are satisfied by this case (guessing, believing,

truth, etc.), and even more are not. Some of the philosophically interesting concepts that are not satisfied, though, are the concepts of justified belief and knowledge. Indeed, this case might well be used to elicit either of the following intuitions:

> Smith does not know that he will get the job.
> Smith's belief is not justified.

But whether or not either of these intuitions are elicited will surely depend on how the case is framed by the argument in which the appeal to intuition is embedded. We can imagine that the first intuition would be elicited if the case were to occur as the crucial part of a paper mounting a refutation of some philosopher's theory that guessing to the truth provides knowledge, for instance.

Here is the second example.

> Smith has applied for a job and, after he is interviewed by the president of the company, the president tells Smith that he has the job. She then asks Smith to immediately sign a job contract, which he does. Smith is then told to report for work early next week.

Once again, depending on how the case is being used within an argument, it could be used to elicit this intuition,

> Smith knows that he has the job.

But were the case being used in service of a different argument - suppose it occurs as part of a paper discussing the legality of asking someone to sign a

contract before they have had the opportunity to consider the contract at length - a different intuition may be elicited - say, that the president did something improper. The concept implicated in this intuition has shifted; it seems to be the concept of legal propriety now.

We can see that in both of these examples it is the use to which the intuition is being put in argument that determines which of the intuitor's concepts 'counts' as the implicated concept. So, more generally, the implicated concept will usually be determined by the pragmatic context in which an appeal to an intuition occurs.

But these examples also show that we need to introduce some more terminology into our discussion of the structure of intuitions. We must consider the *operative presuppositions* that are made about the case, for as we will soon see, these partially determine what the *salient features* of the case will be.

Here's the idea. Consider the case used in our very last example. If an intuitor knows that the case occurs as part of an epistemological argument, she may presume that the purpose of the case is to elicit an intuition that reveals something about knowledge. Because of this, she might further presume that when Smith signs the job contract, he comes to believe that he has a job. However, if instead the intuitor knows that the case occurs as part of an argument in legal theory, she may instead presume that the purpose of the case is to elicit an intuition that reveals something about legal propriety. In this scenario, she might never think of whether or not Smith forms the belief that he has a job. Indeed, because the salient features of the case need not always consist in explicitly mentioned properties of the case, being fixed instead by the intuitor's operative presuppositions about the case combined with the explicitly mentioned properties

of the case, it is important for the person constructing the case to pay attention to what sorts of presumptions about the case will be made by potential intuitors. Usually, this will require the person constructing the case to adopt some presuppositions of her own about the presuppositions that will be made about the case by the intuitor. So, the point here is that both the person presenting the case to an intuitor and the intuitor herself will typically have beliefs that 'fill in' further details of the case, beyond what properties are explicitly described. These beliefs are what I have been calling *operative presuppositions*; and, importantly, which properties of the case are taken by some intuitor to be the *salient features* of the case can be partly determined by the operative presuppositions that the intuitor makes about the case.[13]

We are now in a position to be able to say more precisely what the relationship is between the implicated concept and the propositional content of an intuition. For by now it should seem clear that to have an intuition is nothing more than for the intuitor to apply one of her concepts, namely the implicated concept, to the salient features of a case – where which of the intuitor's concepts counts as the implicated concept is determined by the pragmatic context in which the appeal to the intuitor's intuition occurs, and where the salient features of the case are determined by the explicitly mentioned properties of the case plus the operative presuppositions that the intuitor is making about the case, and where the propositional content of the intuition is determined by whether or not the salient features satisfy the intensional content of the implicated concept. So, the basic idea is that the intuitor of the intuition that Smith does not know that he will get the job gets that intuition because the salient fea-

[13]The operative presuppositions that are involved in developing an intuition seem to be a special case of the phenomena of conversational implicature. And, just as for more ordinary examples of this phenomenon, it may be very hard to explicitly list all of the presuppositions that are, well, operative in the manifestation of any particular intuition.

tures of the case do not satisfy her concept of knowledge, given the intuitor's operative presuppositions about the case; the intuitor of the intuition that the patient does not have a moral duty to remain hooked up to the famous violinist gets that intuition because the salient features of the case do not satisfy her concept of moral duty, given the intuitor's operative presuppositions about the case; and the intuitor of the other intuition that Smith does know that he will get the job gets that intuition because the salient features of the case do satisfy the intuitor's concept of knowledge, given the intuitor's operative presuppositions about the case.

At this point, I want to address the following issue. I have claimed that an intuition amounts to an intuitor applying one of her concepts to a case, and that specifically, the intuitive judgment is produced by whether or not the salient features satisfy the intensional content of the implicated concept. But the satisfaction or not of the intensional content of some concept is not the only way that someone can apply a particular concept. For example, someone might use her implicit beliefs about the extension of a concept in order to mediate the application of that concept, where these beliefs do not constitute the intensional content of the implicate concept. Likewise, some concepts made be applied more or less automatically – as the result of, say, a course of classical conditioning. So, why don't these other kinds of conceptual application count as intuitive judgments? The answer, quite simply, is that not every application of a concept – even when the application is mediated by implicit psychological processes – counts as the manifestation of an intuition in the relevant *philosophical* sense of "intuition".[14] At the risk of stipulating too much, what I am trying to do is (as I said in the introduction) to come up with a terminological framework that captures fairly

[14]Indeed, see Hogarth 2001 for an interesting exploration of a notion of intuition that understands the notion in a much broader sense than philosophers usually do.

accurately the standard philosophical notion of an intuition. And of course, the standard philosophical notion of an intuition is derived from the practice of conceptual analysis, which plainly consists of philosophers appealing their intuitions in an attempt to explicitly characterize or describe the intensional content (or the meaning, or the sense) of some concept or another. Thus, in the relevant philosophical sense, it is the application of a concept mediated by the (dis)satisfaction of its intensional content that properly counts as an intuition.

So, it therefore looks as though something like the following picture must approximate (at a very abstract level) the psychology of an intuition, where "+" and "→" stand in for whatever the real underlying cognitive processes are that operate on the mental representations to produce the propositional content of the intuition.

> Intuitor's apprehension of the salient features of the case, conditioned by the intuitor's operative presuppositions + Intuitor's representation of the implicated concept → Propositional content of the intuition

Now, I believe that this picture captures most of the judgments that are standardly called intuitions by philosophers. Thus, this is not intended to be a picture that captures all of the phenomena that are called intuitions by contemporary researchers; for example, since it is implausible that grammatical competence is represented conceptually, it is unlikely that this picture applies to linguistic intuitions.

1.4 Meaning-Directed and World-Directed uses of Intuitions

In this section, I want to draw attention to a distinction that has important methodological consequences for research that involves intuitions. This distinction is easiest to grasp if we first introduce one last piece of terminology into our examination of the structure of intuitions. So, let us say that a *presentation* of an intuition is composed of the following three things: the case that is used to elicit an intuition, the implicated concept, and the intuition that has been elicited by the case. To illustrate this, we can say that the case from our last example above (which described Smith receiving a job offer from the president of a company), plus the intuition that Smith knows he has a job, plus the intuitor's concept of knowledge together compose a *presentation* of an intuition.

What I want to suggest is that, if the preceding account of intuitions is right, then presentations of intuitions can be used as two logically distinct kinds of evidence. The idea here is that – for any presentation that implicates the intuitor's concept of knowledge – the presentation may only tell us something about the intuitor's concept of knowledge, and not knowledge itself, if we have reason to believe that the intuitor's concept of knowledge does not correspond sufficiently well to knowledge. Of course, the propositional content of the intuition in a presentation may be identical to the propositional content of an intuition in another presentation whether or not either presentation involves a concept of knowledge that corresponds closely enough to knowledge. Even people with fairly unreliable concepts of knowledge may have intuitions about Gettier cases where the propositional content of each intuition is, for instance, that Smith does not know that he has a job. But one of the reasons why it is important to recognize these two different 'evidential uses' of presentations of intuitions

is that this distinction helps make it clear that two presentations that have the same propositional content can (and perhaps often do) differ in their reliability.

Let me try to make the distinction that I have in mind more salient by using a less philosophically tendentious example. Suppose we are interested in the intuitions of mine that implicate my concept of a hadron. Myself, I know very little about high energy physics. About all I know about hadrons is that they are made up of quarks bound together by the strong force. So, if someone presented me with a series of cases designed to elicit intuitions that implicate my hadron concept, it may very well be that I get intuitions that have propositional content specified by these sentences: "That electron is not a hadron", "That neutron is a hadron", "That baryon is not a hadron". But there's no reason to suspect that any of my intuitions are true; and if any of them are, then that is only an accident. The point here is that intuitions of mine that implicate my hadron concept are best interpreted as revealing only something about the representational structure of my hadron concept, and not the nature of hadrons.

However, if we find ourselves examining intuitions that implicate the hadron concept of someone trained in high energy physics, then a completely different interpretation of these intuitions seems plausible. Because of the way that she acquired her concept (i.e. through her training), this person's hadron concept should correspond accurately with hadrons. So, instead of seeing these intuitions as only revealing something about the intuitor's concept of hadrons, it is more natural to interpret these intuitions as revealing something further – they can be treated as evidence about what sorts of things have the property of being a hadron.

Examples like these suggest that, as I say, there are two logically distinct uses

of presentations of intuitions as evidence. A presentation of an intuition can be used as a *meaning-directed probe* or it can be used as a *world-directed probe*. To use a presentation as a world-directed probe is to treat the intuition manifest in the presentation as something like a recognition. That is, to use a presentation as a world-directed probe amounts to interpreting the intuition manifest in the presentation that x is F as providing good evidence that it is true that x is F. However, when using a presentation as a meaning-directed probe, we are not required to consider the truth-value of the propositional content of the intuition manifest in the presentation. So, to return to the example we just used, the trained physicist's intuitions about hadrons could reasonably be used as both meaning-directed and world-directed probes. Even though we probably only care about the later use, presentations of her intuitions that implicate her hadron concept can legitimately be used as either meaning-directed probe (thereby telling us something about the representational structure of her hadron concept) or world-directed probes (thereby telling us something about what things have the property of being a hadron). However, it is only rational to treat presentations of my intuitions that implicate my hadron concept as meaning-direct probes.

When discussing intuitions deployed in philosophical contexts, philosophers quite often make something like the following claim (this from Alvin Goldman), "Intuitions are evidence for the content of an intuitor's concept, or conception, of the term in question."[15] We can also recall Kornblith's comment, that "appeals to intuition are designed to illuminate the contours of our concepts."[16] I suggest that assertions like these be interpreted as assertions about what sort of knowledge you can gain by using presentations of intuitions as

[15]Goldman 2001 p.477
[16]Kornblith 2006 p.11

meaning-directed probes. But that said, it is also plausible that most philosophers who use presentations as evidence are interested in using these presentations as world-directed probes.

Moreover, I think that it is important to make it clear that, while it is true that using presentations of intuitions as world-directed probes is one way of trying to limn the boundaries of philosophically-interesting categories, it is also evident that presentations of intuitions are neither the only nor the most reliable kind of evidence that can be used in service of this kind of inquiry. There are, for instance, obviously better methods to use when trying to figure out what sorts of things might be hadrons; methods that involve much more than just eliciting the hadron-concept-implicating intuitions of high energy physicists. Arguably, the same point holds for the methods appropriate for the study of some philosophically interesting subjects (like knowledge) as well.[17] A similar lesson applies to the methods appropriate for the study of the representational structure of people's concepts. While it is true that presentations of intuitions used as meaning-directed probes are a source of evidence about the intensional content of people's concepts, it is not obvious that using intuitions as meaning-directed probes is either the best or the only way to try to characterize the intensional content of someone's (or some groups of people's) concepts. Depending on what the correct psychology of concept possession is, there may be other methods that are more appropriate.

[17] C.f. Kornblith 2002

1.5 Reliable Intuitions

Enough has been said about the structure of intuitions to allow us to turn to an examination of the epistemology of intuitions – the outlines of which, I am sure readers will have noticed, have already been suggested in the previous sections.

We have now observed that presentations of intuitions can be used as two logically-distinct kinds of evidence: they can be used as either world-directed probes or meaning-directed probes. Our task here is to identify some of the conditions under which it is rational to use a presentation as either a meaning-directed probe or a world-directed probe. So, here are the conditions that I believe should govern the use of presentations as meaning-direct probes.

> It is rational to use a presentation of an intuition as a meaning-directed probe if the intuitor is able to make ordinary operative presuppositions about the case, *and* the intuition manifest in the presentation occurs in favourable circumstances.

By "ordinary operative presuppositions" I just mean the presuppositions that the intuitor makes when deploying the concept implicated in her intuition in everyday discourse. And for technical concepts that are not deployed in everyday discourse, then the ordinary presuppositions would be just those presuppositions that are typically made when deploying the concept in the discourse in which the concept is usually used.

It is easy to see why the first condition matters to the appropriateness of using a presentation of an intuition as (at least) a meaning-directed probe. It may be, for instance, that some case is too weird or too complex, so that it is hard

or impossible for the intuitor to interpret the case as she would interpret more normal cases, and thereby make the presuppositions that she would ordinarily make in the course of deploying the implicated concept. Consider, for instance, a presentation that involves implicating someone's concept of moral responsibility, where the case used to elicit the intuition describes a scenario in which events occur that obviously violate the actual laws of physics. It is not unreasonable to suspect, about such cases, that an intuitor would make operative presuppositions that are different than the operative presuppositions that she would make about more natural cases. So, intuitions about 'unnatural' cases should not be interpreted as revealing anything about the representational structure of the implicated concept as it would be used in actual circumstances by the intuitor, and for this reason it is therefore inappropriate to use presentations that include such cases as evidence.

As for the condition requiring that presentations of intuitions occur in favourable circumstances, the justification for this second condition is the observation that the circumstances in which an intuition is elicited may impact the reliability of the presentation, even if the intuitor is able to make ordinary operative presuppositions about the case (and even if the implicated concept is sufficiently accurate – see below). For example, presentations of someone's intuitions may not be reliable in circumstances that are fraught with heated feeling of outrage or indignation, or in circumstances that present an opportunity to significantly advance one's own interests at the cost of doing the right thing. Certain contexts may offer incentives for an intuitor to get the 'right' intuition, as opposed to the intuition that would actually result from applying the implicated concept to the salient features of the case. Finally, an intuitor may find herself appealing to her intuitions in a distracting environment. There is in fact

some experimental evidence that priming and framing techniques can manipulate the content of a person's intuitions.[18]

As for the other way of using the presentation of an intuition, it seems that in order for it to be rational to use a presentation as a world directed probe, one further condition must be satisfied, namely we must have a good reason to believe that the intuition at the heart of the presentation is very likely to be true. Of course, we will have this reason only if the concept implicated in the intuition is itself relevantly reliable. The upshot, then, is that

> It is rational to use a presentation of an intuition as a world-directed probe if the intuitor is able to make ordinary operative presuppositions about the case, *and* the intuition manifest in the presentation occurs in favourable circumstances, *and* we have reason to believe that the implicated concept is sufficiently accurate.

And of course an intuitor's concept of x will ordinarily be sufficiently accurate if it maps onto the prevailing conception of x, *and* where both the prevailing conception of x is embedded in a conceptual network that has achieved some inductive and/or explanatory success and this success is at least in part explained by the fact that the intensional content of the prevailing conception encodes properties that correspond well enough to the properties that are typically instantiated by xs.

Let me mention two points of clarification to end this section. First of all, I do not mean to suggest that it is impossible for some individual to possess a con-

[18]See Sinnott-Armstrong 2008 for a nice review of the evidence. He also seems to reach the conclusion that this evidence shows that intuitions are not reliable. However, I think that this evidence shows only that we should pay attention to the environment in which we elicit intuitions. Maybe armchairs are, after all, a pretty good place in which to elicit reliable intuitions.

cept that is more accurate than the associated prevailing conception. Situations of this kind may be extremely rare in some discourses (high energy physics, say), but occur frequently in other discourses (such as public political discourse about the moral appropriateness of using military force, and maybe also epistemology – see section 1.7 below). And second, this proposal is not meant to deny that some individual may possess concepts that are more 'sufficiently accurate' than another individual's concepts. Suppose that two people have concepts that correspond well enough, as it were, to the things that fall under the concepts. It is of course possible that the one of these two concepts corresponds more accurately than the other concept. In cases like these, it is appropriate to treat the more accurate concept as the concept that is sufficiently accurate, which in turn implies that presentations of intuitions that implicate the less accurate concept should not be used as world-directed probes.

1.6 Intuitions Implicating Technical Concepts and Expertise

For presentations that involve intuitions that implicate technical concepts, our conditions imply that the following principle should guide uses of presentations as evidence when they are used as world-directed probes: use as evidence only presentations containing the intuitions of experts, where the intuition in question implicates a concept that finds its home in the intuitor's field of expertise.

There are several reasons why this is a sensible principle to follow, the most obvious of which is that it will normally only be experts who have concepts that map onto the prevailing conceptions in the discipline in which the technical concept finds its home. Since technical concepts are, of course, just those con-

cepts that have been developed by a discipline in order to help a theory being pursued by the members of the discipline achieve inductive and/or explanatory success, it will be rare for technical concepts to have any currency outside of the discipline in which that have been established. For example, it will be very hard to find someone who is not an expert about high energy physics and who, despite this, has a hadron concept that agrees with the prevailing conception of hadrons in high energy physics. But furthermore: it will normally only be experts in the discipline in which the prevailing conception associated with the implicated concept finds its home that will be in a position to make the appropriate operative presuppositions. And finally, for quite a few technical concepts, it will only be experts who actually possess instances of the concept in question. While ordinary people may have, amongst others, fairly unrefined hadron concepts, it is not implausible to think, that, for example, no one but a trained philosopher has the concepts of a natural kind, of supervenience, or of t-schemas.

In light of these points, it is interesting to observe that some philosophical uses of presentations already conform to the principle about technical concepts that we have just adduced. Let me offer two examples of presentations that are used as world-directed probes, where the evidential force of the presentation depends on the high probability that a certain kind of expert will share the intuition. The two examples come from a debate in the philosophy of biology concerning the metaphysics of species; both examples consist of presentations that are used to confirm the hypothesis that there can be historically disconnected species.

In our first example, Kristin Guyot makes use of a presentation that is de-

signed to show that, if all living members of *Galeopsis tetrahit*, which is a herb that originally arose as a hybrid from *G. pubescens* and *G. speciosa*, were to die and then another hybridization event occurs between these last two species in the same environment, then the new plants would be *G. tetrahit* re-emerged. In this presentation, the case is the hypothetical scenario where all living *G. tetrahit* die and are almost instantaneously replaced with another hybrid in the same environment; the implicated concept is the concept of *G. tetrahit* qua biological species; and the intuition is, as Guyot says, that "*Galeopsis tetrahit* was resurrected."[19] It is clear that she means to use this presentation as a world-directed probe. But more importantly, it is also clear that she thinks that her own intuition is representative of the intuitions that people who possess a certain amount of training in evolutionary theory would have about the case. Her aim is obviously not to present an intuition that agrees with the pre-theoretical species intuitions of non-experts in evolutionary theory.

Second example. Philip Kitcher uses a presentation that involves a different hybrid species, *Cnemidophorus tesselatus* (a unisexual species of lizard), which arose from crosses of *C. tigris* and *C. septemvittatus*. Once again, we are asked to suppose that all *C. tesselatus* die off, and then another hybridization event occurs soon thereafter in the same environment. Kitcher's intuition is, basically, that the new hybrid population are *C. tesselatus*. And as before, the cogency of Kitcher's argument turns on whether his fellow experts in evolutionary theory share the same intuition. His argument would have no more and no less impact if it turns out that people who are non-experts in evolutionary biology have intuitions that agree or disagree with Kitcher's own intuition.[20] [21]

[19] Guyot 1986 p.114

[20] This example is from Kitcher 1984 p.314-315. I'm grateful to Richard Boyd for pointing me toward these two examples.

[21] Of course, it might be useful to subject the claim that both Kitcher's and Guyot's intuitions

1.7 Intuitions Implicating Non-Technical Concepts

So, I think that the basic argument for deferring to experts when using presentations involving intuitions that implicate technical concepts is straightforward. There are of course a host of complicating issues that a more thorough treatment of expert intuitions should deal with. For example, individual members in a community of experts may have different concepts of one and the same entity; for instance, a population biologist will usually have a different gene concept than a molecular chemist. But at this point, I think that is more important to address the role of intuitions in philosophical inquiry.

For, we now face the following question: do the intuitions normally used in philosophical theory always implicate technical concepts? do the intuitions normally used in philosophical theory always implicate technical concepts? At first blush, it may seem that the answer is "no". Notions like knowledge, justification, belief, moral responsibility, and causation (to list just a few) obviously have currency in discourses outside of academic philosophy. This may be taken to suggest that, while it might be appropriate to defer to the presentations of the intuitions of philosophers (qua experts) when we are considering technical notions like supervenience, it is not clear that we should do this for presentations of intuitions that implicate someone's concept of, say, knowledge. In fact, if philosophers should properly be engage in attempting to clarify ordinary or 'folk' concepts, then prima facie, it seems as though the presentations of intuitions that philosophers do treat as evidence ought to agree with 'folk' presentations of intuitions.

This conception of the proper activity of a philosopher is, I believe, the ul-

do agree with to intuitions of other experts in evolutionary biology to experimental test.

timate source of many experimental philosophers' scepticism about allowing intuitive presentations that have not been shown to agree with related 'folk' intuitive presentations to occupy an evidential role in philosophical inquiry. The worry is summed up rather clearly by the following passage, which is taken from a recent article written by two of the more visible proponents of experimental philosophy, Joshua Alexander and Jonathan Weinberg. They write that "Philosophical practice is not concerned with understanding the nature of knowledge (or belief, freedom, moral responsibility, etc.) in some technical sense, but of knowledge as the concept is ordinarily understood outside of strictly philosophical discourse and practice. If it were concerned only with the technical sense of the concept, it would be divorced from the concerns that led us to philosophical investigation of the concept in the first place and its verdicts would have little bearing on those initial concerns. As such, large and central swaths of philosophical practice must be concerned with the ordinary concepts."[22] From this, they extract the conclusion (though not in these words) that presentations that are used as evidence in philosophy must include intuitions that are consistent with the intuitions of non-philosophers, especially if these presentations are being used as world-directed probes. The upshot is clear: experimental evidence showing that 'folk' intuitions routinely disagree with the intuitions that philosophers have treated as evidence in the past is a compelling reason to be sceptical of the philosophical intuitions, and a fortiori, sceptical of any philosophical theories that used these intuitions as evidence.

However, I think that this line of reasoning should be resisted. Again, it is obvious why philosophers should not care about the ordinary notions of a t-schema or the supervenience relation, since there isn't one available to study.

[22] Alexander and Weinberg 2007 p.57

But I want to argue that, even for ordinary notions like knowledge or justified belief, just in case a certain kind of theoretical accomplishment has been realized, it is appropriate to defer to the presentations of the intuitions of philosophers.

Here's the idea. Suppose that we want to understand the nature of knowledge, and we have some reason to believe that the notion of justified belief may figure into our best explanation of knowledge. Suppose too that, in ordinary discourse, the notion of justified belief has a range of applications. It could be that only one of these applications is best suited for integration into our explanation of knowledge. In these circumstances, it is possible that the ordinary notion of justified belief can undergo something like a process of denotational refinement, as one of the prior applications of the notion of justified belief coalesces, under the pressures of philosophical inquiry, into the new 'semi-technical' conception of epistemically justified belief.[23] Since this new semi-technical concept has more explanatory power in our account of knowledge than the ancestral ordinary notion, it makes sense to use it in our theory of knowledge. But this also means that the semi-technical concept is likely to be more accurate than the ancestral notion (which itself might be fairly accurate). Importantly, this example demonstrates that, unlike technical concepts, the most accurate semi-technical concepts will often *disagree* with the prevailing conceptions from which they have been derived. Still, this is clearly not a reason to suspect that the semi-technical concept is unreliable. So, my point here is this: for presentations used as world-directed probes that involve intuitions that implicate semi-technical concepts, it does not matter, evidentially speaking, whether the intuition agrees with presentations of 'folk' intuitions that implicate concepts which agree with

[23] A conception which itself may undergo further refinement, perhaps turning it into a 'wholly technical' conception.

the prevailing conception from which the semi-technical concept was derived. So long as the semi-technical concept of, say, epistemic justification carries some weight in an explanatorily successful theory of knowledge, it matters little if presentations of intuitions in which this concept is implicated agree with any related 'folk' intuitive presentations.

Of course this argument requires that I reject the conception of philosophical inquiry represented by Alexandar and Weinberg's comments above. But this is a bullet that I am prepared to bite: I think philosophical practice is better understood as a (not always successful) attempt to come up with (at least approximately) true descriptions of various abstract features of the world. And in their attempts to do this, philosophers both invent new technical concepts and refine existing non-technical concepts into semi-technical concepts. So, perhaps there is a sense in which philosophers often enough try to 'clarify' any number of folk concepts. But this may usually amount to an attempt to clean the folk concept up so that it can fulfill a meaningful role in a theory that has more epistemic virtues than some antecedent folk theory.

So, if this line of reasoning is right, then it is unfortunate that in recent years more than a few (though not all) experimental philosophers have put a lot of work into demonstrating that about a very wide range of subjects, non-philosophers consistently have intuitions that do not agree with the intuitions of many philosophers, and that the intuitions of non-philosophers can sometimes vary in accordance with such factors as the ethnic background of the intuitor or the intuitor's socio-economic status.[24] As previously mentioned, some of these experimental philosophers have used these data to argue that philoso-

[24] Alexander and Weinberg 2007 provide a nice introduction to the literature. And since Weinberg, Stich, and Nichols 2001 serves as the paradigm or exemplar for the research programme, this paper is also a helpful resource.

phers should be sceptical of philosophical theories that have been based more or less upon presentations of intuitions that have been treated as world-directed probes, and which implicated concepts that have currency in folk discourse – concepts such as the concepts of reference, of causation, and of knowledge.[25] But since in many of the relevant philosophical cases the implicated concept could be a semi-technical concept, in order to properly assess whether or not the implicated concept is 'sufficiently accurate' – and thereby determine whether or not it is appropriate to use the presentation in which the intuition is embedded as a world-directed probe – requires asking whether the implicated concept carries some weight in an explanatorily successful theory. Of course, there is no guarantee that all of the presentations of intuitions that both disagree with 'folk' intuitions and have been treated by philosophers as a world-directed source of evidence implicate concepts meeting this condition. So there is no guarantee that, for other reasons, the scepticism about philosophical theories based on philosophical intuitions recommended by some experimental philosophers is not appropriate. Speaking for myself, I am optimistic that analytic epistemology has been more or less successful in the relevant sense. But whether or not my optimism is justified, my point here is just that it is not an effective criticism of the practice of relying on presentations of intuitions as world-directed probes in philosophy to point out that the presentations that philosophers do use as world-directed probes may not agree with the intuitive presentations of non-philosophers, or that the intuitions of non-philosophers vary according to, inter alia, the intuitor's ethnic background.[26]

[25] See, e.g., Weinberg, Stich, and Nichols 2001

[26] Indeed, even if, say, the epistemological intuitions of philosophers are also shown to sometimes vary according to ethnic background, for instance, then this would not be be reason to be sceptical of presentations of these intuitions. Again, the crucial test would be to determine which, if any, of the epistemological concepts implicated in the relevant intuitions makes the most fruitful contribution to our epistemological theories.

But things might actually be worse for the efforts of experimental philosophers. We can now see why presentations involving the intuitions of non-philosophers should not be treated as world-directed probes, at least when the intuitions implicate concepts that may have either semi-technical or technical uses in some successful philosophical theory. But maybe all this shows is that the experimental philosophy literature can be read as providing data derived from presentations of intuitions that should be construed as meaning-directed probes. That is, perhaps the work of experimental philosophers tells us something about the various ways in which groups of non-philosophers, organized according to various social parameters, conceive of various subjects that are of interest to philosophers. On this reading, experimental philosophy is a nascent sub-field of social psychology that studies how ordinary people conceive of philosophically interesting subjects.

1.8 Conclusion

So, we can see that the presentations of the intuitions of philosophers can be legitimately used as world-directed probes, even if they are not normally consistent with presentations of intuitions that implicate the concepts of non-philosophers.

But at the same time, I want to stress that this conclusion does not imply that it is *always* appropriate to treat presentations of philosophical intuitions as world-directed probes, even when the intuitions in question are related to a subject about which philosophers are ordinarily recognized as possessing expertise. Perceptions of expertise can be divorced from actual theoretical success,

after all. Consider, for instance, the epistemic status of socially recognized experts on issues closely related to estimates of human potential in the latter half of the 19^{th} century. Because most of these individuals held sexist and/or racist views,[27] they had only a very limited understanding of their subject, and so were improperly treated as experts, at least with respect to issues involving to human potential. The suggestion here is that it is possible that something like this state of affairs may exist for academic philosophy. It may turn out that the social epistemology of philosophy reveals that some of the typical philosopher's concepts are not sufficiently accurate, despite that fact that philosophers may be commonly recognized as possessing expertise about issues closely related to this collection of concepts.

Here's a quick (and to my mind, not implausible) example of how this might occur. Most philosophers in North America work in universities where roughly a third of the way through their career, they are offered a contract that guarantees them permanent employment. If a tenure contract is the only kind of contract that philosophers are normally exposed to, and philosophers are usually not in routine contact with people (such as lawyers and union organizers) who deal with a wider variety of contracts, then it may well be that the typical philosopher's concept of a fair contract is not sufficiently accurate, in virtue of the fact that a tenure contract is radically different than most other kinds of employment contracts. Still, philosophers are commonly recognized as experts on issues pertaining to social justice, and because of this, presentations of philosophical intuitions about fair contracts may be accorded more epistemic standing than they in fact deserve, perhaps resulting in these presentations being erroneously used as world-directed probes.

[27] C.f. Gould 1996

Of course, I do not deny that philosophical intuitions about fair contracts can be helpful if we want to figure out how philosophers conceive of fair contracts. So long as appropriate experimental controls are used, there is no reason why we should not use presentations of such intuitions as meaning-directed probes. But it is probably more interesting – and it is clearly more important – to focus on trying to figure out the nature of a fair contract instead, and we can now see why it is possible that philosophical intuitions, or, the intuitions of philosophers, may be of little use for this kind of inquiry.

BIBLIOGRAPHY

[1] Alexander, J. and Weinberg, J. (2007) Analytic epistemology and experimental philosophy. *Philosophy Compass*, 2 (1), 56-80

[2] Audi, R. (2004) *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton University Press.

[3] Bealer, G. (1996) A priori knowledge and the scope of philosophy. *Philosophical Studies* 81, 121-42

[4] Bealer, G. (1998) Intuition and the autonomy of philosophy. In Depaul, M. and Ramsey, W. (eds.) *Rethinking Intuition: The Psychology of Intuition and Its Role In Philosophical Inquiry*. Rowman & Littlefield Publishers, Inc.

[5] Block, N. and Stalnaker, R. (1999) Conceptual analysis, dualism, and the explanatory gap. *The Philosophical Review*, 108(1), 1-46

[6] Fodor, J. (2008) *LOT 2: The Language of Thought Revisited*. Oxford University Press.

[7] Gettier, E. (1963) Is justified true belief knowledge? *Analysis*, 23, 121-23

[8] Goldman, A. (2001) Replies to contributors. *Philosophical Topcs*, 29, 461-508

[9] Goldman, A. and Pust, J. (2002) Philosophical Theory and Intuitional Evidence. In Goldman, A. *Pathways to knowledge - Private and public*. Oxford University Press.

[10] Gould, S.J. (1996) *The mismeasure of man*. W.W. Norton & Company.

[11] Guyot, K. (1986) Specious individual. *Philosophica*, 31(1), 101-126

[12] Harman, G. (1977) *The Nature of Morality*. Oxford University Press.

[13] Harman, G. (2008) Using a linguistic analogy to study morality. In Sinnott-Armston, W. (ed.) *Moral psychology, Volume 1, The evolution of Morality*. M.I.T. Press.

[14] Horgath, R.M. (2001) *Educating Intuition*. University of Chicago Press.

[15] Huemer, M. (2005) *Ethical intuitionism*. Palgrave MacMillian.

[16] Kitcher, P. (1984) Species. *Philosophy of Science*, 51(2), 308-333

[17] Kornblith, H. (2006) Appeals to intuition and the ambitions of epistemology. In Hetherington, S. (ed.) *Epistemology futures*. Oxford University Press.

[18] Kornblith, H. (2002) *Knowledge and its Place in Nature*. Oxford University Press.

[19] Machery, E. (2009) *Doing without Concepts*. Oxford University Press.

[20] Putnam, H. (1975) The meaning of "Meaning". In Putnam, H. *Mind, language, and reality - Philosophical papers volume 2*. Cambridge University Press.

[21] Sinnott-Armstrong, W. (2008) Framing moral intuitions. In Sinnott-Armstrong, W. (ed.) *Moral psychology, Volume 2, The cognitive science of morality*. M.I.T. Press.

[22] Thompson, J.J. (1971) A Defence of Abortions. *Philosophy and Public Affairs*, 1(1), 47-66

[23] Weinberg, J., Stich, S., and Nichols, S. (2001) Normativity and epistemic intuitions. *Philosophical Topics*, 29, 429-460

CHAPTER 2

AGAINST THE LINGUISTIC ANALOGY

2.1 Introduction

In recent discussions involving both philosophers and psychologists, the phrase "the linguistic analogy" has been used to denote the view that progress in moral psychology can be made by adopting a basically Chomskyan approach to the field.[1] Somewhat more precisely, the linguistic analogy can be conceived as a methodological wager: it holds that, in view of the putative success of Noam Chomsky's nativist linguistic psychology, it is reasonable to expect that the conceptual framework of this theory can be used to fruitfully illuminate the psychology of other kinds of knowledge, such as moral knowledge. So, if the linguistic analogy is apt, adopting a basically Chomskyan moral psychology may provide the intellectual conditions necessary for moral psychology to finally "establish itself as a serious science."[2]

Thus, the initial effect of raising the analogy is to focus attention on the hypothesis that some of the psychological mechanisms that according to the Chomskyan view implement some components of our linguistic knowledge may be importantly similar to the psychological mechanisms that implement some of our moral knowledge.[3] The idea is that some of our moral knowledge and our linguistic knowledge are both implemented in approximately the same

[1] Both Mikhail 2007 and Hauser 2006 provide useful introductions to the approach.

[2] Mikhail 2008 p.354

[3] Here and hereafter, I am, following proponents of the linguistic analogy, using "knowledge" to refer to information that is available to be used by at least some of the mind's cognitive faculties. This use of "knowledge" allows it to denote more than only a particular species of belief.

type of psychological mechanisms. Specifically, the view in the moral case is that there is a developmentally-endogenous domain-specific morality acquisition device (or, morality module) that encodes rules that map representations of intentional actions, for example, onto moral values such as "permissible" and "impermissible". At least some of the rules which specify these mappings are the content of a universal moral grammar, and while the rules that are the universal moral grammar are innately specified, they do admit of parametric variation in order to account for cultural differences of moral view. Furthermore, this morality module is both upstream and encapsulated from the more domain-general cognitive mechanisms that implement conscious reasoning. It is because of this that the morality module's content primarily receives its expression in our moral intuitions. But this encapsulation also implies that it is possible for the moral principles encoded in the morality module to fail to correspond to any of the beliefs encoded in the more general psychological structures accessible to consciousness. So, a further consequence of this view of moral cognition is that someone may be able to manifest systematic moral intuitions even if they lack a worked-out, consciously accessible moral theory.

What is perhaps most striking about the linguistic analogy, however, is that it is an intended consequence of the view that neither explicit moral reasoning nor moral affect are ordinarily able to determine moral judgment.[4] Marc Hauser, one of the most visible defenders of the linguistic analogy, and his team of fellow researchers believe that one of the most central questions in moral psychology is, What causes moral judgments? And on Hauser and his team's understanding of the present debate, the field is dominated by three different models that each provide a different answer to this question.[5] According to

[4]See Hauser 2006 p.45, Hauser et al. 2008a p.117. Also, see section 2.5 below.
[5]See Hauser et al. 2008a p.113-121

their view, the Kantian model holds that moral judgments are caused by explicit moral reasoning, the Humean model holds that moral judgments are caused by moral affect, and the Hybrid model holds that moral judgments are caused by explicit moral reasoning and moral affect working in concert. As an alternative to these views, Hauser and his team introduce what they call the Rawlsian model, which holds that unconscious "action analysis" (i.e. the computations driven by the moral grammar encoded in the morality module) causes moral judgments, and explicit moral reasoning and moral emotion only kick-in after the moral judgment has been made.

I will have more to say about Hauser and his team have set up their vision of the currently credible theories of moral judgment below in section 2.5 – for now it is more important to recognize that, according to this vision of moral judgments, because moral cognition is very heavily regulated by the information encoded in the morality module, affect and the capacities underwriting explicit reasoning are too far downstream from the epistemic centre of moral cognition to have any effect on moral judgment.[6]

I believe, therefore, that it is accurate to portray the three 'core ideas' of the linguistic analogy as these:

[6]One of the experimental strategies that has been used to defend the linguistic analogy involves showing that people's intuitive moral judgments are extremely sensitive to the moral properties of ethical dilemmas, and that people are not very good at justifying these seemingly quite reliable intuitive judgments (sections 2.3 and 2.4). As I have just noted, defenders of the linguistic analogy conceive of the debate that they are initiating with more traditional views in moral psychology as a debate about whether moral judgment is usually determined by explicit cognitive resources, such as either (or both) moral affect or explicit moral reasoning. I believe that they want their experimental data to be interpreted as showing that, when moral agents are able to reliably detect the moral properties of the world, they are able to do this because of their tacit – not explicit – moral knowledge. The intended conclusion, it seems, is that reference to explicit processes of moral cognition is unimportant in any adequate scientific moral psychology, insofar as it is a goal of such a theory to explain how we are able to navigate the moral world.

(a) moral cognition is governed by knowledge of a coherent and consistent system of moral principles covering permissions and prohibitions, (b) this knowledge is tacitly or unconsciously represented in humans,[7] and (c) this knowledge is developmentally-endogenous.

It easy to see that these propositions have a number of important implications for more traditional positions in moral psychology, ethical methodology, and ethical theory.

First of all, the received view amongst most moral philosophers is that an ordinary moral agent's moral knowledge is neither coherent nor consistent. Some people will have moral views that are more coherent and consistent than other people's moral views, but almost no one (who is not either a professional philosopher or a lawyer, anyway) will be taken to have a moral view that almost exactly satisfies either of these two epistemic ideals. And while it is true that proponents of the linguistic analogy can agree with the received view so far as explicit or consciously accessible moral knowledge is concerned, their position implies that an individual's tacit or unconscious moral knowledge, surprisingly, forms both a consistent and a coherent moral theory.

Similarly, many philosophers will not believe that an ordinary agent's moral knowledge encodes mainly just a system of rules of prohibition and permission, or, if it does encode a system of rules, that such a system has an internal structure that approximates the structure of a set of mathematical axioms. Rather, it will strike most philosophers as plausible that an ordinary agent's moral knowledge is realized in what is, by comparison, a much more motley and heterogeneous fashion. Here, the received view will be that this knowledge consists of, inter

[7] Except for explicit psychological and philosophical theories of it, of course.

alia, some explicit moral beliefs and inferential tendencies, some combination of a collection of moral sentiments and passions, and a certain amount of unconscious moral knowledge too. The content of this moral knowledge will, of course, include both beliefs about prohibited and permitted acts and knowledge of some moral principles, but it will also include knowledge of things that are valued or considered to be moral goods, beliefs about the reliability or sensitivity of other moral agent's ethical thought, knowledge about what sorts of practical plans tend to increase or decrease things of value, and so on. And finally, many philosophers will believe both that this body of moral knowledge can potentially undergo very radical changes throughout the course of an individual's life, and that a person's moral knowledge is not in any important sense innate.

The core ideas of the linguistic analogy also have some surprising methodological implications for theoretical ethical inquiry. For example, the core ideas imply that following or pursuing reflective equilibrium is not a reliable strategy to follow in order to improve our ethical theories, where the intended outcome is gradual increases in the coherency or consistency of some body of ethical knowledge. Here, an analogy between theoretical linguistics and ethical theory per the linguistic analogy is helpful. By studying the grammar or syntax of a natural language, a linguist will sooner or later acquire a large body of information that is, basically, an explicit representation of the largely tacitly encoded linguistic knowledge of some community of speakers. But this theoretical knowledge cannot (in any non-trivial sense) be used to figure out how language should be spoken in the community. It is simply a description of the tacit knowledge that explains why, in some linguistic community, language is used that way that it is. The core ideas of the linguistic analogy seem to suggest

that ethical theory is in the same position as theories of grammar and/or syntax for natural languages: the ethical theorist's job is simply to provide an explicit description of the moral knowledge that ordinary moral agents in some community have, where the only data to be explained are patterns in people intuitive moral judgments. Trying to improve people's moral views by, say, engaging with them in explicit moral deliberation is, at best, a harmless activity.

The primary reason for this is that both the practical and the theoretical utility of explicit moral reasoning are undermined by the linguistic analogy. Its core ideas imply that it is a mistake to believe – as many philosophers and psychologists do – both that the ability of a person to improve their moral view is enhanced by moral deliberation, and that doing moral philosophy can be expected to contribute to moral understanding.

Take, for instance, reflecting on the moral questions, How can I be a good person? If the moral psychology of the linguistic analogy is correct, then engaging in explicit moral deliberation – either personally or interpersonally – about such a questions should not be central to moral practice and inquiry. Even if someone were to arrive at a compelling answer to the question of how to be a good person, this would not make a difference to what moral judgments she forms because the answer is encoded too far downstream from the source of moral judgment. The same holds for moral affect. It is easy to see that, for instance, trying to follow the virtue theorist's advice to habituate into one's character certain sentimental tendencies would be a mistake if the linguistic analogy is true, since moral affect kicks-in too far downstream from the real centre of moral cognition for moral affect to make any significant difference to moral judgment. And finally, it seems pointless, if the linguistic analogy is true, to try

to develop any other ethical theory than whatever ethical theory will be able to capture patterns in people's intuitive judgments. Working out some other ethical theory would be something like developing a language that no one can possibly speak, since the linguistic analogy seems to imply both that the one true ethical theory is a description of the moral principles regulating our intuitive moral judgments, and that no other ethical theory can ever hope to win over our moral judgments.

Indeed, along these lines, people who hold broadly Kantian, consequentialist, or Aristotelean moral theories normally take these theories to be compatible with the various aspects of moral psychology that have most fascinated both philosophers and psychologists over the years. Perhaps the single true ethical theory is a theory of rules for action. If so, it need not be a part of this theory that anyone knows what these rules are, or for that matter, that humans will have a particularly easy time following these rules, or that, once we do learn what some of the rules are, people will no longer have moral sentiments that are sometimes in tension with the rules. But as we can see, the core ideas of the linguistic analogy suggest both that some version of a deontological ethical theory is true, and that, surprisingly, people already have knowledge of it.

So, we can see that, because of the psychological, methodological, and theoretical implications of the core idea of the linguistic analogy, it would be an extremely important result if there were shown to be credible scientific evidence supporting the theory. The success of the linguistic analogy would overturn many longstanding ethical theories and assumptions. This is one of the reasons why it is important to investigate the case that has been made for the linguistic analogy.

Another reason, though, derives from the linguistic analogy's popularity. It is becoming an extremely visible theory. It has been introduced to the public by way of a number of articles and favourable reviews published in a variety of leading scientific, philosophical, and lay imprints, including *Science*, *Nature*, *The New York Times*, and *The Wall Street Journal*, and even a cover story for *Time* magazine.[8] The linguistic analogy is seen by some as a prestigious new scientific theory that may be able to finally settle questions that have puzzled philosophers for centuries. So, it is important to investigate whether the general estimation of the scientific importance of the linguistic analogy is deserved.

It is this paper's thesis that there is no evidence for any of the core ideas of the linguistic analogy.[9] Concerning the evidence that knowledge of a moral grammar is innate, no argument has been provided that either would rebut obviously plausible non-nativist alternatives, or that meets currently accepted scientific standards for such evidence (section 2.2). Regarding evidence that moral cognition in ordinary moral agents is rule-governed, no evidence has been provided that would refute the view that in ordinary moral agents moral knowledge consists of a collection of different types of information that exhibits widely varying degrees of consistency and coherency (sections 2.3 and 2.4). And regarding the view that people have unconscious knowledge of a number of moral principles, the experiments designed to show this do not manage to show either this or even the obvious – namely, that moral judgments are sometimes guided by unconscious knowledge (sections 2.3 and 2.4). Of course, someone who disagrees with the core ideas of the linguistic analogy is free to hold that some moral

[8]For favourable reviews in both scientific and popular avenues, see Bloom & Jarudi 2006, Waldman 2006, Holtz 2007, Kluger 2007, Pinker 2008. References to the best defenses of the linguistic analogy in both scientific and philosophical contexts are found throughout this paper.

[9]Because it has already been rebutted (see Prinz 2008), I do not discuss the neurological evidence that has been published in support of the linguistic analogy (see Hauser 2008a).

judgments may be mediated by explicit moral knowledge, and others by unconscious moral knowledge. But unless the moral psychology of the linguistic analogy is thought (implausibly) to be the only moral psychology that allows for the possibility of unconscious moral knowledge, holding the view that people have some unconscious moral knowledge confers absolutely no support for the linguistic analogy (section 2.5).

2.2 A Poverty of the Moral Stimulus Argument

We will begin with an examination of an argument for the linguistic analogy that has been termed by its author "the argument from the poverty of the moral stimulus". This argument is designed to show that humans come equipped with a stock of developmentally-endogenous moral knowledge, and it is supposed to parallel poverty of stimulus arguments as they are deployed in theoretical linguistics.[10]

Proponents of the linguistic analogy usually give John Rawls credit for introducing into contemporary debates the idea that moral psychology may be studied using ideas borrowed from Chomsky's nativist programme in linguistics. But it is John Mikhail who has done the most, in his various publications over the last decade or so, to both raise the academic profile of the linguistic analogy and work our various details of the view, such as providing an account of some of the content of what he calls the universal moral grammar.

[10]Poverty of the stimulus arguments have, basically, this form: we have some knowledge K, and in order to learn K from experience, we would have to be exposed to data D. But D is extremely rare or non-existent in the environment in which learning occurs, so our knowledge of K must be developmentally-endogenous.

It is interesting to note that in his initial publications defending the linguistic analogy, Mikhail was hesitant to claim that there was a poverty of the moral stimulus argument available in support of the view. He writes that we cannot formulate an argument of this kind until a reasonably complete description of the moral knowledge driving some of our moral judgments is developed. Applying Chomsky's distinction between an I-language (an individual's linguistic knowledge) and an E-language (the public language of a community of speakers) to moral knowledge, Mikhail says,

> we cannot even formulate [the question, How is I-morality acquired?] until a presumptively adequate answer to [the question, What constitutes I-morality?] is in hand. This is one reason I have said very little thus far about the argument from the poverty of the moral stimulus.[11]

He later writes that although it may be possible to articulate a poverty of stimulus argument for moral knowledge, "we refrain from drawing any firm conclusions about it here... [The assumption that there is an innate cognitive faculty that encodes a generative grammar for morality] is largely speculative and the issue requires more empirical investigation."[12] So, according to Mikhail, in order to formulate a coherent poverty of the moral stimulus argument, we first need a detailed description of at least some the moral knowledge that people ordinarily have. Let us call this the *knowledge requirement*, in order to be able to easily refer to it.

Now, in both of the monographs just quoted and also in subsequent work,

[11] Mikhail 2000 p.93
[12] Mikhail 2002 p.65

Mikhail has begun to work out a description of some of the moral knowledge that he believes guides some of our moral judgments. Mikhail thinks that he has made some progress describing the content of our I-morality, which he believes has the form of a generative grammar. His work focuses on the type of judgment that philosophers call moral intuition, and in particular his aim is to identify a collection of moral principles that are able to capture patterns of intuitions that people report when presented with variations on the standard run-away trolley cases.[13] It is on the basis of this work, as well as some general findings in developmental moral psychology, that he believes that the knowledge requirement has been satisfied.

Accordingly, he has recently presented a poverty of the moral stimulus argument. In an article entitled *The Poverty of the Moral Stimulus*,[14] Mikhail observes that social psychologists have shown, e.g., that "three- and four-year-old children use intent or purpose to distinguish two acts with the same results", and that they can distinguish between violations of moral conventions and violations of social conventions. He also reports that five and six year-olds are able to recognize that false factual beliefs sometimes excuse someone from moral responsibility, but false moral beliefs about the same circumstances do not; that five and six year-olds "calibrate the level of punishment they assign to harmful acts on the basis of mitigating factors"; that six and seven year-olds react "negatively when punishment is inflicted without affording the parties notice and the right to be heard"; and finally, that when presented with trolley problems,

[13]See, e.g., Mikhail 2000, 2002, 2007. A run-away trolley case is a hypothetical moral dilemma, in which the brakes of a trolley experience a mechanical failure, thereby endangering the lives of a group of people trapped on the rail line. Normally, the circumstances of the dilemma are such that a person to whom the dilemma is presented faces two morally difficult choices. For instance, they can opt to throw a switch diverting the trolley onto a secondary line on which only one person is trapped, or (in different versions of the scenario) they can choose to push a man in front of the trolley in an attempt to halt it and save the lives of the group of people.

[14]Mikhail 2008

"children as young as eight permit killing one to save five, but only if the chosen means is not wrong, the bad effects are not disproportionate to the good effects, and no better alternative is available".[15] He concludes that,

> In these cases[...] to explain the observable data we must attribute unconscious knowledge and complex mental operations to the child that go well beyond anything she has been taught. Indeed, as difficult to accept as it may seem, we must assume that children possess an elaborate system of natural jurisprudence and an ability to compute mental representations of human acts in legally cognizable terms[...] These concepts and the principles which underlie them are as far removed from experience as the hierarchical tree structures and recursive rules of linguistic grammar. It is implausible to think they are acquired by means of explicit verbal instruction or examples in the child's environment.[16]

So, these considerations are intended to show that children have some innate moral knowledge. But what is the evidence that this knowledge consists of a generative moral grammar? For this, Mikhail has an independent argument. He reasons that because we acquire our moral knowledge on the basis of a finite number of experiences, and because we can make a potentially infinite number of moral judgments "about the properties of various acts, institutions, and agents, including judgments in entirely new situations, ones dissimilar from the finite number of situations she has previously encountered",[17] the configuration of our moral knowledge is best conceived of as a kind of generative grammar.

[15] Mikhail 2008 p.354-355
[16] Mikhail 2008 p.355
[17] Mikhail 2000 p.54-55

So, we should believe that the developmentally-endogenous moral knowledge that Mikhail thinks we possess is structured like a generative grammar because this hypothesis offers the best solution to the problem of projection in the moral domain.[18]

Mikhail's view, then, is that there is "a 'Universal Moral Grammar' (UMG) analogous to the linguist's notion [*sic*] of a Universal Grammar (UG), that is, an innate function or morality acquisition device that maps the child's early experience into the system of principles that constitutes the mature state of her moral competence."[19] He goes on to express his belief that cultural variation in moral views may be profitably treated on the model of parametric variation, following in outline Chomsky's principles-and-parameters approach in the theory of natural language syntax.

As I said, my aim in this section is to demonstrate that this argument fails. I will do this by showing briefly that Mikhail has not actually offered a poverty of the stimulus argument that meets the evidential standards used for arguments of this type in debates in theoretical linguistics, and also by challenging his judgment that only a nativist moral psychology is able to provide a plausible explanation of how children are able to acquire the items of moral knowledge that he lists. (In the final section of this paper, I will offer some brief remarks addressing Mikhail's views about what follows from our ability to project in the moral domain.)

[18]It is wrong to conclude that Mikhail's own work on the moral intuitions of adults provides evidence that the moral knowledge he believes to be innate is a generative moral grammar. First of all, there are some serious methodological problem with Mikhail's work. But ignoring these problems, Mikhail has shown that some inter-defined principles are able to capture some patterns in people's moral intuitions, suggesting that at least some of the moral knowledge that guides people's moral intuitions looks vaguely like a generative grammar. He has not shown that these principles are also used by children.

[19]Mikhail 2008 p.355

First, a few words about poverty of stimulus arguments in linguistics. For although there is some debate about the precise details of poverty of stimulus arguments, there is general methodological agreement that these arguments depend upon two different sets of data.[20] The first is a description of the knowledge that is allegedly innate – a description that satisfies what we previously called the *knowledge requirement*. The second is evidence demonstrating that non-nativist learning routines could not be used to acquire the information that satisfies the knowledge requirement. Nativists have historically attempted to provide this kind evidence by arguing on purely theoretical grounds that certain kinds of linguistic data are extremely rare or even non-existent – for, the learning routines cannot function if they are not provided with enough incoming data. But the debate now seems to turn on, e.g., experiments designed to assess whether language learners can use information about the distribution of certain linguistic forms to learn a language,[21] and whether electronically analyzable samples of child-directed speech, ordinary context speech, and edited text contain or lack the relevant kinds of linguistic information.[22]

Let me illustrate this point with the following example. Geoffrey Pullum and Barbara Scholz have recently presented data derived from electronic corpus analyses that appear to show that certain syntactic constructions are not as rare as some advocates of linguistic nativism have claimed. For instance, in perhaps the most colourful example, Pullum and Scholz refute Chomsky's famous assertion that "a person might go through much or all of his life without ever having been exposed to the relevant evidence"[23] necessary to distinguish between two

[20] c.f. Ritter 2002

[21] See, e.g., Reali and Christiansen 2005

[22] Although he mounts a passionate defense of the position that humans do not have any innate linguistic knowledge, Sampson 2005 provides an accessible introduction to both sides of the debate.

[23] Chomsky 1980 p.40

hypotheses concerning the formation of polar interrogatives from certain types of declarative sentences; the evidence in this case are "auxiliary-initial [interrogative] sentences in which the clause-initial auxiliary is not the first auxiliary in the related declarative."[24] In addition to some other corpus data bearing on the rarity of such interrogatives, Pullum and Scholz found that "of the roughly 23,000 questions in [the 1987 *Wall Street Journal* corpus], one must look through only 15 before hitting an example of" a sentence that falsifies Chomsky's claim.

This example allows me to demonstrate the first problem with Mikhail's poverty of the stimulus argument. He seems to have misunderstood that poverty of stimulus arguments are, in their present form in theoretical linguistics, sophisticated empirical arguments. Few of the linguists and psychologists involved in this debate, whether they agree with Chomsky or not, are willing to follow Chomsky's tendency to simply announce from the armchair what linguistic information may or may not be available in the environment of language learners and users. Mikhail is of course correct that in order to be able to properly assess the question of whether or not some body of moral knowledge is innate, we must first do the empirical work required to produce a sufficiently rich description of this knowledge. But an explicit description of the knowledge that is allegedly innate is obviously not tantamount to a demonstration that the knowledge is innate. Yet, save for his judgment about the abstractness of the knowledge, this is all that Mikhail has produced so far. So, in order to actually give a poverty of the stimulus argument, Mikhail needs to go further – he needs to demonstrate using the best available empirical methods that the only scientifically plausible explanation of how moral agents acquire some of their moral knowledge is the nativist theory that he favours.[25] Otherwise, it is rad-

[24] Pullum and Scholz 2002 p.37

[25] The closest a supporter of the linguistic analogy has come to offering such data is, to my

ically premature to conclude that "we must attribute unconscious knowledge and complex mental operations to the child that go well beyond anything she has been taught".

This brings us to the second problem with his argument. For as Mikhail's argument stands now, it rests on, as I just noted, his own judgment that the moral knowledge he has described is too abstract to be learned. Because of this, Mikhail's judgment can be easily refuted by producing a plausible non-nativist explanation of how people are able to learn the items of moral knowledge that he lists. Let me stress here that the following argument is not intended to show that the non-nativist explanation is true, but rather to make the point that – as a matter of scientific confirmation – Mikhail is not entitled to his nativism until he can rule out all plausible non-nativist alternatives to his view.[26] His argument can be refuted, then, by listing some non-nativist alternatives.

So, consider the following 'naturally occurring' episodes of moral pedagogy, each of which may be the source of the items of moral (or quasi-moral) knowledge that Mikhail lists and believes to be innate. First of all, three- and four-year-old children may learn to use intent or purpose to distinguish two acts with the same results from experience with their own intentions. In the process of implementing some plan (drink milk), a child could cause the plan to fail intentionally at one time (by intentionally knocking the milk over), and at a later time see that the plan fails without her intending (the glass slips from the child's hand). In both of these cases, the act is the same (milk is spilt), but, from

knowledge, Dwyer 1999 & 2003. Joyce 2006 defends a nativist moral psychology, and although he indicates that some of his arguments may be construed as a poverty of stimulus argument for morality, he also denies "that morality *with a particular content* is innate," and accepts "that mechanisms of cultural transmission play an enormous and perhaps exhaustive role in determining the content of an individual's moral convictions." (Joyce *forthcoming*) p.1

[26]Of course, he also needs to rule out plausible nativist alternatives as well. Not every moral nativist believes that there is a universal moral grammar.

the child's perspective, only one of the acts was intended. Experiences like these could be the basis from which children learned to distinguish between intended and unintended acts. And a child might learn that there is a moral difference between intended and unintended acts by being punished for intended and wrong actions, but not being punished by accidental and wrong acts.

As for knowledge of the distinction between moral and social conventions, this can be acquired on the basis of the exercise of imaginative reflection: a child who is punished for doing something morally wrong but is merely informed or gently scolded for violating a social convention could come to recognize in rough the distinction between violations of moral norms and violations of social norms by associating the former, but not the later, with more painful punishments.[27] Children could also acquire knowledge of this distinction by attending to the frequency of punishment. It seems plausible that moral violations are punished far more often than violations of social norms, and it seems equally plausible that parents ordinary patrol their children's moral behaviour more intently than they patrol their non-moral behaviour. So we would need evidence to the contrary in order to be certain that these patterns of behaviour are not the source of children's knowledge of the distinction between moral and social conventions.

Moving along, more mature children may be able to recognize that false factual beliefs but not false moral beliefs sometimes excuse someone from moral responsibility by simply observing and learning from the practices of moral praise and blame that occur amongst their family, at their school, or in their broader social environment. By the age of six, children will have witnessed countless

[27] But see Shweder et al. 1987 for a view that, put roughly, defends scepticism about the reality of a universal distinction between social norms and moral norms.

examples of this kind of behaviour (especially if they have been exposed to television). For instance, perhaps on one occasion a parent withholds punishment for a moral violation because the parent learns of a gap in the child's factual (i.e. non-moral) understanding – and instead of punishing the children, the parent instead tries to correct the child's knowledge. Observation and learning may also be the point of departure for knowledge that punishment is commonly adjusted on the basis of mitigating factors. One again, children may simply be emulating their parents: perhaps six and seven year-olds have a rough concept of moral fairness that they have acquired as the result of being punished for reasons that they cannot discern, and it is this concept that they use to guide negative reactions to examples of punishment that occur without the punishers telling the punishees why they are being punished.

These examples show that it is actually relatively easy to come up with a plausible non-nativist explanation of how children can acquire the moral knowledge that Mikhail's believes must be innate. Mikhail has offered no evidence whatsoever that any of these non-nativist scenarios do not occur. Mikhail is therefore wrong to believe that a nativist moral psychology is the only plausible explanation of how children can acquire the items of moral knowledge that he mentions. As I say, Mikhail needs to be able to refute these (and perhaps quite a few other) plausible alternatives before he can claim that only the particular kind of nativist moral psychology that he favours can account for the social psychological data.

2.3 Moral Competence and Justification

In the previous section, I reported that Mikhail is confident that his research has shown that the knowledge that guides some people's moral intuitions has a structure that approximates the structure of a generative grammar. However, other proponents of the linguistic analogy have adopted a more cautious attitude towards this issue. These researchers have planned and conducted a study that is intended to confirm the hypothesis that a modularized and encapsulated psychological representation of some moral principles guides our intuitive moral judgments. The researchers beliefs that the results "thus far lead, we think, to the intriguing possibility that *some* forms of moral judgment are universal and mediated by unconscious and inaccessible [moral] principles."[28] This is, it is important to note, a weaker hypothesis than the view that these principles are also collectively organized in a structure something like that of a generative grammar. But all the same, I will use this section and the next to argue that the results of these experiments cannot be used to support this weaker hypothesis, and so a fortiori they cannot be used as evidence in support of the stronger hypothesis that a developmentally-endogenous, modularized, and encapsulated system of moral principles guides our intuitive moral judgments.

The study in question is called the Moral Sense Test. It is operated by Marc Hauser and a number of his graduate students, and it is designed both to find patterns in the judgments that philosophers call moral intuitions and to test whether people can justify their moral intuitions. Experimental participants in the Moral Sense Test visited a website, where they were presented with pairs of descriptions of moral dilemmas that are designed to 'target' knowledge of

[28] Hauser et al. 2008a p.134, emphasis theirs.

moral principles such as, e.g., "it is less permissible to cause harm as an intended means to an end than as a foreseen consequence of an end."[29] Participants were asked (though not in exactly this language) to use their moral intuitions to determine whether a choice derived from the details the various scenarios that they read was morally permissible or morally impermissible. Participants were then asked to justify their moral intuitions. Some of the data from the study is published in both Hauser et al. 2007 and 2008a. These articles have primarily explored three independent issues: first, the degree to which moral intuitions are shared across various different demographic indicators; second, the degree to which the patterns in the moral intuitions observed 'fit' the targeted moral principles; and third, the ability of experimental participants to offer justifications for their moral intuitions, especially when their intuitions seem to 'fit' one of the targeted moral principles. The guiding idea appears to have been that if people have intuitions that are consistent with the targeted moral principle, but they are unable to articulate this moral principle when asked to justify their intuitions, this shows that they have unconscious knowledge of the moral principle in question. As far as I can tell, Hauser et al. have not seriously considered the hypothesis that moral intuitions might not be guided by psychological representations of moral rules, and are instead guided by information that has a different logical form.

The scenarios are variations on the standard run-away trolley theme, and Hauser and his team have published analyses of people's moral intuitions elicited by four of these scenarios. In scenario 1, Denise faces a 5 versus 1 dilemma: he can flip a switch to spare the lives of five persons trapped on the tracks in front of an out-of-control train, but the switch will direct the train onto

[29] Hauser et al. 2007 p.15

a side track where one person is trapped and will be killed by the train. In scenario 2, Frank faces a similar dilemma, except that to spare the lives of the five people trapped on the track, his only option is to shove a "sufficiently heavy" weight in front of the train, and the only such weight available is a fat man standing on a footbridge over the main tracks on which the train is running. In scenario 3, Ned can flip a switch that will cause the speeding train to be diverted onto a side track that loops back to the main track; but on the side track is a fat man. In this case, if Ned does nothing, the five people trapped on the main track will be killed by the train. However, if Ned flips the switch, the train will be diverted to the sidetrack and thereby kill the fat man, but be sufficiently slowed for doing so to allow the five people trapped on the main track to escape. Finally, scenario 4: this scenario is the same as scenario 3, except that on the side track are both a normally sized man and a heavy weight. The man is standing in front of the weight and will be killed if the train is diverted by Oscar onto the side track. However, the weight is itself heavy enough to sufficiently slow the train to give the five trapped on the main track time to escape. So, in this case, the man will be killed not as a means of saving five, but as an anticipated side-effect of intending the train to hit the weight, which is the cause or the means by which the lives of the five are saved.

Hauser et al. found that 85% of their respondents judged it permissible to divert the train in scenario 1, while only 12% judged it permissible to shove the man in scenario 2. Judgments about scenarios 3 and 4 were much closer: 56% of respondents said that it is morally permissible for Ned to throw the switch in scenario 3, while 72% said that is is morally permissible for Oscar to throw the switch in scenario 4.[30] They also found that this pattern of judgments was

[30]Hauser et al. 2007 p.6

shared across almost all demographics they examined, and that participants did not do very well when it came to offering justifications of their moral intuitions.

However, I think that it is premature to conclude that the experimental data produced by the Moral Sense Test actually support this conclusion. But before explaining why I believe this, I want to first address an interesting problem in Hauser et al.'s articulation of the details of the moral psychology that they predict by way of the linguistic analogy. The problem concerns the conceptual framework that Hauser et al. are importing into their research on moral psychology from Chomskyan linguistic psychology. For one of the features of the linguistic analogy that makes it a scientifically plausible theory is that it, if it succeeds, it will generate consilience between Chomskyan linguistic psychology and moral psychology. However, this can only happen if the conceptual framework deployed within Chomskyan linguistic psychology is deployed in the same way in moral psychology.

Now, one of the conceptual resources that Hauser et al. believe can be imported into moral psychology is Chomsky's well-known competence / performance distinction. Hauser et al. write that within the conceptual framework of the linguistic analogy they

> expect a dissociation between our competence and performance – between the knowledge that guides our judgments of right and wrong and the factors that guide what we actually say or do; when confronted with a moral dilemma, what we say about this case or what we actually would do if confronted by it in real life may or may not map onto our competence.[31]

[31] Hauser et al. 2008a p.125. Compare this with their description of the competence / perfor-

Hauser et al. identify moral competence with the knowledge that guides our judgments of right and wrong, and performance with "the factors" that guide what we actually say or do. And they think that competence does not map onto performance. The dissociation that they have in mind is given by the title of their 2007 article (*A Dissociation Between Moral Judgments and Justifications*),[32] where, again, they report that when experimental participants are asked to justify their intuitive moral judgments, most are unable to articulate the rules or principles that Hauser and his team believe are guiding experimental participant's moral intuitions. It would appear, then, that Hauser and his team have produced some experimental data supporting the dissociation between competence and performance, or in other words between moral judgment and justification, thereby vindicating the application of Chomsky's distinction in the domain of moral psychology. The idea here is not just that there is some unconscious knowledge of moral principles but, as suggested by using the term 'competence' to denote this knowledge, this unconscious moral knowledge is both epistemically and causally central to moral cognition.

But there is both a small mistake and a large mistake here. The small mistake is that Hauser and his team have not used Chomsky's terminology correctly. Chomsky never uses "competence" and "performance" to denote a body of knowledge and a body of additional factors respectively.[33] Rather, his usage

mance distinction in linguistics, at Hauser et al. 2008a p.110-111.

[32]For more evidence supporting this interpretation of Hauser et al.'s use of the competence performance distinction, see Hauser 2006. This is Hauser's public introduction to the moral psychology of the linguistic analogy. Some of the tropes of this work are, e.g., "the moral organ" and "a moral instinct", which are both names for "a faculty of the human mind that unconsciously guides our judgments concerning right and wrong, establishing a range of learnable moral systems, each with a set of shared and unique signatures.... The notion of a universal moral grammar with parametric variation provides one way" of understanding how moral systems are implemented in human cognition. (Hauser 2006 p.425)

[33]Chomsky has not changed how he uses these terms. In one of his earlier works, *Aspects of the Theory of Syntax*, we read that,

of these terms is by comparison both more precise and more abstract: he uses "competence" to denote a body of knowledge and "performance" to denote actual uses of some of the knowledge denoted by "competence". So, for example, someone's knowledge of a syntactic rule would count as a part of their linguistic competence, and their use of the syntactic rule in producing or interpreting speech would count as a performance of that knowledge.

Now, one of Chomsky's reasons for introducing the distinction is to capture the idea that, because of the complexly integrated nature of human cognition, performances cannot be interpreted as perfect reflections of the properties of competence. Someone's grammatical knowledge may allow for sentences of unbounded length, but because of the limitations of memory, sentences that are 1,000 words in length might be (erroneously) judged by this person as ungrammatical. Examples like these make Chomsky's key point, namely that we should not read our theories of an underlying competence straight off the content of its various performances, even though the only way we can achieve epistemic access to an underlying competence is by way of its performances. But the smaller

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogenous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and error (random or characteristic) in applying his knowledge of the language in actual performance[...] To study actual linguistic phenomena, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one.[34]

He continues, "We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). Only under the idealization [mentioned above] is performance a direct reflection of competence."

Then, in *Rules and Representations* he writes, "Theories of grammatical and pragmatic competence must find their place in a theory of performance that takes into account the structure of memory, our mode of organizing experience, and so on... To the extent that we have an explicit theory of competence, we can attempt to devise performance models to show how this knowledge is put to use." (Chomsky 1980 p.225) And in *The Minimalist Program* (which, incidentally, is one of the main sources of inspiration for the linguistic analogy), Chomsky writes "We distinguish between Jones's *competence* (knowledge and understanding) and his *performance* (what he does with that knowledge and understanding)." (Chomsky 1995 p.14)

point that I want to stress here is that it is not consistent with Chomsky's usage to use "performance" to denote "the factors" guiding what we say or do, where what we say or do are intended to *not* be uses of people's underlying moral competence, because competence is "dissociated" from performance. For, using Chomsky's terminology and supposing that Hauser and his team are right, people's moral intuitions are all a type of performance, *even if* they are better reflections of some underlying epistemic competency.

So, the failure to accurately import Chomsky's distinction into the domain of moral psychology damages the case that can be made for consilience between Chomskyan linguistic psychology and moral psychology according to the linguistic analogy. But this is a small error, and it is easy to remedy.

Behind this terminological issue, however, there is an important philosophical issue that is worth paying more attention to. For I believe that what Hauser and his team really want to communicate by way of the competence / performance distinction is the idea that psychological processes that are mediated by explicitly accessible moral information are not regulated by people's actual moral competence, which itself consists of a body of unconsciously represented moral knowledge, and which is only expressed by way of people's intuitive moral judgments. Put another way, what we say or do[35] in the moral domain are (surprisingly) not manifestations of our moral competence, but our moral intuitions are. So, the issue masked by Hauser and his team's terminology is this: does the fact that people are not very good at justifying their seemingly accurate, seemingly rule-guided intuitive moral judgments indicate that in humans the epistemic centre of moral cognition is just our unconscious moral knowledge? Should "moral competence" and its cognates be used to denote in humans just

[35]Except, presumably, for manifesting moral intuitions.

a body of unconscious moral knowledge, roughly the same as "grammatical competence" is used by Chomskyean linguists to denote a body of unconscious linguistic knowledge?

Perhaps the easiest way to see why this issue matters is to contrast Hauser and his team's conception of moral competence with what might be called the standard view of moral competence amongst philosophers. (Here, we return to a number of issues that were raised in section 1.) For it is common for moral philosophers to hold that people's moral knowledge is consciously accessible to them in varying degrees, and that, also in varying degrees, it is realized both affectively and non-affectively. Furthermore, it is common to hold that people's moral judgments, choices, actions, inferences, et cetera, are determined sometimes by affectively-realized moral knowledge, and at other times by non-affective and unconscious moral knowledge, and so on through the possible combinations. And on top of this, because people often engage in public moral deliberation, a person might form some particular moral judgment because, say, someone she perceives as a moral expert has formed a similar judgment, or because she has just heard a particularly compelling argument in favour of some particular moral judgment. In view of these points, the philosophical use of "moral competence" would denote the rather lumpy and heterogeneous collection of a person's variously conscious, unconscious, affective, non-affective, socially-acquired, and privately-acquired items of moral knowledge. So, the philosophical category of moral competence is noticeably broader than Hauser and his team's conception of moral competence.[36]

[36]Of course, this is not to deny that various traditions in moral philosophy place different amounts of emphasis on these various components of moral competence. But all the same, Humeans will not deny that explicit moral beliefs are a common feature in people's moral psychologies, and I am sure that no deontologist has even denied that people are often motivated by their moral emotions.

However, by setting up (and seemingly vindicating experimentally) the competence / performance distinction so that "competence" is implicitly used to denote only people's unconscious moral knowledge, Hauser and his team are mounting a challenge to the standard view. Their argument seems to be this: consciously and affectively represented moral knowledge is not responsible for producing some seemingly accurate moral judgments (because the contents of the responses that experimental participants provided to the justification prompt do not refer to the properties that Hauser and his team believe experimental participant's intuitive moral judgments were sensitive to), but unconscious moral knowledge is, so we should narrow the category of moral competence to unconsciously represented moral knowledge only.[37]

But this is not a very compelling argument. Evidence that people are not very good at justifying their intuitive moral judgments – in the sense of being able to explicitly state the information that is presumably guiding their judgments – does not provide any reason to prefer the 'narrow' conception of moral competence that Hauser and his team favour to the 'wider' conception of moral competence. This is so because it is easy to find alternative explanations of this evidence that are not inconsistent with the 'wide' view of moral competence, but at the same time are inconsistent with some of the core ideas of the linguistic analogy.

It may be that, for instance, people's moral knowledge is not realized psy-

[37]Of course, Hauser and his team also think that they can show that this unconscious knowledge consists of knowledge of moral principles (like the principle of double-effect) on the basis of demonstrating that people's intuitive judgments are consistent with the principles that have in mind. As for the evidence that this body of knowledge is consistent and coherent, so far as I can tell the case for this view rests on nothing more than an optimistic projection from the terms of Chomsky's linguistic psychology. That said, as we will soon see, the view that people's moral knowledge is both consistent and coherent plays an important role in Hauser et al.'s assessment of some of their experimental evidence.

chologically in a way that makes it easy for them to be able to justify (in the appropriate sense of justification) their moral judgments. Perhaps people do have unconscious knowledge of some moral rules, and when they form judgments using these rules, they have a hard time justifying their judgments because the rules that have used are not consciously accessible. Someone could accept this much and still deny that what unconscious moral knowledge people have consists in knowledge of a consistent and coherent system of moral rules, as is suggested by proponents of the linguistic analogy. Likewise, someone who accepts even that people do have unconscious knowledge of some moral rules can still deny that the knowledge of these rules is developmentally-endogenous – hockey referees, after all, are paid to manifest accurate 'referee-intuitions' that are guided by unconscious knowledge of a collection of rules, where this knowledge is obviously not innate. Or, someone might hold the position that, while people do have unconscious knowledge of some learned moral rules, their unconscious moral knowledge is not exhausted by this knowledge; that people have unconscious moral knowledge that has a variety of logical forms.

Similarly, it could also be that people's moral knowledge is not structured in a way that makes it easy for them to be able to justify (again, in the appropriate sense) their moral judgments. Proponents of the 'wide' view of moral competence are certainly not required to hold that people's moral knowledge – whether it is unconsciously, consciously, affectively, or non-affectively represented, and whether it is socially or privately acquired – is even remotely consistent or coherent. Someone's conscious moral beliefs may be inconsistent with some of the unconscious moral rules they have learned, or they might have conscious knowledge of a hodgepodge of moral rules, particular beliefs, rules of thumb, and borrowed judgments from perceived moral experts – where these

various items are neither wholly consistent with one another, or form a completely coherent moral theory. A moral psychology that allows for conflicts amongst the various items of a person's moral knowledge can explain rather easily why people are ordinarily unable to explicitly justify their intuitive moral judgments, even if they are aware of the moral knowledge that caused a sequence of their judgments.

So, it is easy enough to see that evidence that people are not very good at justifying their intuitive moral judgments possess no threat to the 'wider' view of moral competence. In fact, it might even be a prediction of the various moral epistemologies held by many philosophers that people typically will not be very good at justifying their moral judgments, intuitive or not.

But I do not want to pursue that thought here. Instead, let me return to the idea that what is really going on here is that Hauser and his team want to show that, when it comes to successfully tracking the moral properties of the world, explicit moral reasoning and moral affect both perform extremely poorly compared to the tracking abilities of moral intuition. The idea, then, is that in order to explain what success moral cognition does enjoy, we only need to refer to people's unconscious or intuitive moral knowledge. Hence, if, as it seems reasonable to do, we should use "moral competence" to denote only the knowledge that actually plays a role in helping people navigate the moral world, then evidence that only people's unconscious moral knowledge plays such a role would seem to warrant using "moral competence" as Hauser and his team do.

But it is odd that justification prompts should be the test used to assess whether or not moral reasoning and moral affect are reliable. It is not clear

how the fact that people are not very good at justifying even their most accurate moral judgments can be used to show that only people's unconscious moral knowledge is accurate. This is because there is a difference between being able to use explicit moral reasoning to produce reliable moral judgments, and being able to use explicit moral reasoning to justify one's moral judgments. Someone can be quite good at the first (in the sense that her moral reasoning – and for what it is worth, her moral sentiments – tend over time to produce in her judgments that are typically true), and be quite bad at the second (in the sense that she cannot normally explicitly justify her moral judgments). Suppose for sake of argument that some broadly Kantian ethical theory is correct. If so, Kantians are free to think that unreconstructed utilitarians still achieve pretty good descriptions of our duties, because after all, many of the duties that should motivate us are connected in subtle and complicated ways with considerations of utility. Of course, Kantians will think that at an abstract level utilitarians are using the wrong conceptual framework in which to peruse ethical questions, and that because of this utilitarians do not normally get it right when it comes to explicitly justifying their ethical conclusions. But neither of these points mean that Kantians must reject all of the reasoning produced by the utilitarians as unreliable. There is room for the Kantian to interpret many of the utilitarian's conclusions as basically correctly, *and correct because of the truth-tracking abilities of the deliberative reasoning engaged in by the utilitarians*. So, the moral reasoning that the people engage can have an epistemically causal role in producing approximately-true moral knowledge even if the justifying conceptual or theoretical framework that they rely on is not correct.

Examples like this show that there is no a priori inference from evidence that people are not very good at justifying their intuitive moral judgments to the

conclusion that people are not very good (in the sense of reliable) explicit moral reasoners. Indeed, a process of explicit reasoning can be said to be reliable if it is sensitive to, or somehow governed by, the world's moral properties – and this holds even if all of the moral judgments it produces over a period of time are false (because the person manifesting the judgments is still learning), or if the reasoning process makes uses of cognitive resources that are neither perfectly coherent or consistent at some time (because moral inquiry is still continuing), or if the reasoning process makes use of imperfect cognitive resources (because we have not yet developed epistemically ideal moral concepts and inferential strategies, or the ethical theory that provides a framework for our inquiry is not correct).

Here's an analogy concerning inductive reasoning and Bayesian epistemology that should help further illustrate the point I'm trying to make. Bayesian epistemologists believe that scientists should update the probability that they assign to hypotheses they are considering by applying Bayes' theorem. But these epistemologists are of course aware that scientists do not use Bayes' theorem when reasoning inductively, and also that very few scientists could appeal post hoc to Bayes' theorem to justify accepting some hypothesis and rejecting another. And, more importantly, these epistemologists also believe that scientific reasoning is one of the most reliable forms of reasoning we have discovered. So the fact that scientists cannot usually use Bayes' theorem to justify their inductive reasoning shows absolutely nothing about whether or not in scientists such reasoning is by and large reliable. The fact that some people may be unable to explicitly justify some pattern of reasoning or judgment tells us nothing about whether or not the pattern of reasoning or judgment is reliable.

Anyway, it is time to move on. In the next section, I will demonstrate that the experimental data reported by Hauser and his team with the intention of establishing both that (a) some people's intuitive moral judgments are guided by knowledge of moral rules or principles, and that (b) people are ordinarily unable to verbally justify these judgments does not actually support either of these propositions. In fact, we will soon see that Hauser and his team's experimental data even fails to establish a proposition that is independently obvious, namely that intuitive moral judgments are guided by unconscious knowledge.

2.4 The Moral Sense Test

I noted above that Hauser and his team tested whether their experimental participants could provide adequate justification for their (the experimental participant's) moral intuitions. Hauser et al. found that their participants were generally not able to provide adequate justifications for intuitions that seemed to fit their target principles, and this result is their primary evidence for the dissociation that we have been discussing.

We have also seen that Hauser et al.'s method involves constructing scenarios or cases that differ in terms of abstract moral categories, so that differing judgments about these cases can be explained by positing knowledge of principles framed in terms of these moral categories. Now, if you have done enough moral philosophy, then the most salient difference between scenarios 3 and 4, for example, will be probably this: scenario 3 involves treating a man's death as a means to a good outcome, while scenario 4 involves treating a man's death as a foreseen consequence of choosing a means of achieving a good outcome. Part of

what the principle of double-effect holds is that actions or choices with bad foreseen consequences are morally permissible, just so long as the consequences are not the means to a good outcome, and that the 'goodness' of the good outcome sufficiently outweighs the 'badness' of the foreseen consequences. Accordingly, as Hauser et al. write, "scenario 3 ... and scenario 4 were designed to probe just ... the principle of the double effect",[38] and they report a statistically significant difference between the number of participants who judged it permissible to turn the switch in scenario 3 and the number of participants who judged it permissible to turn the switch in scenario 4. This difference can be explained by the hypothesis that some experimental participants do use the principle of double effect when arriving at their moral intuitions about each case.

When Hauser et al. examined their experimental "subject's ability to explicitly articulate the principle(s) responsible for their pattern of judgments",[39] they found that few participants who manifest contrasting intuitions about scenarios 1 and 2, and about scenarios 3 and 4, were able to provide statements that - according to Hauser et al.'s standard of justification - adequately justified this contrast. To arrive at this conclusion, Hauser et al. coded the putative justifications of the participants who provided contrasting judgments into one of three categories: sufficient justification, insufficient justification, and discountable justification. The authors defined the sufficient justification category to include "factual differences between the two scenarios and claimed the difference to be the basis of moral judgment",[40] while "an insufficient justification was one that failed to identify a factual difference between the two scenarios."[41] The examples of factual differences that they give are:

[38]Hauser et al. 2007 p.15
[39]Hauser et al. 2007 p.12
[40]Hauser et al. 2007 p.13
[41]Hauser et al. 2007 p.13

> in scenario 1, the death of the one man on the side track is not a necessary means to saving the five, while in scenario 2, the death of the one man on the bridge is a necessary means to saving the five; ... in scenarios 1, 3, and 4, an existing threat (of the train) is redirected, while in scenario 2, a new threat (of being pushed off the bridge) is introduced; ... in scenarios 1, 3, and 4, the action (turning the train) is impersonal, while in scenario 3, the action (pushing the man) is personal or emotionally salient.[42]

An insufficient justification included responses like " 'I don't know how to explain it', 'It just seemed reasonable', and 'It struck me that way.' " Hauser et al. also included in this category those respondents who "explained their judgment of one case using utilitarian reasoning (maximizing the greater good) and their judgment of the other using deontological reasoning (acts can be objectively identified as good or bad) without resolving their conflicting responses", and finally, subjects who "referred to principles, or moral absolutes, such as (1) killing is wrong, (2) playing God, or deciding who lives and who dies, is wrong, and (3) the moral significance of not harming trumps the moral significance of providing aid."[43] As for the third category, discountable justification, Hauser et al. included within it those justifications that were either blank or those that included additional assumptions, such as "men walking along the tracks are reckless... a man's body cannot stop a train... the five men will be able to hear the train approaching and escape in time."[44]

Using these coding standards, Hauser et al. found that of the 597 respon-

[42]Hauser et al. 2007 p.13
[43]Hauser et al. 2007 p.13-14
[44]Hauser et al. 2007 p.14

dents who provided justifications for differing judgments about scenario 1 and scenario 2, 267 were in the category of discountable justification, and of the remaining 330 participants, 70% provided insufficient justifications, and 30% provided sufficient justifications. Put another way, nearly half of the respondents provided discountable justifications for differing judgments about scenarios 1 and 2, while about 38% of the total respondents provided insufficient justifications, and only about 11% of the total respondents provided sufficient justifications. Hauser et al. do not provide the ratio of blank responses to additional assumption for the category of discountable justification, nor do they provide data about the distribution of responses like, e.g., "a man's body cannot stop a train" in either the category of discountable justification or insufficient justification.

The data on differing judgments for scenarios 3 and 4 is less clear. Hauser and his team write, "the proportion of subjects who judged scenarios 3 and 4 differently within a single session was quite small (5.8%), [... so, i]n order to generate a new, larger sample of subjects with potentially conflicting judgments in scenario 3 and 4, we re-contacted subjects who had been presented with only one of the scenarios and who had judged scenario 3 as impermissible or scenario 4 as permissible, and asked them to make a judgment on the corresponding case."[45] 207 participants responded, and "33% judged the foreseen case (scenario 4, Oscar) permissible and the intended [or means] case (scenario 3, Ned) impermissible."[46] Using the same three categories of justification, "between 2% and 34% of individuals who perceived a difference between these scenarios would be able to provide a sufficient justification for their judgments."[47]

[45]Hauser et al. 2007 p.14
[46]Hauser et al. 2007 p.15
[47]Hauser et al. 2007 p.15

So, those participants who manifest differing moral intuitions about the two pairs of scenarios were generally not very good at providing sufficient justifications for their two intuitions, at least in the terms that Hauser et al. were expecting.

But notice that, although Hauser et al. do not give more than a few examples of "insufficient" or "discounted" justifications, most of the examples they do actually report appear to be able to justify the patterns of intuitions that Hauser et al. observed, so long as a reasonably relaxed standard of justification is employed. To mention two examples, recall that Hauser et al. graded as insufficient justification any reasoning that "explained their judgment of one case using utilitarian reasoning ... and their judgment of the other using deontological reasoning ... without resolving their conflicting responses."[48] But maybe some people hold a hybrid utilitarian-deontological moral theory; there is no a priori reason why this could not be the case. (One can see here the role that the idea that people's moral knowledge is both consistent and coherent is playing in Hauser and his team's assessment of the justifications provided by their experimental participants.) Or maybe people's ordinary moral theories are not versions of either deontology or consequentialism, meaning that it is just a mistake to try to interpret their moral views through the lenses of these two families of ethical theory. Secondly, Hauser et al. also coded as insufficient justification any reasoning that "referred to principles, or moral absolutes, such as (1) killing is wrong, (2) playing God, or deciding who lives and who dies, is wrong, and (3) the moral significance of not harming trumps the moral significance of providing aid."[49] Each of each of these principles could plausibly explain a difference in judgments about the scenarios, especially if we allow ourselves some addi-

[48]Hauser et al. 2007 p.15-16
[49]ibid.

tion assumptions about how experimental participants perceived each scenario.

Moreover, it is hard to see what the relevant epistemological difference is between the principle of double effect and the principle that says that not harming trumps providing aid. It is disconcerting that Hauser et al. coded these responses as insufficient justifications, given that the purpose of the justification test is to determine whether, again, experimental subjects have the "ability to explicitly articulate the principle(s) responsible for their pattern of judgments". And it is even more worrying that these responses were classified as insufficient justification given that Hauser et al. are occasionally reluctant to claim that they have compelling evidence that, e.g., the principle of double-effect and not some other principle is actually guiding people's moral intuitions.[50] For all they have shown is that the principle of double effect could sometimes be unknowingly used by some people in developing their moral intuitions. But it seems that Hauser et al.'s experimental participants are supplying them with a wealth of alternative principles that they should incorporate into their analysis of the data provided by the Moral Sense Test. It is possible, after all, that some of the principles supplied by the experimental participants are able to capture some of the observed patterns in people's moral intuitions just as well as Hauser et al.'s "target principles".

In fact, these responses are a real problem for Hauser and his team's research strategy, because they seem to show that experimental participants have conscious access to some of the moral principles that are responsible for their moral judgments. True enough, many of the principles that the experimental participant's do report are not nearly as high-minded or abstract as the principle of

[50]See, e.g., Hauser et al. 2008a p.141; compare this passage with the main argument in Hauser et al. 2007.

double effect. But the fact that some experimental participants did produce these alternative principles implies either that (a) Hauser and his team were not actually experimenting on only moral intuitions, or that (b) it is wrong to think that moral intuitions are caused by only unconscious knowledge. Hauser and his team should accept (a) if they also accept that moral intuitions are caused by unconscious knowledge, but this requires them to abandon the conclusion that they think the Moral Sense Test supports. However, to accept (b) is to be forced to deny something that is independently obvious on phenomenological grounds, namely that intuitions are caused by unconscious knowledge. So, it seems that Hauser and his team have worked themselves into a dialectical situation in which they may be forced to reject something that is independently obvious.

Leaving this line of thought aside, though, it is unfortunately the case that similar criticisms apply to the category of discountable justification. It is odd that responses like "a man's body cannot stop a train" should be graded by Hauser et al. as illegitimate additional assumptions, since one of the obvious factual differences between scenario 1 and 2 and scenarios 3 and 4 is this: both 2 and 3 contain choices of outcomes that involve nomological impossibilities, while 1 and 4 do not. Consider scenario 2, for example. It describes a situation that obviously violates at least two basic physical and physiological generalizations that almost everyone will have knowledge of. By hypothesis, in all the scenarios the train must be initially moving with enough velocity to kill five ordinary sized (I assume) adults. So, in order to actually prevent the deaths of the five by shoving a large man onto the tracks, the man Frank must shove must have a mass at least greater than the combined mass of five normal sized adults. But that is true only assuming that there exists a train that could in fact

be stopped by the mass of a single man, even one heavier than five ordinary sized adults. However, neither of these two assumptions are at all plausible – and so it is hard to understand why Hauser et al. discounted justifications that took into account these two pieces of obviously accurate background information.

Indeed, given that people cannot be prevented from making additional assumptions about cases like those used in the Moral Sense Test,[51] and that it would be extremely hard to control for what additional assumptions people do in fact make (especially when the cases are weird and far-removed from ordinary life as Hauser et al.'s trolley cases are), it is rather puzzling that Hauser et al. should think that any vaguely plausible additional assumptions could be ruled out as illegitimate. After all, it seems as though some of these additional assumptions were involved in producing at least some of their experimental participant's moral intuitions, and perhaps even some of the articulated justification for these moral intuitions as well.

Let me try to better illustrate this point with an example borrowed from a paper written by Jerry Fodor and Charles Chihara. The example specifically is the rule in basketball that says that a field goal occurs when the basketball passes through the hoop. Fodor and Chihara ask, "Suppose ... that a player takes a long two-handed shot and that the ball suddenly reverses its direction, and after soaring and dipping through the air like a swallow in flight, gracefully drops through the player's own basket only to change into a bat, which immediately entangles itself in the net. What do the rules say about that?"[52] My intuition is that it is not a field goal. But – and this is my point – I did not get

[51] See Chapter 1
[52] Fodor & Chihara 1981 p.40

this intuition by relying on just my knowledge of the rule governing field goals in basketball alone. Perhaps I relied on some background beliefs about the rules of basketball only being in force when the games is suitably stable, or maybe I made some assumptions about how strictly to interpret the rule. The important point here is just that examples like this show that, even if our intuitive moral judgments are guided by knowledge of moral principles, it is unlikely that just this knowledge alone – i.e. knowledge of whatever moral principle is behind a judgment – is used to generate the judgment. We will need more than just a description of these moral principles in order to be able to adequately account for the psychological causes of whatever patterns may be observed in people's ordinary moral judgments.

Summing up: Hauser et al. have used a standard of justification that clearly favours the hypothesis they are setting up to test. Hauser and his team have applied an overly strict standard of justification.[53] I believe that they have done this because they have simply assumed that people's moral knowledge is largely consistent and coherent, and that people's moral intuitions are guided by unconscious knowledge of moral principles (and not unconscious knowledge that has some other logical form). But both of these propositions cannot be assumed until after an adequate defense of the core ideas of the linguistic analogy has been mounted. We must conclude therefore that the data published to date from the Moral Sense test does not show whether or not people are able to justify their moral intuitions. Similarly, these data cannot be used to support the view that some moral judgments are guided by unconscious knowledge of certain moral principles.

[53] And just to be clear, my complaint is not that they used overly fictitious or unrealistic scenarios – even though, plausibly, people's ability to justify their moral judgments is tied to how familiar the moral situation which elicits their judgment is.

2.5 Models of Moral Judgment

I have now considered in some detail the case defenders of the linguistic analogy have made for two of the three ideas that I believe form the 'core' of the linguistic analogy. In this section, I want to return to an issue that was first raised back in section 2.3 – the issue of unconscious moral knowledge.

In section 2.1, I reported that Hauser and his team understand one of the central questions in moral psychology to be: what causes moral judgments? I also reported that the believe that they think the field is dominated by three different models that each provide a different answer to this question.[54] To briefly review, the Kantian model holds that moral judgments are caused by explicit moral reasoning, the Humean model holds that moral judgments are caused by explicit moral affect, and the 'Hybrid' model holds that moral judgments are caused by explicit moral reasoning and moral affect working in concert. Hauser and his team favour a new model, the Rawlsian model, which holds that unconscious computations undertaken by the morality module are the source of moral judgment.

The point that I want to stress here is that these models seem too simple. For it is hard to see why an obvious alternative model of moral judgment has been overlooked by Hauser and his team – namely a model that holds that moral judgments are sometimes caused by either (i) explicit moral reasoning, (ii) unconscious moral knowledge, (iii) conscious moral affect, (iv) unconscious moral affect, or (v) deference to the judgments of another moral agent, or (vi) some combination of the previously mentioned potential causes. This model is, obviously, much more complicated than any of the models described by Hauser and

[54] See Hauser et al. 2008a p.113-121

his team. But it seems doubtful to me that anyone who has thought seriously about the complexity of human motivational psychology in the moral domain has ever denied the accuracy of a model such as this – least of all Humeans and Kantians.

But that said, we can see that Hauser and his team conceive of their favoured moral psychology as the only position in the debate that takes seriously the role that unconscious moral knowledge plays in producing some moral judgments. Hauser and his team's way of framing the debate makes it appear as though all of the plausible alternatives to the linguistic analogy's moral psychology deny the role of unconscious moral knowledge in producing some moral judgments. Because of this, unsophisticated readers of Hauser and his team's work might conclude, once they come to appreciate the (obvious) fact that people have some unconscious moral knowledge and that this knowledge sometimes plays a role in moral cognition, that the moral psychology of the linguistic analogy is the only scientifically (and therefore philosophically?) credible moral psychology on offer. However, since the fact that people have some unconscious moral knowledge is, frankly, unsurprising, by itself this fact gives us no reason whatsoever to believe that the linguistic analogy is apt.

Indeed, we have good reason to believe that unconscious knowledge is ubiquitous. It would be weird if our knowledge of moral phenomena were to turn out to be the only kind of knowledge that humans have that was *not* partially unconscious. Here, it is helpful to be reminded of some of Thomas Kuhn's remarks. Kuhn believed that becoming competent in any scientific field involving acquiring a body of tacit knowledge "which is learned by doing science rather than by acquiring rules for doing it."[55] Kuhn goes on to emphasize that our

[55]Kuhn 1996[1969] p.191

ability to reliably interpret the world in any domain depends upon a large and inarticulable body of knowledge that cannot be captured by any system of rules. If Kuhn is right (and I believe he is), then living a moral life will sooner or later – just as it does in other domains – produce in an individual a body of tacit or unconscious moral knowledge that will sometimes be responsible for some of that individual's judgments.

So, we should expect that any scientifically and philosophically plausible moral psychology will refer to unconscious moral knowledge. Of course, what makes the linguistic analogy an original view in moral psychology is not just that it draws attention to the fact that people have unconscious moral knowledge, but that it also puts forward the bold proposal that moral cognition in humans is almost entirely implemented in only people's unconscious moral knowledge. However, we have now seen that there is little compelling evidence that this is true.

2.6 Conclusion

My assessment of the plausibility of the evidence offered to dare for the three core ideas behind the linguistic analogy is now complete. However, in the proceeding sections we encountered two important issues that should be addressed: what to think of the evidence that there are some widely-shared intuitive moral judgments, and what to think of the fact that people can successfully project in the moral domain. By way of conclusion, I want to offer a few remarks on these two issues.

As we saw, Hauser and his team presented evidence that some patterns of

moral intuitions are widely shared across many different demographic indicators. Of course, in virtue of its commitment to a developmentally-endogoenous set of principles that develop into our moral competence, the moral psychology behind the linguistic analogy predicts that there will be some more or less universal patterns of moral intuitions. But many other plausible non-nativist moral psychologies share exactly this same prediction, and just about any moral psychology whatsoever is formally compatible with the existence of widely-shared patterns in people's moral intuitions.

Indeed, consider how a moral sceptic might account for the existence of widely-shared patterns of moral intuitions. For even on the assumption that there are no moral facts to constrain the construction of moral theories, there will be other constraints that the sceptic can appeal to in order to explain how the content of two differing moral systems can in some places be the same. I have in mind such constraints as those imposed by the shared framework provided by human rational and emotional capacities, but also various kinds of shared background knowledge in non-moral domains, shared problems of social coordination, and perhaps even shared economic, political, and religious institutions.

As for the issue of projectibility in the moral domain, it is of course possible that it is solved by the presence of an innate moral grammar. But then again, it is not necessary to go in for a whole moral grammar when any kind of innate moral knowledge will do just as well. For instance, some psychologists who hold the view that morality has evolved through natural selection endorse the hypothesis that we have some innate moral sentiments – that we are "prepared" to feel, e.g., disgust about some kinds of events and not others.[56] Or, if there is

[56]See, e.g., Haidt 2001

not a form of moral nativism that strikes you as plausible, it is possible to explain how it is that we are able to successfully project in the moral domain using a theory that relies on cognitive resources that are used in a range of epistemic domains, beyond morality. Maybe a broadly Humean moral psychology is correct. Perhaps we bring ourselves into the world of reliable moral cognition by exercising imagination and self-interested reasoning together. We see another child in pain and then imagine what it would be like to experience that pain for oneself. Maybe this is all that is required in order to begin the process of inculcating other-regarding moral sentiments, which then sooner or later leads to practical reasoning about how to avoid activities that cause negative moral sentiments and also philosophical reasoning about whether the moral sentiments felt in response to certain phenomena are, in fact, the appropriate sentiments to have. Perhaps in the fullness of time this thinking can ultimately lead to behaviours that are designed to habituate into one's character only the appropriate moral sentiments. All this to say, it is not the case that innate knowledge of a system of moral rules is the only possible, or plausible, explanation of how we are able to solve the problem of projection for moral thinking specifically.

BIBLIOGRAPHY

[1] Bloom, P. and Jarudi, I. (2006) The Chomsky of Morality? *Nature* 443(26), 909-910

[2] Chomsky, N. (1980) Opening the Debate. In Piattelli-Palmarini, M (ed.) *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Routledge and Keegan.

[3] Chomsky, N. (1968) *Language and Mind*. Harcourt, Brace, and Jovanovich.

[4] Chomsky, N. (2000) *The Architecture of Language*. Oxford University Press.

[5] Dwyer, S. (1999) Moral Competence. In Kumiko Murasugi & Robert Stainton (eds.) *Philosophy and Linguistics*. Westview Press.

[6] Dwyer, S. (2003) Moral Development and Moral Responsibility. In *The Monist*, 86, 181-199

[7] Fodor, J. and Chihara, C. (1981) Operationalism and Ordinary Language. In Fodor, J. *Representations*. MIT Press.

[8] Haidt, J. (2001) The Emotional Dog and its Rational Tail. In *Psychological Review*, 108(4), 814-834

[9] Hauser, M. (2006) *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*. Harper Collins.

[10] Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., and Mikhail J. (2007) A Dissociation Between Moral Judgments and Justifications. In *Mind & Language*, 22(1), 1-21

[11] Hauser, M., Young, L., and Cushman, F. (2008a) Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions. In Sinnott-Armstron, W. (ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. MIT Press.

[12] Hauser, M., Young, L., and Cushman, F. (2008b) On Misreading the Linguistic Analogy: Response to Jesse Prinz and Ron Mallon. In Sinnott-Armstron, W. (ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. MIT Press.

[13] Hotz, R. L. (2007) Scientists draw Link between Morality and Brain's Wiring. *The Wall Street Journal*, May 11.

[14] Joyce, R. (*forthcoming*) Preçis of *The Evolution of Morality*, *Philosophy and Phenomenological Research*

[15] Joyce, R. (2006) *The Evolution of Morality*. MIT Press.

[16] Kluger, J. (2007) What makes us Moral? *Time Magazine* Nov. 21, 2007.

[17] Kuhn, T. (1996[1969]) *The Structure of Scientific Revolutions*. University of Chicago Press.

[18] Mikhail, J. (2002) Aspects of the Theory of Moral Cognition: Investigating Intuitive Knowledge of the Prohibition of Intentional Battery and the Principle of Double Effect. unpublished working paper, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=762385`.

[19] Mikhail, J. (2008) The Poverty of the Moral Stimulus. In Sinnot-Armstrong (ed.) *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness*. MIT Press.

[20] Mikhail, J. (2000) *Rawls' Linguistic Analogy: A Study of the 'Generative Grammar' Model of Moral Theory Described by John Rawls in "A Theory of Justice"*. PhD dissertation, Cornell University.

[21] Mikhail, J. (2007) Universal Moral Grammar: Theory, Evidence, and the Future. *Trends in Cognitive Sciences*, 11(4), 143-152

[22] Pinker, S. (2008) The Moral Instinct. *The New York Times*, January 13.

[23] Prinz, J. (2008) Resisting the Linguistic Analogy: A Commentary on Hauser, Young, and Cushman. In Sinnott-Armstron, W. (ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity* MIT Press.

[24] Pullum, G. and Scholz, B. (2002) Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 19(1-2), 9-50

[25] Reali, F. and Christiansen, M.H. (2005). Uncovering the richness of the stimulus: structure dependence and indirect statistical evidence. *Cognitive Science* 29, 1007-1028

[26] Ritter, N. (2002) Introduction. *The Linguistic Review*, 19(1-2), 1-7

[27] Sampson, Geoffrey (2005) *The 'Language Instinct' Debate*. Second Edition, Continuum International.

[28] Shweder, R., Mahapatra, M., and Miller, J. (1987) Culture and moral development. In Kagan & Lamb (eds.) *The emergence of morality in young children*. University of Chicago Press.

[29] Waldmann, M. (2006) A case for the moral organ? *Science*, 314, 57-58

CHAPTER 3

AGAINST THE SOCIAL INTUITIONIST MODEL OF MORAL JUDGMENT

3.1 Introduction

The social intuitionist model of moral judgment is a scientific theory of morality, and its basic commitments are these. Moral judgments are derived from, and both personal moral cognition and interpersonal moral deliberation are regulated by, five moral modules that cause quick and automatic affective responses to stimuli. These affective responses are called "moral intuitions." Evolutionary processes favoured these modules because, with the help of moral reasoning (although "reasoning" is not the right word; more on this in section 3.2), they created and maintained advantageous patterns of social co-operation.[1] In fact, the evolutionary function of moral reasoning was to create shared patterns of moral intuitions, as it was the stable patterns of intuitive affect that allowed humans to implement the beneficial patterns of social co-operation. Furthermore, moral reasoning continues in present society to have the same function as its evolutionary function: in order to sustain co-operative patterns of behaviour, the purpose of moral reasoning is primarily the creation and maintenance of shared patterns of moral affect. It is a mistake to think that moral reasoning properly aims either at the truth or any other truth-related epistemic virtues, such as coherency, plausibility, increased explanatory breadth, or consistency. (To simplify exposition, I'll refer to both truth and members of this truth-related collection of epistemic virtues with: $^{*}truth^{*}$.)

[1] See Haidt and Joseph 2004, Haidt 2007

Philosophers should care about the social intuitionist model of moral judgment because, as its proponents have argued, it has several interesting philosophical implications. For example, supporters of the social intuitionist model claim that their theory vindicates the position that "all ought statements must be grounded, eventually, in an is statement."[2] The same proponents also claim to have shown on scientific grounds that traditional utilitarianism and Kantian ethics are both mistaken, because they "produce psychologically unrealistic systems that most people will reject",[3] that moral scepticism and moral relativism have both been shown to be false, and that engaging in moral philosophy cannot be expected to contribute to moral understanding.[4] The social intuitionists also believe that their theory vindicates an unusual form of projectivism about moral properties. Most philosophical moral projectivists honour the requirement that, even if correspondence moral truth is not actually in the cards, moral reasoning nevertheless should aim to promote aspects of moral $^{*}truth^{*}$. But the social intuitionists disagree. They think that, in fact, it misleading and potentially damaging to think that one of the aims of moral reasoning is the promotion of moral $^{*}truth^{*}$. So, since many people, moral philosophers included, often engage in moral reasoning as if it aims at moral $^{*}truth^{*}$, it is important to determine whether or not the social intuitionist model of moral judgment is correct.

It is this paper's thesis that a credible case for the social intuitionist model has not yet been made. Specifically, I will show that the argument that proponents of the social intuitionist model have offered for the existence of the five moral modules is extremely weak (section 3.3); it is therefore radically premature for proponents of the model to claim that they have "discovered that there

[2]Haidt and Bjorklund 2008a p.214
[3]Haidt and Bjorklund 2008a p.215
[4]Haidt and Bjorklund 2008a p.213-216

are five innate psychological systems upon which cultures build their moral systems."[5] Furthermore, the social intuitionists have adopted an overly simple conception of the model of moral judgment that they presume is held by their opponents. This has allowed them to interpret some experimental results as providing confirmation for the social intuitionist model. But once we have the opportunity to consider what is probably the standard model of moral judgment amongst moral philosophers, we will discover that these experimental results can be easily accommodated by the standard model; a fortiori, the results cannot be used to confirm the social intuitionist model (section 3.4). I will also show that the social intuitionist's argument that a proper aim of moral reasoning is not the promotion of moral *truth* presents no evidence that would change the mind of someone who does think that the *truth* of someone's or some group's moral view is amongst the proper aims of moral reasoning (section 3.4). Deprived of its claims about both the five moral modules, its model of moral judgment, and its claims about the function of moral reasoning, the social intuitionist model falls apart. It cannot be used to sustain either the philosophical implications mentioned above or some further implications that I will discuss below.

3.2 Philosophical Implications of the Social Intuitionist Model

Before turning to a criticism of some of its central ideas, it is important to begin with a more thorough introduction to the social intuitionist model. Because of the complexity of the model and its relevant philosophical implications, I want to do this in stages. First, I want to begin to work through some of the imme-

[5] Haidt 2009

diate implications of the conceptual scheme for moral psychology that has been adopted by proponents of the social intuitionist model. After this, I will describe the model itself and then spell out some of the more significant implications of the model, including its implications about the aims of moral reasoning and the nature of moral inquiry. Finally, I'll draw out some of the differences between the social intuitionist model of moral judgments and, at least amongst moral philosophers, the received and very nearly obviously true model of moral judgment.

Conceptual Scheme – Although proponents of the social intuitionist model believe that their view of moral judgment is "some kind of intuitionism",[6] the model has very little in common with the views of ethical intuitionists like Moore and Ross. One key difference is that, whereas the classical ethical intuitionists were ethical non-naturalists, social intuitionism is quite obviously a kind of ethical naturalism. Indeed, it is far more accurate to think of social intuitionism as a theory that follows in the tradition of Hume, Smith, and other British moral sense theorists. For what the theory holds, basically, is that the five moral modules each take a particular class of stimuli as input and, as output, generate affective states that, in turn, lead individuals to regard the input stimuli as morally salient. The idea is that the human mind is "prepared" to quickly manifest either negative or positive affective responses to actions, events, choices, et cetera, that fall under the categories harm/care, fairness/reciprocity, authority/respect, purity/sanctity, and in-group/out-ground boundaries. As noted above, these quick affective responses are what proponents of the social intuitionist model call moral intuitions; and as the main pro-

[6]Haidt and Bjorklund 2008a p.181

ponent of the social intuitionist model, Jonathan Haidt, writes, "moral intuition is therefore the psychological process that the Scottish philosophers talked about, a process akin to aesthetic judgment: one sees or hears about a [person's action or a] social event and one instantly feels approval or disapproval."[7]

Some of the key concepts in the social intuitionist model have been explicitly defined by Haidt. According to him, moral judgments are "evaluations (good versus bad) of the actions or character of a person that are made with respect to a set of virtue [i.e., values] held by a culture or a subculture to be obligatory."[8] Moral reasoning is defined as "conscious mental activity that consists of transforming given information about people in order to reach a moral judgment."[9] And moral intuitions are defined as "the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion."[10] Note that it follows from these definitions that moral reasoning can never cause moral judgments. However, Haidt, in more recent writings, also distinguishes between a moral intuition and a moral judgment;[11] so the official view, in fact, is that moral intuitions almost always cause moral judgments, and that all moral reasoning is ultimately derived from moral intuitions. Haidt asks, rhetorically, "It is undeniable that people engage in moral reasoning. But does the evidence really show that such reasoning is the *cause* of moral judgment, rather than the consequence?"[12] The answer, according to Haidt and the other social intuitionists, is almost always "no".

[7]Haidt 2001 p.818
[8]Haidt 2001 p.817
[9]Haidt 2001 p.818
[10]Haidt 2001 p.818
[11]See Haidt and Bjorklund 2008a p.217
[12]Haidt 2001 p.817

This conceptual scheme has some important consequences. First of all, the scheme implies that judgments about whether or not a policy or an institution has moral worth are not (surprisingly) moral judgments. Haidt only allows moral judgments to apply to properties that hold of individual persons within the context of social values. He writes, "for example, 'eating a low fat diet' may not qualify as a moral virtue for most philosophers, yet in health-conscious subcultures, people who eat cheeseburgers and milkshakes are seen as morally inferior to those who eat salad and chicken."[13] But the judgment that it is cruel for the state government to allow cheeseburgers on the elementary school lunch menu is not a moral judgment, as it about the actions of an institution. This point will have some consequences for our discussion in section 3.5 of whether we have independent reason to think that moral reasoning aims at and sometimes yields $^{*}truth^{*}$.

Likewise, it follows from the social intuitionist model's conceptual scheme that if someone engages in explicit deliberation, with either herself or with some other people, about whether or not some particular policy or institution is fair or just, she is not (again, surprisingly) engaging in moral reasoning. Moral reasoning, then, does not cover, e.g., empirical investigation into the sort of political arrangements that might make a causal difference towards improving human flourishing, fairness, and/or freedom.

This really is a striking implication. It is not clear why Haidt and the social intuitionists have set up their model of morality so that, for instance, the ideologies of most political movements are not subsumed within the ambit of morality. Most political movements claim moral justifications for their views, after

[13]Haidt 2001 p.817 He cites Stein & Nemeroff 1995

all.[14] But one possible explanation can be found in the social intuitionist's claim that morality evolved by natural selection and that the function of moral reasoning has not significantly changed since morality involved. For, in the EEA[15] no one faced questions like Is capitalism just? or, How can we better realize different kinds of equality in our society? Because of this, evolution would not have had the opportunity to select whatever psychological traits underwrite human moral cognition because these traits were able to answer these questions. Instead, evolution presumably favoured whatever psychological capacities allowed people to engage in sustained patterns of co-operation; presumably all this required was the evaluation of individual behaviours according to a public set of norms. However, it does not follow from this that whatever psychological traits were selected for cannot be used (perhaps with the help of other traits) to attempt to answer questions pertaining to, e.g., the fairness of capitalism. But more on this and related issues in section 3.3.

The Model Itself – For now, I want to continue to examine the details of the social intuitionist model. The most important aspect of the social intuitionist model is that it provides a map of the causal relations that, beginning with

[14]In fact, the social intuitionist have something like a debunking posture towards political ideologies. Haidt and Joseph 2007 argue that the 'culture war' in the U.S. can be explained by opposing political strategies that appeal to different sets of moral intuitions. Liberals appeal to a smaller set of intuitions than conservatives, and the failure of liberals to acknowledge the wider range of intuitions guiding conservative judgments about, e.g., the moral permissibility of abortion explains the persistence of the debate. Thus, according to the social intuitionists, the two main political ideologies in the U.S. are generated by separate and incommensurate clusters of moral affect, and instead of trying to reason out social policy that reflects a fair compromise between the ideologies, the 'war' could presumably be solved by an effective propaganda campaign. However, see Wenz 2009 for an account of the 'culture war' that analyzes it as a rational debate between, not liberal and conservative camps, but 12 different political ideologies.

[15]"EEA" means the environment of evolutionary adaptation. For some phenotypic trait that is an adaptation, there will have been an environment in which individuals in generations of organisms that had this phenotypic were more fit than individuals in generations of organisms that lacked the trait. This environment is the EEA.

moral intuitions, give rise to a moral system within a particular group of people.[16] Here is a reconstruction of the model.

Step 1: Moral judgment begins when a situation elicits in an individual some moral intuition. This affective flash then usually causes a matching moral judgment: a negative intuition leads to a disapproving judgment, and a positive intuition leads to an approving judgment. The individual who forms the judgment will also normally form a belief that the eliciting situation contains something that is wrong or something that is right. However, in cases in which someone already has standing beliefs that are in tension with the moral intuition, the individual may not form the judgment and/or belief that matches her intuition.[17]

Function: The function of the link between moral intuitions and moral judgment seems to be to allow the mind to detect and register the world's moral properties. But it is important to keep in mind that Haidt and other proponents of the social intuitionist model subscribe to a form of moral projectivism. More on this below.

Step 2: After a moral judgment has been formed as the result of an initial moral intuition, an individual may begin to engage in moral reasoning in response to her moral judgment.

Function: Haidt and fellow social intuitionist Fredrik Bjorklund write that once we have formed a moral judgment, "we often feel a need to justify [the judgment] with reasons."[18] And so "we search for arguments that will support

[16] See Haidt and Bjorklund 2008a p.187 or Haidt 2001 fig. 2 for a picture of the model.
[17] From Haidt and Bjorklund 2008a p.187-188
[18] Haidt and Bjorklund 2008a p.189

an already-made judgment."[19] This reasoning is post-hoc. There are no moral reasons prior to or independent of our moral intuitions; there are only those that we invent after the fact in order to try to justify our moral judgments. They continue by saying that the "human tendency to search only for reasons and evidence on one side of a question is so strong and consistent in the research literature that it might be considered the chief obstacle to good thinking."[20] So, the function of individual moral reasoning is not to change moral judgment in responses to new facts or arguments; rather, its function is to provide a post-hoc rationalization for moral judgment. And as we will see in a moment, good moral thinking amounts to nothing more than thinking that produces shared moral affect, whether or not it also promotes any aspect of moral $^{*}truth^{*}$.[21]

Step 3: Reasoned persuasion occurs. An individual who has proceeded through steps 1 and 2 may then try to elicit intuitions in other people that match her own moral judgments. But she need not try to elicit in others explicit awareness of good reasons for moral judgment, and the epistemic virtues (or vices) of her

[19]Haidt and Bjorklund 2008a p.189

[20]Haidt and Bjorklund 2008a p.190. They cite Kuhn 1991, Kunda 1990, and Perkins, Farady, and Bushey 1991. The general thrust of this literature is that people do not readily change their moral point of view. However, Masnick 1999 has results that seem to be inconsistent with the general thrust of the literature cited by Haidt and Bjorklund – that is, she has evidence that people do change their mind on moral issues if presented with good arguments, and given enough time to consider the arguments.

[21]Note that Haidt and Bjorklund assume that moral reasoning cannot genuinely justify moral judgment unless in considers all sides of a question. To philosophers, this will seem odd. Most epistemologists will hold that moral reasoning can genuinely justify moral judgments if it can provide evidence that the judgment is true. However, it is possible to make sense of Haidt and Bjorklund's remarks once we remember that it is their view that the legitimate function of moral reasoning is to create inter-subjective patterns of stable moral affect, not $^{*}truth^{*}$. The idea, then, is that moral reasoning can genuinely justify a judgment if it serves its legitimate function, namely it motivates people to align their moral judgments with widely-shared patterns of moral affect. Of course, in order to accomplish this, a moral reasoner first must find out if members of her group have different moral intuitions than her own; and if some members of her group have different intuitions, that the social intuitionist model implies that it would be appropriate for her to try to find ways to iron out the differences.

moral advocacy are irrelevant. Her goal is simply to use verbal reasoning to align another person's affective responses with her own.

Function: As the social intuitionists acknowledge, it is a misnomer to call this the "reasoned persuasion" link. For, Haidt and Bjorklund write that "it is important to note that 'reasoned persuasion' does not necessarily mean persuasion via logical reasons. The reasons that people give to each other are best seen as attempts to trigger the right intuitions in others."[22] The goal, then, of this reasoning is not "to reach correct conclusions or to create accurate representations of the social world."[23] Instead, it functions to iron-out potential differences in people's moral intuitions. Moral reasons should not be evaluated according to whether or not they promote moral $^{*}truth^{*}$; a good moral reason is one that effectively changes its audience's moral intuitions. The model implies that you would be reasoning improperly – and indeed, that your reasoning would be futile – if you tried to change someone's moral judgment without first attempting to change the moral intuition underlying their moral judgment.

Step 4: Social persuasion occurs. This may happen immediately after a person forms a moral judgment in response to a moral intuition, bypassing any individual-level moral reasoning. Here, someone tries to change someone else's moral intuitions by a strategy or mechanism other than giving reasons.

Function: Social persuasion occurs because humans are "ultrasocial". "Only human beings cooperate widely and intensely with nonkin, and we do it in part through a set of social psychological adaptations that make us extremely sensitive to and influenceable by what other people think and feel... [and] the social

[22]Haidt and Bjorklund 2008a p.191
[23]Haidt and Bjorklund 2008a p.190

persuasion link captures this automatic unconscious influence process."[24] So, the fact that someone forms a particular moral judgment may cause other people in the first person's social group to form the same moral judgments. And the function of this process is to create stable patterns of shared moral affect in some group.

Step 5: Reasoned judgment may occur, less frequently than steps 1-4. "People may at times reason their way to a judgment by sheer force of logic, overriding their initial intuition. In such cases reasoning truly is causal and cannot be said to be the 'slave of passions'."[25] It is also possible for reasoned judgment to conflict with moral intuition – and in this event, "the reasoned judgment may be expressed verbally, yet the intuitive judgment [i.e., the moral intuition] continues to exist under the surface, discoverable by implicit measures such as the Implicit Association Test."[26]

Function: Reasoned judgment results from our intellectual desire "to derive coherent and consistent moral systems by reasoning out from first principles."[27] That is, it occurs as a response to cultural demands for increased moral understanding. However, "when these reasoned moral systems violate people's other moral intuitions, the systems are usually rejected or resisted."[28] So, reasoned judgments exists only because human cultures have an apparently hopeless desire for an epistemically virtuous moral system. Presumably, resisting our moral intuitions in exchange for an epistemically appealing moral system would be something like trying to resist hunger or thirst in exchange for a mas-

[24]Haidt and Bjorklund 2008a p.192-193
[25]Haidt and Bjorklund 2008a p.193
[26]Haidt and Bjorklund 2008a p.193-194
[27]Haidt and Bjorklund 2008a p.194
[28]Haidt and Bjorklund 2008a p.194

terful painting of a meal.

Step 6: Private reflection may occur, less frequently than steps 1-4. This happens when a situation may trigger incompatible moral intuitions, and so long as the individual with the opposing moral intuitions consciously attempts to resolve the conflict, this counts as private reflection. But private reflection may also occur "in those rare cases where a person has no intuition at all (such as on some public policy issues where one simply does not know enough to have an opinion)."[29]

Function: Private reflection seems to function to iron out differences amongst a person's own moral intuitions, or in rare cases, to actually create moral intuitions.

Further Implications – It should be quite clear that this model of the causes and functions of moral judgment has some very surprising implications.

First of all, the view of moral cognition suggested by the model holds that moral cognition is very heavily regulated by moral intuition, and that, as I've said before, the promotion of moral $^{*}truth^{*}$ is not a proper aim of moral reasoning. Indeed, the social intuitionists are upfront about the fact that their models undermines the received view that moral reasoning – both practical and theoretical – can be expected to contribute to moral understanding. Haidt and Bjorklund write, "The model [states] that moral reasoning is less trustworthy than

[29]Haidt and Bjorklund 2008a p.195. Since this suggests that institutional-level properties are not the proper domain of a person's moral intuitions, this looks to be further indication that it is the social intuitionist's view that debates about, e.g., the fairness of a tax policy are not truly moral debates; a fortiori, reasoning about the fairness of a policy is not moral reasoning.

many people think, so [moral] reasoning is not a firm enough foundation upon which to ground a theory – normative or descriptive – of human morality."[30]

This is a very surprising position to adopt. Amongst other things, it implies that there is no point in trying to find the right answer to questions such as, What is a fair way of dividing up our wealth?, How can I be a good person?, Is it in my self-interest to act from just moral motives?, Are the received moral views of my community correct?, What principles, if any, should guide my actions?, Do I have special duties to my family and friends that I don't have to strangers or people that I don't know? or even, Are my moral intuitions appropriate? More practically, the social intuitionist's view of moral reasoning suggests that ordinary moral reasoning that is undertaken because an individual hopes to produce reliable plans of action that will help her acquire things of moral worth, or in an effort to mediate conflicts between her duties and obligations, or in an effort to improve the epistemic virtues of the individual's own moral view is ultimately misguided. It is, at best, a waste of time. For once we understand that moral reasoning does not properly aim at moral $^{*}truth^{*}$, and that its real function is to iron out either tensions in a person's own moral intuitions, or to iron out inter-subjective differences of moral intuition, it no longer seems rational to pursue answers to questions like these, either by means of private reflection or systematic philosophical inquiry.

Of course, Haidt and his colleagues often emphasize that their model allows for a "causal role for moral reasoning",[31] and that it is therefore no threat to human dignity.[32] But they do not seem to understand the criticisms of philoso-

[30]Haidt and Bjorklund 2008a p.216. It is not clear if Haidt et al. used any moral reasoning to come up with their descriptive theory. Presumably this quote implies that they didn't.

[31]Haidt and Bjorklund 2008a p.181

[32]See Haidt and Bjorklund 2008a p.216

phers and psychologists who have pointed out that it is standardly taken to be the truth, plausibility, coherency, consistency, or perceived trustworthiness of a pattern of moral reasoning that influences the production of moral judgment.[33] What social intuitionists deny is that when people are engaged in moral reasoning, they are generally sensitive to how the reasoning can create a body of moral knowledge that instantiates certain epistemic virtues. Social intuitionists see interpersonal moral reasoning as, instead, a brute causal process that succeeds when it creates shared patterns of moral affect in a group of people, whether or not it does this through, for instance, an increase or a decrease in the overall coherence of the group's moral view.

It is because of this that it is misleading – as the social intuitionists themselves acknowledge – to say that the social intuitionist model offers a picture of the role of moral reasoning in human moral cognition. For according to one of its ordinary uses, "reasoning" and its cognates have a normative connection with $^{*}truth^{*}$. The basic idea is that someone is reasoning if amongst the aims of some of their cognitive activity is at least some aspect of $^{*}truth^{*}$. But the social intuitionist's position is that $^{*}truth^{*}$ is never the proper aim of moral reasoning, which in turn implies that strictly speaking humans should never engage in (to use the term with its normative sense here) moral reasoning. So, when talking about the social intuitionist views about moral reasoning, it would be more appropriate to place scare-quotes around the phrase.

On final implication of the social intuition model concerns where it would locate normative ethical theory amongst the other natural sciences. For it follows rather straightforwardly from the social intuitionist's model that normative ethical theory reduces to a branch of applied human social psychology. Normative

[33]See Jacobson 2008 p.222, Haidt and Bjorklund 2008b p.243-244

ethics is not an autonomous discipline that can define its own concepts and methodological principles. Here's why. Suppose that a moral theorist claims to have discovered that the true moral theory says that something is morally permissible if it increases human happiness and morally impermissible if it decreases human happiness. According to the social intuitionist model, the moral theorist cannot hope to convince people of her theory. The only way that she can convince people to believe her theory is if she finds a way of molding their moral intuitions to her view; but she will be unable to do this. The social intuitionist model holds that people have moral intuitions that respond to five categories of (non-moral) properties, and it is obvious that considerations of human happiness will not always align with these categories. For example, an act that is favoured because it elicits approving in-group moral intuitions might decrease human happiness, while an act that is disapproved of because it elicits negative purity moral intuitions might significantly increase human happiness. Of course, it need not be a part of the true moral theory that people are actually able to follow it. But all the same, it seems that instead of trying to discover the one true ethical theory, moral theorists should properly be concerned with discovering what stimuli prompt people's moral intuitions and, in light of this knowledge, finding ways of creating widely-shared and stable patterns in people's moral intuitions. In effect, then, moral inquiry becomes a branch of applied social psychology.

To be clear, this is not the view that moral inquiry needs input from, amongst a variety of other disciplines, applied social psychology. It really is the view that the discipline should be eliminated in favour of the social psychology of moral intuitions. Let me illustrate this point by quoting some of Haidt and Bjorklund's remarks about deontological approaches to ethics. Although they are confused

about the difference between a normative ethical theory and a meta-ethical theory, their meaning is nonetheless plain to see.[34]

> If Greene is correct in his analysis of the psychological origins of deontological statements [i.e., that deontological judgments are really just post-hoc rationalizations of moral emotions], then even metaethical work must be marked with an asterisk, referring down to a particular understanding of human nature and moral psychology. When not properly grounded, entire schools of metaethics can be invalidated by empirical discoveries, as Greene may have done.[35]

So, it seems to be the view of social intuitionists that an ethical theory is not valid unless it is methodologically "grounded" in psychological premises. The social intuitionists hold that there is no such thing as an informed moral judgment. According to them, moral reasoning should be eliminated by applied social psychology.

Social Intuitionist and Projectivism – The model of moral judgment provided by the social intuitionists does not allow moral facts to play a causal role in moral cognition, and it also holds that, properly understood, moral reasoning should not be thought to produce true moral theories. These are two of the reasons why I believe that the social intuitionist model should be officially read as subscribing to a form of projectivism about moral properties.[36] However,

[34]Indeed, the confusion is suggestive: the social intuitionists think that the applied social psychology of moral intuitions is *ethics*, properly conceived. In turn, this might make deontological or consequentialists theories seem like meta-ethical theories, insofar as, if the social intuitionist model is correct, these schools of thought are attempts to provide post-hoc justifications for patterns in people's (or, philosophers) moral intuitions.

[35]Haidt and Bjorklund 2008a p.214-215. They refer to Greene 2008.

[36]I am not here distinguishing between cognitivist and non-cognitivist versions of moral pro-

because of their unique views about what philosophical attitudes we should adopt regarding ordinary moral reasoning, the social intuitionists part ways from some of more traditional and more popular versions of moral projectivism. And since the difference between the projectivism of the social intuitionists and traditional philosophical versions of projectivism is subtle, it is worthwhile spending some time making the difference perspicuous.

First of all, the social intuitionists believe that there is a reflexive relationship between the initial computational routines of the five modules that generate human moral intuitions and the "virtues" (that is, the moral values) of a particular group of humans. The five morality modules are initially set to develop so that certain stimuli (but not moral properties instantiated by the stimuli, since there are none) cause either positive or negative moral intuitions, and cultures build their moral systems as elaborations upon patterns in their moral intuitions. Of course, some cultures may emphasize some moral intuitions and suppress others (by, for instance, eliminating certain classes of stimuli), but there are no virtues or moral values that are not somehow constrained at a distance by the five moral modules.

But moral judgments are not just the expression of intuitive approval or disapproval; the social intuitionists are not emotivists. Haidt and Bjorklund write that "when people make moral claims, they are pointing to moral facts outside of themselves – they intend to say that an act *is in fact wrong*, not just that they disapprove of it."[37] However, their intentions are mistaken. People form these judgments not because they apprehend what moral facts there may be,

jectivism, since this distinction does not matter for understanding the social intuitionist's metaethical position. A properly thorough treatment of moral projectivism should, of course, address this important distinction.

[37] Haidt and Bjorklund 2008a p.214

but because, as we have seen, some situations elicit negative or positive moral intuitions which, in turn, produce judgments that behave from the perspective of the moral agent *as if* they are about moral facts, when in fact they are not. So, it looks so far as though social intuitionism is a fairly standard version of moral projectivism.

However, as I said above, there is important difference between the moral projectivism of the social intuitionists are more familiar philosophical version of moral projectivism. The difference concerns that value or the rationality of engaging in moral reasoning as if one of its proper aims is to promote moral $^{*}truth^{*}$. Many philosophical moral projectivists believe that it is rational for ordinary moral thinkers to engage in moral reasoning as if moral realism is true. However, as we have just seen above, it is an intended implication of the social intuitionist model that people should not engage in moral reasoning as if their reasoning aims at moral $^{*}truth^{*}$. In fact, as we will see more fully in section 3.5 below, the social intuitionists hold that adopting a realist attitude about moral reasoning can make it nearly impossible to achieve some important moral ends. As I said above in section 3.1, it seems to be the social intuitionist's view that realist attitudes (and presumably quasi-realist attitudes too) can be morally damaging.

So, let me try to make the a difference between standard philosophical projectivism and the projectivism of the social intuitionist little more clear. A good place to start is with a statement of a view that I will call *minimal moral realism*.

> *Minimal Moral Realism*: Moral judgments can be true or false, where the operative notion of truth is correspondence truth, and where the moral facts that satisfy the truth-conditions for moral judgments

play a causal role in regulating a person's or a group's moral judgments.

Now, moral projectivists have at least two possible areas of disagreement with a proponent of minimal moral realism. The projectivist can argue that correspondence truth is not the right notion of truth generally, or for moral discourse specifically. Or, a projectivist might hold the view that the minimal moral realist has the wrong ontology because, in fact, there are no moral facts available to causally regulate our moral judgments. Thus, a projectivist may suggest as an alternative to minimal moral realism a view that holds that the situations which elicit approving or disapproving moral sentiments in humans lead us to project moral properties onto the situations that elicit the sentiments. No acts are cruel, so the moral judgment that Sally's actions were cruel does not, strictly speaking, express a moral fact. Rather, because they cause disapproving moral emotions in us, acts are seen as if they were cruel, and so we unknowingly project moral properties like cruelness onto the world.

So, we can distinguish minimal moral realism from minimal moral projectivism, but with an important caveat. We need to break the projectivist's view into two parts. Here's the first part:

> *Minimal Moral Projectivism - Part 1*: Moral judgments are not true or false, but the operations of our moral sentiments makes it seems from the 1^{st} person perspective as if they are true or false.

Now, it is also a common theme in the writing of moral projectivists that it is important that moral judgment and moral discourse both 'earn the right' to be evaluated along realist lines. It is important to be able to interpret moral

judgments *as if* they have truth-conditions, and there is no problem if ordinary moral reasoners continue on in their deliberative practices about moral issues *as if* this reasoning aims at moral $^{*}truth^{*}$. So we find that Simon Blackburn, for instance, holds the view that ordinary moral agents are entitled to, inter alia, to assert propositions like murder is wrong, and reason that, if murder is wrong, then so is allowing someone to die if they could have been saved, or that murder is wrong no matter who does it.[38]

Part of the motivation for this maneuver is surely the Frege-Geach Problem.[39] But another source of motivation may be an awareness that sometimes patterns of moral reasoning conducted by people who believe that their reasoning aims at $^{*}truth^{*}$ can help people realize some of their moral goals. Put another way, appeals to patterns of reasoning about moral questions that are conducted under the assumption that this reasoning aims at the $^{*}truth^{*}$ can sometimes be used to explain people's ability to realize certain moral achievements. Even if a realist attitude about moral discourse is mistaken on philosophical grounds, the fact that ordinary moral thinkers both engage in moral reasoning and often evaluate moral discourse as if moral realism is appropriate may be an extremely important premise in any explanation of how participants in moral discourse are able to achieve some of their moral goals. (I'll offer an analogy in a moment that should further clarify the point I'm developing here.)

[38] C.f. Joyce 2007

[39] Put very crudely, this is the problem that, if moral statement do not have truth-conditions because they function only to express evaluative and/or prescriptive attitudes, it looks as though it follows that "it is wrong to tell lies" and the antecedent of "if it is wrong to tell lies, then it is wrong to ask another person to tell lies on your behalf" should have different meanings. But it seems fine if these two statements as used as premises in an argument the conclusion of which is "Therefore, it is wrong to ask another person to tell lies on your behalf", which in turn suggests that the original statement and the antecedent of the conditional statement have the same meanings, which in turn implies that the emotivist interpretation of moral statements is incorrect.

So, it makes sense to write out the second part of minimal moral projectivism.

> *Minimal Moral Projectivism - Part 2*: Ordinary moral agents are entitled to the view (even though it is mistaken) that their moral judgments can be evaluated for their truth and falsity. This attitude is part of the explanation of how moral reasoning is able to make a causal difference to moral successes and failures.

But it is here that the social intuitionist parts ways with the minimal moral projectivist. For the social intuitionists agree only with part 1 of minimal moral projectivist. And instead of assenting to part 2, they seem to want to augment part 1 with the following view.

> *Social Intuitionism - Part 2*: It is because ordinary moral agents think that moral reasoning aims at moral $^{*}truth^{*}$, and because (as a result of this) ordinary moral agents think that moral judgments can be evaluated for their truth and falsity, that ordinary moral agents fail to achieve many important moral goals. Therefore, ordinary moral agents should not believe that moral reasoning aims at moral $^{*}truth^{*}$ and, consequently, that moral judgments can be evaluated for their truth and falsity.

This makes explicit the difference between the moral projectivism of the social intuitionists and the more traditional philosophical versions of moral projectivism.

But here's an example that I hope clarifies the justification for part 2 of min-

imal moral projectivism. I noted that standard philosophical projectivists can hold that patterns moral reasoning carried out under realist attitudes actually make a positive causal difference to the achievement of certain moral successes, even though the realist attitudes are, speaking philosophically, mistaken. The position here is roughly analogous to some historical patterns of biological reasoning about homologous traits (traits of two different organisms that are similar because they are inherited from a common ancestor). Biologists were able to recognize what we now call homologous traits long before the articulation of modern evolutionary theory because, instead of relying on knowledge of common evolutionary lineage, the biologists instead relied on 'knowledge' of common design. That is, before rise of evolutionary theory, judgments about trait similarity were guided by principles concerning the design choices that a God-like individual would make if she were creating species. These pre-evolutionary-theory biologist's judgments were made *as if* organisms had been designed by a God, and the important point here is that, even though they were mistaken, these guiding principles of were able to make a causal contribution to the biologist's ability to reliably detect what we would now call homologous traits. So too for realist attitudes and principles about moral reasoning: they may be mistaken, but they still may be able to make a positive causal contribution to the achievement of moral successes. But social intuitionists deny this (again, as we will see in section 3.5); whereas I think this position is acceptable to most philosophical moral projectivists.

Alternative Models of Moral Judgment – There is one last point that I want to make by way of introducing the social intuitionist model. Haidt and his fellow social intuitionist's see themselves as operating in a theory-choice space in

which the dominant model of moral judgment is a model that I'll call, following Haidt's terminology,[40] simple-minded rationalism. This model holds that moral reasoning is the only cause of moral judgments. It is important to keep in mind that this means to Haidt *conscious* moral reasoning. So, it follows that the social intuitionist's think the dominant model of moral judgment does not allow some moral judgments to be caused by unconscious moral knowledge. Simple-minded rationalism does not allow moral affect to be the cause of moral judgment either. This is odd, however, because it excludes from the choice space what is in fact nearly everyone's prima facie position, namely the position that we have unconscious (and sometimes fairly reliable) non-modularized moral information, the epistemic virtues of which may be improved by explicit deliberation.

Indeed, in the paper that first introduced the social intuitionist model, Haidt begins by asking, rhetorically, "What model of moral judgment allows a person to know that something was wrong, without knowing why?"[41] Haidt seems to be attributing to the rationalist not only the view that conscious mental states cause all moral judgment, but the plainly absurd view that only conscious mental states that are able to justify the judgment are the cause of moral judgments. All the same, Haidt proceeds to put forward the social intuitionist model as just such a model – that is, the social intuitionist is intended to be (perhaps the *only* scientifically plausible) model of moral judgment that allows a person to form a moral judgment without explicit knowledge of why the judgment is correct. And more importantly, as we will see in section 4, his belief that a number of experiments provide confirmation for the social intuitionist model depends on the assumption that there is no scientifically or philosophically plausible model

[40] See Haidt 2001
[41] Haidt 2001 p.814

of moral judgment available that allows either unconscious moral knowledge or moral affect to cause moral judgments.

But all this shows is that Haidt has overlooked what is the received model of moral judgment, at least amongst moral philosophers. For the received model holds that moral judgments are sometimes caused by either (i) explicit moral reasoning, (ii) unconscious moral knowledge, (iii) conscious moral affect, (iv) moral affect guided by unconscious psychological processes, (v) deference to the judgments of another moral agent, (vi) acceptance of the moral reasoning of another moral agent, or (vii) some combination of the previously mentioned potential causes. According to this model, moral judgments can be caused by motley array of different psychological processes, none of which need be assumed to be fundamental; and like the social intuitionist model, the received model allows unconscious moral knowledge and moral affect to cause some moral judgments. But there are two key differences between the received model and the social intuitionist's model: the received model allows a person's explicit moral reasoning to cause another person's moral judgment directly (i.e. a person does not have to change a person's moral intuitions in order to change their moral judgment); and the received model allows for cases of mixed causation.

There is an important clarification I need to mention. By calling this alternative model the received model of moral judgment amongst moral philosophers, I do not mean to deny that different schools of moral thought will disagree about what reforms, if any, are appropriate. For example, perhaps affect should play a smaller role in our moral reasoning, and reasoning about moral principles a much larger role. Or maybe we should reason less about what justice and duties are, and instead consider more examples of how to maintain and increase

human flourishing. It might be that we should work to align our feelings of sympathy with our moral judgments, and then work to extend our feelings of sympathy to progressively larger groups of people. Or, perhaps the right thing to do is to try habituate into ourselves a suite of admirable affectively-mediated character traits and, then, try to ensure that our moral judgments are aligned with behavioural dispositions that follow from these traits. It could even be that we should try some combination of these different proposals. My point here is that the received model is the common 'point of departure' for these different normative moral psychologies.

This brings us to one last philosophically interesting implication of the social intuitionist model of moral judgment. For we have already seen that moral cognition according to the received model is implemented in a more motley and heterogeneous suite of psycho-social capacities and processes than the social intuitionist model allows. But the view that, even ideally, moral cognition should continue to be implemented in a motley suite of psycho-social capacities and processes is compatible with the received model, and incompatible with the social intuitionist's model. This is a result of the combination of the social intuitionist's view about the proper aims of moral reason with their view about what constitutes the psychological or cognitive centre of human moral cognition. For as I've noted, they think that the later is constituted by five moral modules and that generate affective outputs,[42] and we have seen now that the social intuitionist's think that moral reasoning should only occur when there are differences in the affective outputs generated by these modules. So, ideally, once a group of people have managed to create nearly universally shared moral intu-

[42]Haidt and Bjorklund 2008a p.205 think that modules can spawn numerous other modules, so this claim can also be interpreted as the view that the epistemic centre of human moral cognition is a system of moral module.

itions – a kind of affective consensus – it will not be rational for them to engage in moral reasoning. Moral cognition, then, will ideally be entirely implemented in people's moral modules. For once the moral affects of some group were to come into alignment, there is no point for members of the group to try to use explicit moral deliberation in order to increase either the coherency, consistency, or correspondence of the group's moral view – there is no point for them, that is to say, to try to promote moral **truth** . Of course, this is not to deny that, as a matter of psychological reality, people will not continue to construct elaborate moral codes. My point is that once affective consensus has been achieved in some group, the social intuitionist model implies that there is no point in engaging in this activity.

So, we now have a fairly clear view of both the fundamental commitments of the social intuitionist model of moral judgment as well as some of its more important philosophical implications. It is now time to begin our evaluation of the case that has been made for the social intuitionist model. I will focus on the following three issues: the evidence that the social intuitionists have offered in support of the existence of the five morality modules, the evidence that the social intuitionists have offered on behalf of their model of the causes of moral judgment, and the evidence that the social intuitionist have offered in support of their views about the proper aims of moral reasoning.

3.3 Whence the Five Domains of Moral Cognition?

As I noted above, Haidt believes that social intuitionists have shown that humans have five innate morality modules that generate quick affective evaluations of human actions and choices. So, let us investigate how Haidt and his colleagues achieved this result.

It is useful to separate this investigation into two components. First, I want to see how Haidt and his colleagues arrive at the claim that human moral cognition is about five different categories of moral phenomena. Second, I want to see how Haidt arrives at the view that there are five moral modules that encode affective responses to phenomena falling within these five different categories. I will show that there are problems in the arguments that have been mounted for each of these views.

3.3.1 Five Domains of Morality

To many philosophers, the claim that morality is "about" harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity will be puzzling. The standard view amongst philosophers is that, if morality is "about" anything, it is about the good and the right and the relationship between the two.

However, in order to understand both Haidt's and the other social intuitionist's thinking, it is important to bear in mind that their work is derived from an earlier debate in social psychology and anthropology, between, on one side, the view that human moral systems are defined only by issues of harm, rights, and

justice, and on the other side, the view that each culture defines its own moral code. The debate eventually settled on the difficult issue of how to describe the situations which elicited people's moral judgments.[43] Described one way, a judgment may seem to be entirely about cultural values. Perhaps the gods forbid women from entering the temple, and so people judge that it is impermissible for a woman to enter the temple. Looked at this way, the judgment seems to be derived from a culturally specific moral code. However, upon further investigation, it may turn out that the gods will punish any woman who does enter the temple. Looked at this way, the initial judgment may really be a judgment derived from considerations relating to harm.[44]

Haidt and the other social intuitionists have responded to this debate by stepping back from it and adopting an extremely broad view of morality. As we read in section 3.2, social intuitionists believe that a moral system exists just in case people in some cultural group can be found making evaluative judgments using obligatory values constructed by their cultural group. The social intuitionists do not make any assumptions about the correctness or rationality of any of these judgments or the guiding values. And using this broad conception of morality, Haidt and his fellow social intuitionists believe that they have "discovered" that all human moral judgments cluster around, or are constrained by, five categories of phenomena.

Here is how Haidt summarizes the method by which he and fellow social intuitionist Craig Joseph arrived at the conclusion that there are (probably; more on this in a moment) five different categories of phenomena upon which all of the world's moral systems are constructed:

[43] C.f. Haidt and Bjorklund 2008a p.196
[44] See Shweder et al. 2003 for more on this debate.

> [we] set out to list [the] common principles [of the world's moralities]... by reviewing five works that were rich in detail about moral systems... including Frans de Waal's survey of the roots or precursors of morality in other animals primarily chimpanzees... We... simply listed all the cases where some aspect of the social world was said to trigger approval or disapproval; that is, we tried to list all the things that human beings and chimpanzees seem to value or react to in the behavior of others. We then tried to group the elements that were similar into a smaller number of categories, and finally we counted up the number of works (of out five) that each element appeared in.[45]

I will simply ignore the question of whether the conception of moral judgment implicit in Haidt and Joseph's description of their method is the same as the official definition of a moral judgment we discussed previously in section 3.2. Accordingly, I propose to ignore the difficult question of how Haidt and Joseph were able to determine that the evaluative judgments of the chimpanzees studied by de Waal were derived from a set of obligatory cultural virtues.

Instead, I want to focus for the moment on a different problem with this method. For an obvious worry with this approach is that the categories it produced may be largely conventional, insofar as they reflect only the similarity judgments of Haidt and Joseph and not some underlying psycho-social causal processes. For, if Haidt and Joseph proceeded by first sorting the expressions of approval or disapproval into a large number of groups (and let us suppose that there were 15 such groups) and then attempted to refine this larger num-

[45]Haidt & Bjorklund 2008a p.202-203

ber into a smaller number of groups using their judgments of similarity, then it seems that Haidt and Joseph's method provides as much reason to believe that there are 15 categories of moral phenomena as it provides reason to believe that there are 5 moral categories. Likewise, it is hard to see what prevents us from carrying out this exercise one step further. For example, it seems that the fairness/reciprocity category and the authority/respect category can be collapsed into the more abstract category of norm-abiding/norm-violating, so long as we, following the social intuitionists, do not recognize a strong and universal distinction between social norms and moral norms.[46] Without some additional empirical test of the five categories, there is reason to suspect that five categories reflect only the arbitrary similarity preferences of Haidt and Joseph.

But more on this issue in a moment. First, I also want to call attention to another problem. We have seen now that it is the social intuitionist's view that there are five categories of phenomena that serve as the basis for all of the world's moral systems. And it would seem, given what I have reported so far, that the social intuitionists arrived at this view because all five works in their literature review reported evaluative judgments prompted by the five different categories of phenomena. But this is not what the social intuitionists found. Instead, Haidt and Joseph found that only three of the five categories showed up in all five of the studies they reviewed.

> The winners, showing up clearly in all five works, were harm/care (a sensitivity to or dislike of signs of pain and suffering in others, particularly in the young and vulnerable), fairness/reciprocity (a set of emotional responses related to playing tit-for-tat, such as negative

[46] See Shweder et al. 1987 for more on the problems with this distinction.

> responses to those who fail to repay favors), and authority/respect (a set of concerns about navigating status hierarchies, e.g., anger toward those who fail to display proper signs of deference and respect).[47]

As for the remaining domains,

> There were two additional sets of concerns that were widespread but that had only been mentioned in three or four of the five works: concerns about purity/sanctity (related to the emotion of disgust, necessary for explaining why so many moral rules relate to food, sex, menstruation, and the handling of corpses) and concerns about boundaries between in-group and out-group.[48]

So the second problem here is that, to common sense, this result better supports the hypothesis that there are three categories of phenomena that subsume human moral systems, not five. Indeed, this result might well refute the five category hypothesis. For it seems that Haidt and Joseph are committed to the following proposition: if there are five categories of phenomena upon which human moral cognition (and, presumably, Chimpanzee moral cognition too) is based, then evaluative judgments matching these five categories should be reported in all sufficiently-detailed works of comparative moral anthropology (and primatology). But Haidt and Joseph's results only show that the five categories are reported in either 60% and 80% of the anthropological literature that they review. Assuming the accuracy of Haidt and Joseph's categorization judg-

[47] Haidt and Bjorklund 2008a p.203
[48] Haidt and Bjorklund 2008a p.203

ments, this result provides better support for the view that three, not five, categories regulate moral cognition in humans.

So we have two problems. First, there is the worry that the categories that Haidt and Joseph arrived at are largely conventional, and second, that their results offer better support for the three category hypothesis than the five category hypothesis. This situation really does call for an empirical test. We should find a number of human cultures that, first of all, are at different levels of economic and social development, and secondly, have not been described in the five works that Haidt and Joseph reviewed and then perform an independent analysis of their moral taxonomies. If all these taxonomies yield either the three or the five categories that Haidt and Joseph believe to be both real and universal, this would be evidence in favour of (one of) their views. If not, this would suggest that the conventionality worry is ultimately justified.

However, this test needs to take into account a further complication. When Haidt and Bjorklund refer to the initial publication of the results of the social intuitionist's literature review, they write that

> Haidt and Joseph ... talked about only the first four moral modules, referring to the in-group module only in a footnote stating that there were likely to be many more than four modules. In a subsequent publication (Haidt and Joseph, in press), we realized that in-group concerns were not just a branch of authority concerns and had to be considered equivalent to the first four.[49]

Here is what seems to have happened. Haidt and Joseph studied five contem-

[49] Haidt and Bjorklund 2008 p.217

porary works of comparative human and primate morality, and on the basis of extrapolation from descriptions of judgments approval or disapproval contained in these works, Haidt and Joseph arrived at the view that, first of all, only three "sets of concerns" could capture most of the reports of approval or disapproval mentioned in all five works, and secondly, that one additional category, purity-sanctity, showed up in only some of the works. However, a final category, in-group/out-group, was later created post-hoc, as it were, out of one of the hierarchy category, which in turn created the authority category.

So, the categories themselves shifted after the initial analysis was performed. But this is not the only post-hoc shift to have occurred. It is interesting to note that Haidt and Joseph claim that on the basis of their initial literature review, "The winners, showing up in all five works, were suffering/compassion, reciprocity/fairness, and hierarchy/respect."[50], but four years later, this same literature review is cited by Haidt and Bjorklund as showing, again, that "the winners, showing up clearly in all five works, were harm/care ... , fairness/reciprocity ... , and authority/respect ... ."[51] I submit that a neutral and open-minded reader would regard harm/care and suffering/compassion as two substantially different categories. As I say, the creation of the in-group/out-group category is not the only significant terminological shift that has occurred.

So, it seems that Haidt and his fellow social intuitionists are having a hard time pinning down exactly which categories of phenomena subsume moral cognition. This terminological fluidity is, of course, exactly what you would expect if you thought that the social intuitionist's various moral categories are largely conventional.

[50]Haidt and Joseph 2004 p.58
[51]Haidt & Bjorklund 2008a p.202

And when it comes to the details of the empirical test I proposed earlier, this terminological fluidity shows that we need to take into account a further hypothesis. For we can see now that the social intuitionists actually have three different hypothesis on offer: (a) three categories of phenomena underwrite all forms moral cognition in humans, and they are suffering/compassion, reciprocity/fairness, and hierarchy/respect; (b) three categories of phenomena underwrite all forms moral cognition in humans, and they are harm/care, fairness/reciprocity, and authority/respect; and (c) five categories of phenomena underwrite most forms of moral cognition in humans, and they are harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity.[52] Perhaps (a) and (c) are the most plausible of the bunch, but all the same, Haidt and his colleagues have offered no evidence that would be able to decide between these two.

Thus, Haidt and the other social intuitionists are quite some distance from being able to claim that they have "discovered" that five categories of phenomena subsume all human moral cognition. I have suggested that their own methods and evidence better support the hypothesis that three categories of phenomena subsume human moral cognition. But I have also shown that there is good reason to suspect that categories may be largely convention, insofar as there is no evidence that these categories correspond to any deep features of either human moral psychology or human social practice. So much, then, for the hypothesis that, in humans, morality is "about" harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity.

[52]See Haidt and Joseph 2004 table 1 and Haidt and Joseph 2006 table 1 for the most explicit definitions of the relevant categories that I can detect in the social intuitionist literature. The tables are strikingly different.

3.3.2 Five Moral Modules

It will likely have occurred to some readers by now that the view that there are no moral modules whatsoever is compatible with the hypothesis that all human moral systems share some categories. The reason, quite simply, is that there are many, many alternative explanations – ranging from various shared problems of social co-ordination, various shared political, economic, and religious ideologies and practices, and various shared values and practical interests, shared emotional and cognitive capacities, and implementations of moral systems that have varying degrees of sensitivity to common moral facts – that posit no moral modules and can nevertheless provide explanations of why all human moral systems share a certain number of moral categories. No one holds that the cultural evolution of the various different human moral systems is a completely random process, after all.

What's more, most of these alternative explanations will be compatible with the hypothesis that human morality first evolved through natural selection. This hypothesis functions as something akin to an axiom for the social intuitionists. Haidt writes that "morality... is a major evolutionary adaptation... [that is] built into multiple regions of the brain and body, that is better described as emergent than as learned yet that requires input and shaping from a particular culture. Moral intuitions are therefore both innate and enculturated."[53] So what I want to do here is to show that, even if we assume with Haidt that morality did evolve by natural selection,[54] this provides no reason whatsoever to believe that humans have five (or three) morality modules. The key issue is whether evidence that a trait (or a suite of traits) was favoured by selection shows that

[53] Haidt 2001 p.826
[54] C.f. Haidt 2007

the trait is "built in", that is, innate and relatively non-malleable. But more on this issue in a moment.

First, I want to offer a few comments that serve to better position the social intuitionist model in the extremely large contemporary literature on human moral competence. Haidt has intentionally allied his model with earlier, "first-generation" work in human sociobiology.[55] It was common for first-generation human sociobiologists to posit innate psychological resources as explanations of patterns of behaviour in the EEA, but they generally held that these resources operated in parallel with various non-innate cognitive and cultural achievements. Thus, when applied to moral judgments specifically, the earlier view was that not all moral judgments would be uses of whatever innate psychological resources were involved in establishing moral behaviour in humans. However, Haidt and his colleagues seem to be unaware of the numerous criticisms of first-generation human sociobiology that have been mounted over the last quarter-century. Specifically, they seem to be unaware of the point that it is very plainly a mistake to hold that a psychological disposition established by natural selection has to be developmentally-endogenous.[56] For the social intuitionists go further than first-generation human sociobiology. As we have seen, their proposal is not just that, within the domain of moral cognition, there are some innate psychological resources operating in parallel with broader cognitive and cultural resources, but rather that the epistemic centre of moral cognition is the innate cognitive resources. According to the social intuitionist model, then, nearly all moral judgments are (albeit sometimes indirect) uses of innate psychological resources dedicated to implementing moral cognition.

[55] Compare, e.g., Haidt 2007 with Wilson 2004/1978

[56] See, e.g., Kitcher 1985, Boyd 2001

Now, Haidt and the social intuitionists offer two arguments for the existence of the five modules that they believe underwrite moral cognition. The first argument is this: the hypothesis that there are some moral modules is true because it can explain why children are receptive to some forms of normative instruction and not receptive to others. For instance, the fact that children cannot be taught "to prefer being liked by their peers to being approved of by adults, or to prefer hitting back to loving their enemies."[57] can be explained by the existence of a module that has outputs consistent with hitting back and enjoying the admiration of one's peers over the approval of adults. However, this argument only supports the existence of the putative morality modules if their existence is the best explanation of this phenomena. Of course, Haidt and Bjorklund do not consider any other explanations of this phenomena, and so they can mistakenly be read as suggesting that moral modules are the only – and therefore best – explanation of children's resistance to some forms of moral instruction. But all the same, there are plenty of other explanations available that do not posit the existence of any modules whatever. For, no one believes that the human mind is a blank slate (as Haidt and Bjorklund themselves note),[58] or that the mind is unlimitedly malleable. The debate is about how much of the mind's contents are, at some time, the product of mostly developmentally-endogenous processes (i.e., modules coming online) versus mostly developmentally-exogenous processes (i.e., learning-routines creating more complex information structures). And, importantly, there is no reason to suspect that information that has been acquired from exogenous processes cannot function as a constraint on further development. So, it should be no surprise that children are not malleable with respect to some kind of instruction, and the fact that they are not malleable tells

[57] Haidt and Bjorklund 2008a p.201
[58] Haidt and Bjorklund 2008a p.183

us nothing in particular about the ontogenesis of whatever psychological structures or resources are implementing the non-malleability. Learned information can be as much as a constraint as information encoded in an developmentally-endogenous module.

Haidt's second argument is better. However, it depends on two assumptions. First that there are five categories of moral phenomena subsuming human moral cognition, and secondly, that if a trait is an adaptation, then the trait must be innate and relatively non-malleable. Haidt and Joseph write that "a useful set of terms for analyzing the ways in which such [evolutionary advantageous] abilities [as our ability to, for instance, avoid rotting corpses] get built into minds comes from recent research into the *modularity* of mental functioning."[59] I have already explained why the first of these two assumptions is suspect, and before examining the second of Haidt's arguments for the existence of the five moral modules, I want to say a few words about the implausibility of the second assumption. The point here is that we have no reason to believe that traits or capacities that are adaptive must be "built into minds" in the appropriate sense.

Here's why. First, we need to distinguish between two different types of explanation, *proximate* explanations and *ultimate* explanations. These two types of explanation are commonly recognized by evolutionary biologists as conceptually fundamental to their discipline.[60] These explanations apply to behaviour. Proximate explanations posit psychological mechanisms (modules, learning-routines, developmental pathways, et cetera) that are responsible for causing behaviour. Ultimate explanations, by contrast, identify the evolutionary mech-

[59]Haidt and Joseph 2004 p.59
[60]C.f. Alcock 2001 p.12-13 ; West-Eberhard 2003 p.10

anisms (individual-, kin-, sexual-, group-, or stabilizing-selection, genetic drift, gene flow, et cetera) that are able to explain why a particular pattern of behaviour emerged in the EEA for some organism. More importantly, evolutionary biologists agree that proximate explanations and ultimately explanations are largely methodologically independent. This is because, as a general rule, if an evolutionary biologist is able to identify an ultimate explanation for a pattern of behaviour, there will be many different proximate explanations for the pattern of behaviour that are all compatible with the ultimate explanation. The same goes for proximate explanations: if a scientist is able to identify, say, a developmental pathway that causes a particular pattern of behaviour, then there will normally be a number of different ultimate explanations for the pattern of behaviour that are compatible with the proximate explanation.

Now, not every proximate explanation available for either human or non-human behaviour – either in the EEA or outside of it – posits psychological mechanisms that are innate and relatively non-malleable. The proximate causes of behaviours that solve evolutionary problems need not be built into the mind. For one thing, humans have available to them sophisticated language-using abilities and deliberative skills. So you may well expect that a species that can use language to engage in deliberative reasoning might have been able to use this ability to implement further capacities that were subsequently favoured by evolution, where quite obviously these further capacities are neither innate nor relatively non-malleable. The point, then, is that if we have settled on an ultimate explanation for a pattern of behaviour – say, the implementation of morality, because of group selection[61] – then, by itself, this does not allow us to determine whether or not the psychological capacities responsible for this pattern

[61] See Haidt 2007

of behaviour are innate or relatively non-malleable. We cannot, for instance, determine that the proximate causes of moral behaviour in the EEA involved only language-aided deliberative reasoning. For there is no inference from evidence that some capacity is adaptive to the claim that the psychological mechanisms through which the capacity is implemented are all innate or relatively non-malleable. Evolution need not 'build in' solutions to survival problems.[62]

So, let us turn now to the second argument that Haidt offers for the existence of the moral modules. The argument seems to be basically this: we should believe that there are five modules corresponding to the five categories of phenomena that serve as the foundation for all forms of human moral cognition because evaluative judgments about these categories are human universals, and it is easy to see how it would be evolutionarily advantageous to have five modules whose function is to detect phenomena falling within the five categories. Thus, Haidt and Bjorklund write that "these five sets of intuitions should be seen as the foundations of intuitive ethics. For each one, a clear evolutionary story can be told and has been told many times. We hope nobody will find it controversial to suppose that evolution has prepared the human mind to easily develop a sensitivity to issues related to"[63] the five moral domains.

But in case you do find it controversial, some of the evolutionary stories were told earlier by Haidt and Joseph. For example, they write of the purity/sanctity module that,

> culturally widespread concerns with purity and pollution can be traced to a purity module evolved to deal with the adaptive chal-

[62]Note that this is an exaggerated version of the error committed by first-generation human sociobiologists. See Boyd 2001.

[63]Haidt and Bjorklund 2008a p.203

> lenges of life in a world full of dangerous microbes and parasites. The proper domain of the purity module is the set of things that were associated with these dangers in our evolutionary history, things like rotting corpses, excrement, and scavenger animals... [But] over time, this purity module and its affective output have been elaborated by many cultures into sets of rules... regulating a great many bodily functions and practices, including diet and hygiene. Once norms were in place for such practices, violations of those norms produced negative affective flashes, that is, moral intuitions.[64]

So the inference here is this: humans faced a number of challenges or problems in the environment in which they (or an ancestral species) evolved, and modules like the purity module were selected for because they allowed humans (or an ancestral species) to solve said problems by engaging in certain behaviours such as, e.g., the avoidance of rotting corpses. Of course, the range of stimuli that triggered the modules' output in the EEA for morality may be different than the range of stimuli that triggers the modules' output today, but all the same, we should expect that the modules are still triggered by the same category of stimuli.

This is a fairly standard pattern of inference in evolutionary psychology. But it is important to keep in mind that, as stated, the inference is too quick in a number of different ways. First of all, it is not plausible that the psychological mechanisms that contributed to, say, stable social relations were selected for exclusively because they are able to contribute to stable social relations. We need perception, for example, in order to be able to co-operate with one an-

[64] Haidt and Joseph 2004 p.60

other, or for that matter, to avoid decaying matter. But our perceptual capacities evolved long before they were used to implement patterns of human cooperation. So, even if there are moral modules, their existence does not provide the whole proximate explanation for moral behaviour in the EEA. Likewise, it is not reasonable to think that innate psychological capacities can only be implemented by unconscious individual-level psychological processes. Hunger and thirst have been subject to continued selection, but it does not follow that the way these capacities are implemented in humans does not involve explicit or even social cognition. So, maybe morality is innate; yet, if so, it need not be implemented in only unconscious psychological processes. And more generally, given that humans are able to use both language and reason, it is wildly implausible that whatever moral sentiments we have emerged in a context in which these capacities were not operating. So, one of the default proximate explanations of moral behaviour in the EEA is certainly the view that we used capacities derived from the use of language and reason in combination with our sentimental abilities to implement the relevant patterns of moral behaviour.[65] It is simply naïve to think that whatever psychological mechanisms were used to realize patterns of moral behaviour in the EEA must be modularized, which is to say, innate and relatively non-malleable.

But let us return to Haidt and Joseph's argument. Again, it is basically this: in the EEA humans (or groups of humans) who were able to avoid things like rotting corpses were more fit than humans (or groups of humans) who were unable to do so, which explains why evolution favoured, amongst other kinds

[65] And even then, supposing that the moral sentiments did evolve when these capacities were not yet available, there is no reason now to think that it does not matter that moral cognition in is present form occurs in communities with extremely rich language and extremely complicated cognitive capacities.

of behaviour, rotting-corpse avoidance behaviour.[66] And the social intuitionists think that this behaviour was caused by a purity module. However, this brings us to the fundamental problem in Haidt and his fellow social intuitionist's argument for the existence of both the purity module and other morality modules as well.[67] We have seen now both why you should not assume that there will be only a single proximate explanation available for how an organism was able to implement behaviours that, in turn, were able to solve an evolutionary problem, and we have seen also why you should not assume that whatever mechanisms are posited by be the available proximate explanations must all be innate and relatively non-malleable. Again: traits that are favoured by selection need not be 'built into the mind'. If we now apply these two insights to the social intuitionist's argument for the existence of the morality modules, we can see that is easy to come up with alternative proximate explanations for our target pattern of behaviour – namely, the avoidance of things like rotting corpses in the EEA – that do not posit the existence of a purity module. So, for example, maybe early humans in the EEA had cultural or religious customs that required them to dispose of corpses by burning. Or maybe humans simply learned on the basis of past experience to avoid rotting corpses. Rotting-corpse avoidance behaviour is, after all, a rather simple problem for an organism capable of complex social learning. Or, perhaps humans were able to avoid rotting corpses because some module did emerge, but instead of outputting feelings of disgust, the module outputted feelings and sadness and despair. Or maybe a module emerged that had output with no affective valence whatsoever; it simply caused the appropriate avoidance behaviour more or less automatically. Then again, perhaps

[66] As previously noted, Haidt thinks that morality evolved by way of group-selection. But see Barash 2007 for a rebuttal.

[67] The argument for the other four modules is the same: they exist because they would have been able to solve evolutionary problems faced by humans (or proto-humans). See Haidt and Joseph 2004 p.59 ff.

humans have no moral modules whatsoever – and various different cultural solutions to the problem of rotting corpses were implemented. Maybe some hunter-gatherer groups had social customs that required them to bury corpses, and in other hunter-gatherer groups, the dead were tossed into a swiftly flowing river, while a third group counted on nearby scavenger animals to help them avoid rotting corpses. The important point here is that these alternative proximate explanations are equally compatible with the supposition that humans engaged in rotting-corpse avoidance behaviour in the EEA, and this means that we cannot use the fact that the ability to avoid rotting-corpses was adaptive to determine which of these competing proximate explanations is true.

What's more, Haidt and the social intuitionists have not offered us any evidence that would be able to decide between their view and the competing hypotheses, including both the non-modularized and modularized hypotheses. They are correct, of course, that it could have been a purity module and its affective outputs that caused the advantageous patterns of behaviour in the EEA. But all this means is that the social intuitionists have identified one of the many different possible proximate causes of rotting-corpse avoidance in early humans. The social intuitionists are not entitled to claim that any particular moral modules exist until they can rule out both the competing proximate explanations I've offered and the many other than are easily to come up with for the advantageous patterns of behaviour.

So much, then, for the social intuitionist's argument for the existence of the five (or three) moral modules.

3.4 Experimental Evidence and the Social Intuitionist Model

In addition to their views about the five moral modules, Haidt and the other proponents of the social intuitionist model need their claims about the function of moral reasoning and the primary causes of moral judgments in order to sustain the philosophically relevant implications of their theory. In this section, I will examine the experimental work that social intuitionists have suggested supports their views about both of these matters. However, I'll argue that none of their experimental results pose any trouble for either the received model of moral judgment or for the view that one of the legitimate functions of moral reasoning is the promotion of moral $^{*}truth^{*}$.[68]

So, let me begin by briefly reminding you what the received model of moral judgments holds. This model says that moral judgments are caused by either (i) explicit moral reasoning, (ii) unconscious moral knowledge, (iii) conscious moral affect, (iv) moral affect guided by unconscious psychological processes, (v) deference to the judgments of another moral agent, (vi) acceptance of the moral reasoning of another moral agent, or (vii) some combination of the previously mentioned potential causes.

Now, here are the first four experimental results that impress the social intuitionists. First, some fMRI evidence shows that areas of the brain responsible for affect are implicated in the production of some moral judgments. Second, it is a received view in neuroscience that ordinary patterns of moral cognition cannot be implemented by individuals with significant damage to regions of

[68] I also think that the simple-minded rationalist model of moral judgments is probably consistent with the experimental results produced by the social intuitionists – but since it is likely that no one has ever held the simple-minded model, I do not think it is important to spend time explicitly evaluating it.

the brain that subsume affective responses.[69] So these two results show that, basically, an individual cannot achieve ordinary moral competence if they lack certain affective capacities. Third, experimental subjects have been observed making stronger moral judgments about scenarios that are vividly disgusting compared to the judgments made about scenarios that are not vividly disgusting.[70] And fourth, groups of experimental participants that were hypnotized to feel disgust when they read either the word "take" or the word "often" tended to make more severe moral judgments about morally salient vignettes, and that a third of these subjects responded to a vignette with no apparent moral violation but containing a character whose actions were described using the words "take" and "often" by saying that the actions of the character were "somewhat morally wrong."[71]

Let me speak to the fMRI data first. The fact that moral cognition is partially implemented in people's affective capacities is plainly obvious. Indeed, if the fMRI data did not indicate that sometimes moral judgments implicate affective capacities, we would probably suspect that the machine was broken. It is obvious that the received model of moral judgment is compatible with this evidence.

That said, it is also important to keep in mind that most affective states carry propositional content. For instance, someone can be disgusted, overjoyed, worried, happy, elated, suspicious, jealous, or angry that the man across the street is eating a cheeseburger. I mention this because the social intuitionist's may think – as some other psychologists do[72] – that evidence that people use either af-

[69] Haidt and Bjorklund 2008a p.199
[70] Haidt and Bjorklund 2008a p.199
[71] Haidt and Bjorklund 2008a p.199
[72] See Hauser et al. 2007

fective or unconscious psychological processes to form some moral judgments shows that, more generally, moral reasoning is unreliable. So, the point here is that, since affective states carry propositional content, they can therefore be evaluated as to whether they are able to promote an aspect of moral *truth*. And we often engage in this kind of evaluation. Suppose that a mother is overjoyed that her son has received a job offer. The son may point out to his mother that her joy is inappropriate, because in fact he has received no such offer. A man might realize that his creeping suspicion that his bridge partner is cheating is incompatible with his belief that the partner is an honest person, and then try to reconcile these two attitudes. Someone might think that her joy upon hearing that a distant battle has been won is too superficial a response, and instead be brought to sadness that the battle was even fought in the first place. Indeed, perhaps there are some moral facts that can only be detected by our affective capacities; maybe some forms of tragedy can only be grasped through the appropriate kinds of grief and sorrow. In all of these cases, we can ask if the affective state in question is appropriate, where the relevant standard of evaluation is whether the state promotes or detracts from the some aspect of moral *truth*.

Turning to the fact the ordinary moral competence requires certain affective capacities, it is easy to see that there is no incompatibility between the received view of moral judgment amongst moral philosophers and the experimental work illustrating this fact. One of the reasons why a philosopher will hold the received view is, of course, the sense that certain sentimentally-realized character traits are essential components of the morally-worthwhile life. Someone who is never angered can be readily taken advantage of; someone who is unable to feel sympathy will have difficulty recognizing her duties to kin and comrades; someone who cannot feel guilt will not quickly to learn how to avoid

harmful actions; and so on.

But moreover, the proposition that moral competence requires certain affective capacities might be exactly what you arrive at if you extrapolate from other competencies. It would be very surprising if moral competence did *not* require a range of affective capacities. Consider the following examples of different kinds of competence. Conductors must be emotionally sensitive to how the score will make an audience feel (no one wants to hear an up-tempo and jaunty *E lucevan le stelle*); many scientific fields require of their practitioners that they develop extremely refined emotional capacities (a primatologists needs to understand, amongst other things, what gestures and behaviours annoy the great apes; some physicists and technicians may need to develop a largely tacit and partially affectively mediated relationship with an extremely delicate piece of detection equipment;[73] and most if not all scientists need to become emotionally invested in their own work in order to be sufficiently motivated to pursue it through such difficulties as grad-school, peer-review, grant writing, et cetera); the reliability of a double-blind drug trial might depend in some cases on the emotional capacities of the researchers performing a trial (cool or indifferent researchers may, as a result of depressing the mood of experimental participants, indirectly cause the drug to be less effective); and, lest she cause offense, a good diplomat will of course need to be extremely perceptive of other's emotions (itself a task that is mediated by the mind's affective capacities) and be able to marshal nearly full control of her own affective states and their outward indications. My point: it really is unsurprising that moral competence requires certain affective capacities, since just about every other kind of competency does as well.

Finally, neither the hypnosis study nor the more general experimental evi-

[73] C.f. Sidmondo 2004 p.88-89

dence that people use affective cues to calibrate their deployment of moral concepts provide any reason to suspect that the received model of moral judgment is false. The received model quite obviously allows for people's moral judgments to be caused by different kinds of moral and non-moral affect. Furthermore, since affective states can be evaluated for their ability to promote some aspect of the *truth*, it is useful to keep in mind that proponents of the received model can be reliabalists about these states. Thus, someone who thinks that moral reasoning - by way of, say, advocacy, clarification, or persuasion - functions to promote, amongst other ends, the moral *truth* can hold that most of the time the mind's affective capacities generate judgments that tend towards some aspect of moral *truth*. But all the same, a person who holds these views will also admit that there are situations in which a person's various moral-psychological capacities will not be reliable. For example, perhaps someone's affective capacities are not reliable if, as the result of hypnosis, their affective routines have been 'programmed' to misfire. Indeed, the hypnosis study gets its bite from the fact that it appears to show that people use only their emotional cues to calibrate their moral judgment, and neither their conscious awareness of moral reasons nor their perception of any genuinely moral properties. That said, the study does not show that, under normal circumstances, people's affective cues are not typically *truth* promoting. For it could be that by hypnotizing their experimental participants, Haidt and his team have high-jacked one of the normal – and normally reliable – process by which moral judgments are caused. Compare: Haidt would not consider the social intuitionist model of moral judgment refuted if we conducted an experiment in which the participants were given a post-hypnotic suggestion that required them to blurt out explicit moral justifications before forming any moral judgments.

Let me now turn to the one experiment conducted by social intuitionists that seems to bear only on the question of whether moral **truth** is a proper aim of moral reasoning. Here is a summary of the experimental results in question. When they are presented with a moral dilemma, asked to make a judgment about the dilemma, and then asked to justify or explain the judgment, experimental participants do not "behave like (idealized) scientists, looking for the truth and using reasoning to reach their judgments"; instead, Haidt and other social psychologists have observed that experimental participants usually behave "like lawyers, committed from the start to one side and then search only for evidence to support that side."[74] That is, experimental participants often stick to their initial judgments even if (unlike lawyers, I might add) they are unable to supply explicit reasons for their judgments. Social intuitionists think that this is evidence that moral reasoning does not properly aim at **truth**.

The first point that I want to make concerns a presupposition that guided Haidt and his colleague's experimental design. For Haidt and Bjorklund write that behind their experiments on moral reasoning, "the key question was whether subjects would behave like (idealized) scientists... or whether they would behave like lawyers."[75] So, prior to the experiment, the social intuitionists took the only two plausible theories that someone might hold about the function of people's ordinary moral reasoning to be either the position that people reason like *ideal* scientists or they reason like lawyers. Whereas truth – if not **truth** – is the only aim of ideal scientific reasoning, sticking to an original position and defending it as far as possible is the only aim of legal reasoning. So, the social intuitionists presupposed that either moral reasoning either aims at the truth or its aims at defense, and they think that experimental evidence indi-

[74]Haidt and Bjorklund 2008a p.198
[75]Haidt and Bjorklund 2008a p.198

cating that sometimes people engage in lawyer-like moral reasoning is evidence that the first half of the disjunction is false.

However, the fallacy in this line of inference is obvious. These two hypotheses are clearly not the only scientifically or philosophically plausible views that someone might have about the aims of our ordinary moral reasoning. True, perhaps the importance of many moral problems is a reason to think that people ought to, if they can, reason like ideal scientists about moral problems. But moving past this, the most plausible initial theory to hold would say that ordinary moral reasoning has many different legitimate functions – that include, inter alia, the production and maintenance in an individual reasoner of an epistemically virtuous moral view, the production of islands of moral agreement between an individual and other members of the groups that the individual associates with, and the production of reliable plans, both at the practical and the theoretical level, for achieving things of moral worth. If this 'multiple legitimate function' hypothesis is right, then it is hard to see how evidence that ordinary moral reasoning sometimes does not look as though it is aiming at the truth shows that, all the same, it never does.

But even if we overlook these issues, there are further problems. Here is how Haidt and Bjorklund summarize the experimental results that grounds their view of the function of ordinary moral reasoning:

> The experimenter presented [a moral dilemma to experimental participants and asked them to form a judgment about the dilemma], and then played devil's advocate, arguing against anything the subject said... Results show that [in response to the first dilemma] people did seem to use some reasoning, and they were somewhat respon-

> sive to the counterarguments given by the experimenter... However, responses to [moral and emotionally salient dilemmas were different]: very quick judgment was followed by a search for supporting reasons only; when these reasons were stripped away by the experimenter, few subjects changed their mind, even though many confessed that they could not explain the reasons for their decision.[76]

So, by the social intuitionist's own admission, sometimes moral reasoning does seem to point towards $^{*}truth^{*}$. However, this point seems to be defeated by evidence that people are less able to provide explicit arguments in defense of their more emotional moral judgments. This "puzzled inability to justify a moral conviction" is what social intuitionists call moral dumbfounding.[77]

Let me make two points here. First, if one of the functions of moral reasoning is the production and maintenance of largely accurate moral views, and a moral reasoner thinks that, historically speaking, she has been reasoning well, it would be entirely appropriate for her to resist changing her mind. There is clearly a gap between how much trust Haidt and his fellow researchers invest in the moral judgments of their experimental participants (on theoretical grounds, nearly none), and the amount of trust that their experimental participants seem to be displaying for their own judgments (a fair amount, presumably on practical grounds). Indeed, it seems to follow from the social intuitionists views about what constitutes good moral reasoning (see section 3.1) that, when faced with disagreement, a moral agent should continue to engage in moral reasoning until the disagreement disappears. But if the experimental participant trusts her own judgment, meaning she thinks it is true (or at least close enough), then of course

[76] Haidt and Bjorklund 2008a p.198
[77] Haidt and Bjorklund 2008a p.197; c.f. Haidt and Hersh 2001.

there's no reason for her to keep searching out justification for her judgment – especially if she figures out that she is trying to reason with a sceptic. For what Haidt and his team have discovered is that people sometimes behave as if they trust their moral judgments, even judgments that are caused by either responses guided by either unconscious or affective psychological processes – and keep in mind that both of these processes can be assessed for their reliability or their ability to promote moral $^{*}truth^{*}$. Occasional cases of moral stubbornness are hardly unexpected from the position of someone who thinks that moral reasoning aims at the $^{*}truth^{*}$. Or, put another way, it is appropriate for both the ideal scientist and the ordinary moral reasoner to sometimes reason like a well-paid trial lawyer: namely, when both have acquired (at least something close enough to) the truth.

My second point is that it is pretty easy to find cases of 'scientific dumfounding', and it is hard to see how this by itself suggests that the production of true – or $^{*}true^{*}$ – theories is not amongst the aims of scientific reasoning. Consider: a mathematician may feel something is wrong with a proof long before she can explicitly identify the error, a biologist may be able to accurately sense that a trait of some organism is homologous all the while lacking any conscious awareness of a biological theory that implies what she senses, and an engineer may look at a wall and rightly have sense of dread that the wall is about to fall. The mathematician might need a weekend to check the proof carefully, the biologist might need to look up some details about the evolutionary lineage of the organism that shows that her judgment is justified, and the engineer might need to do a host of difficult calculations. Indeed, imagine that, before any of these individuals have the opportunity to do the subsequent work that shows explicitly why their judgments were right, a sceptic starts to argue with them about their re-

spective judgments. The scientists may find themselves in a position where they are unable to provide the sceptic with explicit reasons supporting their intuitive judgments. Yet in the face of the sceptic's inquiry, all of these scientists may be quite confident in their initial intuitive judgments – and perhaps out of nothing more than a sense that, in the past, their intuitive judgments have been reliable enough. In this case, the scientists would initially be puzzled and unable to explicitly justify their scientific convictions. And while these examples do show that sometimes scientists will justify their judgments after they have been made, it is nevertheless obvious that these examples do not show that the production and maintenance of largely accurate scientific theories is not one of the legitimate functions of scientific reasoning. The fact that a sceptic will sometimes be unable to dissuade a scientist of one of her intuitive judgments does not show that the judgment is unreliable or that, more generally, scientific reasoning does not aim at the truth, amongst other things. Haidt and the social intuitionists are working with a double-standard here.

So much, then, for the experimental evidence that is supposed to show both that moral reasoning does not aim at the $^{*}truth^{*}$ and that the social intuitionist's model of moral judgment is correct.

3.5 Explanations of Moral Disagreement

But there is more going on here. I think we can identify one old philosophical chestnut that has been perhaps only tacitly accepted by the social intuitionists, and I believe that this is part of the explanation of what motivates the social intuitionist model.

The chestnut, as it were, is the view that moral debates are significantly different from scientific debates, and this shows that truth is not an aim of moral reasoning. Haidt writes,

> If moral reasoning is generally a post-hoc construction intended to justify automatic moral intuitions, then ... the bitterness, futility, and self-righteousness of most moral arguments can now be explicated. In a debate on abortion, politics, consensual incest, or on what my friend did to your friend, both sides believe that their positions are based on reasoning about the fact and issues involved. Both sides expect the other side to be responsive to such reasons... When the other side fails to be affected by such good reasons, each side concludes that the other side must be closed-minded or insincere.[78]

It looks as though Haidt's view, then, is that people's mistaken belief that moral reasoning sometimes gets at aspects of moral $^{*}truth^{*}$ is far from benign: this mistaken belief is actually the source of a great deal of suffering. Thus, in order to avoid the suffering, we should give up trying to promote moral $^{*}truth^{*}$; and this consequence explains in turn why, as we saw in section 3.2, it is not right to think that social intuitionism is an ordinary kind of moral projectivism.

Anyway, there are two very similar reasons why this inference is no good. First of all, there are alternative explanations of the persistence and bitterness of moral debate that are incompatible with the social intuitionist model's assessment of the function of moral reasoning. Of course, the persistence of disagreement over moral issues could be explained by the hypothesis that, in fact, there is no fact of the matter or requirement of consistency available to settle the

[78] Haidt 2001 p. 823

disagreement. This seems to be the social intuitionist's position, albeit with the twist that disagreement is ultimately caused by differing moral intuitions. But it can also be explained by the hypothesis that moral inquiry is extremely difficult, or by the hypothesis that moral inquiry and debate is susceptible (in ways that many forms of scientific inquiry are not) to political, corporate, or religious ideologies that may have an interest in perpetuating a debate, or by the hypothesis that all sides share a great many fairly accurate moral beliefs, but that the shared beliefs are not perfectly consistent, and so debate arises over differing proposals about how to resolve the inconsistencies. So the social intuitionist's views about the function of moral reasoning is hardly the only explanation available for the persistence of moral disagreement; and, importantly, Haidt needs to be able to rule out these other explanations before he can use an abductive argument to show that the social intuitionist model is correct about the function of moral reasoning.

The second problem with this inference is that it overlooks the fact that scientific debates can be drawn out, hard to resolve, and extremely bitter. Consider the following examples. Both Micheleson and Morely held onto their respective beliefs in luminiferous ether long after many other physicists and chemists had concluded that their famous experiment showed that no such substance existed. B.F. Skinner evidently stuck to his views in the psychology of language long after they were refuted by Chomsky. But then there's the 60 year old debate, in which neither side has yielded much, between Chomsky and his followers and various groups of Piagetians, connectionists, and non-nativists about the ontogensis of human linguistic knowledge. And then there's the vociferous debate between phyletic gradualists and supporters of punctuated equilibrium in the 1980's, and the quite personal debate that C.V. Raman and Max Borne, both

of whom won the Nobel prize for physics, sustained for nearly 20 years about some technical issues in crystal dynamics.[79] Or, think about the heated and extremely complicated debate about, amongst other issues, whether the definitions in the upcoming DSM-V should be defined through categories derived from clinical consensus (as they are for the DSM-IV) or from categories derived from biological and neurological science. It seems that some members of an older generation of psychologists, psychiatrists, and other mental health professionals have a very strong preference for the clinical consensus route. For these people, this debate could well be resolved by human mortality, not rationality. The point: it is not plausible to think that only, or even the best, explanation of the existence and persistence of these debates is that scientific reasoning is just the post-hoc attempt to justify scientific intuitions. So it is hard to see why the persistence and bitterness of moral debate, but not scientific debate, gives us reason to suppose that the former, but not the latter, does not properly aim at $^{*}truth^{*}$.

Maybe I have misunderstood Haidt's thinking, then. Maybe his view is that the relevant difference between moral debates and scientific debates does not concern the nature of the debate itself, but rather what the debate has the real potential to produce. After all, the construction of extremely accurate, consistent, coherent, explanatorily powerful – in short, very $^{*}truthful^{*}$ – scientific theories does not depend upon these theories winning universal, or even majority, consent amongst all of the relevant scientific experts. So what I am suggesting here is that maybe it is Haidt's view that the relevant difference between moral debates and scientific debates is that moral debates do not produce $^{*}truthful^{*}$ moral theories, whereas scientific theories do, and this shows that the various

[79] C.f. Sur 1999

aspects of moral *truth* are not amongst the proper aims of moral reasoning.

At this point is important to keep in mind that it is either very convenient or very unfortunate that (as we saw in section 3.2) the conceptual framework of the social intuitionist model of moral judgment has been set up so that reasoning about what institutions and social policies to adopt does not count as moral reasoning. For it seems plausible that, persistent moral debates notwithstanding, the most *truthful* moral theories that we have developed concern the nature of the policies, institutions, and procedures that are able to implement and sustain, inter alia, such things as peaceful relations, patterns of equality amongst people of different races and genders, religious tolerance, access to education, the arts, and clean and healthy food, as well as the policies, institutions, and procedures that are able to significantly ameliorate, inter alia, such things as violence, war, poverty, starvation, injustice, and human isolation. All this to say, some of the most *truthful* moral theories are surely our theories about what policies make a causal difference to sustaining flourishing, freedom, and fairness.

However, even if we accept Haidt's definition of moral judgment for the sake of argument, it is not hard to come up with examples that fall within the social intuitionist's conception of morality and nevertheless seem to be perfectly good cases of *truthful*, albeit mundane, moral theories. So, let us suppose that Haidt is right that health-food subcultures have their own moral system, and let us suppose that eating unhealthy food is seen by members of this culture as morally-prohibited because it prompts negative affective intuitions. Members of this health-food culture judge that eating cheeseburgers is morally impermissible because, for members of the culture, the sight of someone eating a cheeseburger automatically causes them to feel a pang of disgust.

Now, it is obvious that the continued existence of this subculture depends on its member's ability to successfully implement plans for practical action that are, at the very least, compatible with whatever ends they see as morally worthwhile. These people must have fairly accurate beliefs about how to put their values into practice lest they, amongst other things, fail to buy healthy food, eat an insufficient number of protein-derived calories per day, buy non-food products from large multi-brand corporations that have policies contrary to the values of the healthy food community, and so on. And since they may also have debates about whether, e.g., eating a certain brand of tofu should elicit the feelings of disgust that it normally does, they may also develop fairly accurate beliefs about what moral intuitions are appropriate. All this to say, it is simply not plausible to think that the practical-moral reasoning, even if it begins with, e.g, a feeling of disgust prompted by the sight of a whole family having their next meal at the local In-and-Out, is just a post-hoc justification of some cluster of affective intuitions. Indeed, even if it is sometimes a post-hoc attempt at providing justification for some initial intuitions, the fact that this reasoning leads in the fullness of time to a network of beliefs that actually makes a causal contribution toward enabling members of the health-food culture to live out a healthy-food life style is very good evidence that these beliefs are largely $^{*}truthful^{*}$, and that therefore providing a post-hoc justification of the original intuition is not the only function of practical-moral reasoning. So, here's my point: practical reasoning about how to achieve things of moral worth can produce ordinary moral views and attendant belief-systems that are largely accurate, and it is very nearly impossible to understand how this could occur if ordinary moral reasoning did not aim at, amongst other things, the $^{*}truth^{*}$.

3.6 Conclusion

I have now completed my evaluation of the evidence that has been offered for the most fundamental tenets of the social intuitionist model, and I want to conclude with the following two observations.

First, both Haidt and the other psychologists involved in what might be termed "the new moral psychology"[80] are motivated by a commitment to naturalism. And yet none of these scientists seriously consider what might be taken to be the first position for a naturalist to occupy. The position that I have in mind is the view that the psychology of moral cognition might turn out to be extremely similar to the psychology of scientific cognition. I find this striking. Instead, the new moral psychologists are fascinated with moral intuitions; the consensus is that moral intuitions are absolutely fundamental aspects of moral cognition. Of course, intuitions play an important role in scientific inquiry, but all the same, no one would ever think that all of a person's scientific knowledge is either implemented in or derived from their scientific intuitions alone.

I am also struck by the gap between the credibility of the evidence that the new moral psychologists have offered for their various theories and the acclaim with which their views have been received. The public does not praise nascent or developing theories in cognitive psychology, political economy, or molecular biology – even when the developing theory has a clear human interest angle. However, the social intuitionist model has been praised as a great and significant achievement, and it is becoming extremely influential. Haidt, for instance, is occasionally treated as a moral luminary; he certainly does not shy away from offering advice to parents and education policy makes about how to properly

[80] See Hauser 2006, Greene 2008, Mikhail 2008, Haidt 2001, Haidt and Bjorklund 2008a

engage in moral pedagogy.[81] But keep in mind that he can't get the difference between a normative ethical theory and a metaethical theory straight.

For any other scientific field, this would simply be annoying. But when shoddy scientific work on morality is received without due criticism and scepticism, it runs the risk of harming people. So, I think that something very weird is going on here. Obvious and common-sensical positions are being overlooked, and obvious and even some elementary problems in theories are going undetected in the peer review process. The ordinary mechanisms of scientific criticism seem to have broken down. One wonders what the explanation for this could be.

[81] Haidt and Joseph 2004 p.64-66

BIBLIOGRAPHY

[1] Alcock, J. (2001) *The triumph of sociobiology*. Oxford University Press.

[2] Barash, D. (2007) Evolution and group selection. In *Science*, 317, 596-597

[3] Boyd, R. (2001) Reference, (In)commensurability and Meanings: Some (Perhaps) Unanticipated Complexities. In Hoyningen-Huene, P. and Sankey, H. (eds.), *Incommensurability and Related Matters*, Boston Studies in the Philosophy of Science, 216.

[4] Greene, J. (2008) The secret joke of Kant's soul. In Sinnott-Armstron, W. (ed.) *Moral Psychology, Volume 2: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. MIT Press.

[5] Haidt, J. (2001) The Emotional Dog and its Rational Tail. In *Psychological Review*, 108(4), 814-834

[6] Haidt, J. (2007) The New Synthesis in Moral Psychology. In *Science*, 316, 998-1002

[7] Haidt, J. (2009) The righteous mind. Retrieved February 14th, 2009 from `http://www.righteousmind.com/`.

[8] Haidt, J., and Bjorklund, F. (2008a) Social Intuitionists Answer Six Questions about Moral Psychology. In Sinnott-Armstron, W. (ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. MIT Press.

[9] Haidt, J. and Bjorklund, F. (2008b) Social intuitionists reason, in conversation – Reply to Jacobson and Narvaez. In Sinnott-Armstron, W. (ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. MIT Press.

[10] Haidt, J. and Craig, J. (2006) The Moral Mind: how five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In Carruthers, P., Laurence, S., and Stich, S. (eds.) *The Innate Mind* vol. 3.

[11] Haidt, J. and Graham, J. (2007) When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98-116

[12] Haidt, J. and Joseph, C. (2004) Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus*, Fall, 55-66

[13] Haidt, J., and Hersh, M. (2001) Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31, 191-221

[14] Hauser, M. (2006) *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*. Harper Collins.

[15] Mikhail, J. (2008) The Poverty of the Moral Stimulus. In Sinnot-Armstrong (ed.) *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness*. MIT Press.

[16] Jacobson, D. (2008) Does social intuitionism flatter morality or challenge it? In Sinnott-Armstron, W. (ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. MIT Press.

[17] Joyce, R. (2007) Projectivism and quasi-realism. In Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*.

[18] Kitcher, P. (1985) *Vaulting Ambition*. MIT Press.

[19] Kuhn, D. (1991) *The skills of argument*. Cambridge University Press.

[20] Kunda, Z. (1990) The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498

[21] Masnick, A. M. (1999) *Belief patterns and the intersection of cognitive and social factors*. PhD Dissertation, Cornell University.

[22] Perkins, D. N. Farady, M. and Bushey, B. (1991) Everyday reasoning and the roots of intelligence. In Voss, J. F. Perkins, D. N. and Segal, J. W. (eds.) *Informal Reasoning and Education*. Erlbaum.

[23] Shweder, R. A. et al. (1987) Culture and moral development. In Kagan, J. and Lamb, S. (eds.) *The emergence of morality in young children*. University of Chicago Press.

[24] Shweder, R. A. et al. (2003). The 'Big Three' of Morality (Autonomy, Community, Divinity) and the 'Big Three' Explanations of Suffering. In Shweder, R. A. *Why Do Men Barbecue?: Recipes for Cultural Psychology*. Harvard University Press.

[25] Sismondo, S. (2004) *An introduction to science and technology studies*. Blackwell Publishing.

[26] Stein, R., and Nemeroff, C. J. (1995) Moral overtones of food: Judgments of others based on what they eat. *Personality and Social Psychology Bulletin*, 21, 480490

[27] Sur, A. (1999) Aesthetics, authority, and control in an Indian laboratory : The Raman-Born controversy on lattice dynamics. *Isis*, 90(1), 25-49

[28] Wenz, P. (2009) *Beyond Red and Blue: How Twelve Political Philosophies Shape American Deabtes*. MIT Press.

[29] West-Eberhard, M.J. (2003) *Developmental Plasticity and Evolution*. Oxford University Press

[30] Wilson, E.O. (2004/1978) *On Human Nature*. Harvard University Press.

CHAPTER 4

MORAL INTUITIONS AND MORAL INQUIRY

4.1 Introduction

Frank Jackson has recently defended a conception of moral inquiry that aims to vindicate what is probably the most common methodological presupposition made about moral intuitions,[1] namely the view that moral intuitions ought to occupy an epistemically privileged evidential role in moral inquiry.

According to Jackson, moral inquiry starts off with conceptual analysis – that is, moral inquiry begins with appeals to moral intuitions. For, by appealing to our moral intuitions and then describing whatever patterns emerge from the appeals, we are, according to Jackson, able to write out an explicit description of the principles that are constitutive of our common sense, or folk, morality. Then, once we have an explicit account of our folk morality, we can subject our account of our folk morality to rational scrutiny, with the aim of improving its epistemic virtues. So, systematic moral inquiry starts off with conceptual analysis, and it continues with the goal of transforming folk morality into what Jackson calls mature folk morality, where the later is "the best we will do by way of making good sense of the raft of sometimes conflicting intuitions about particular cases and general principles that make up current folk morality."[2] Importantly, Jackson thinks that it is unlikely that moral deliberation subsequent

[1] An important clarification: "intuition" here refers to a kind of spontaneous classificatory judgment. This is a different sense of "intuition" than was standardly employed by classical ethical intuitionists. This paper does not consider the viability of either ethical intuitionism or the epistemology of intuitions as they were understood by the ethical intuitionists. See Stratton-Lake 2002, and in particular Sturgeon 2002 and Huemer 2005, for more on the recent interest in ethical intuitionism.

[2] Jackson 1998 p.133

to our intuitive appeals will lead us to give up many of our initial moral intuitions. Instead, it will help us decide "the nature and frequency of [permissible] exceptions"[3] to both our initial moral intuitions and any moral principles that are able to capture them. Thus, once we have an account of our initial moral intuitions, we can then begin to pursue reflective equilibrium, and the particular conception of reflective equilibrium articulated by Jackson construes it as a process largely concerned with both preserving and making sense of our existing moral intuitions. On Jackson's view of it, therefore, moral intuitions occupy an epistemically privileged evidential role in moral inquiry.

It is this paper's thesis that Jackson's conception of moral inquiry should be rejected, as there are two serious problems with it. The first problem is this: the account of the reference of moral language that Jackson relies on in order to explain why moral intuitions deserve a central evidential role in moral inquiry is inconsistent with a growing body of empirical evidence (sections 4.2 and 4.3). The second problem arises from the fact that, since (as Jackson acknowledges) our initial moral judgments are something of a mess, our understanding of how to pursue reflective equilibrium needs to take into account considered judgments about what particular methods of ethical inquiry are most reliable. Since questions about the promotion and stabilization of human flourishing are important to nearly any plausible conception of normative ethical theory (not just, that is to say, to consequentialists – although it is worth noting that Jackson subscribes to some form of consequentialism), examples of the various different mechanisms by which we are able to successfully carry out flourishing-related calculations suggest that the source of our moral intuitions, namely our ordinary linguistic competence with moral terminology, should not be seen as an

[3]Jackson 1998 p.132

especially important source of epistemic access to moral properties (section 4.4). Taken together, the two problems imply that we should reject the view that moral intuitions ought to occupy a special evidential role in moral inquiry, and, in so doing, adopt a more expansive conception of reflective equilibrium than Jackson's own conception (section 4.5).

4.2 Folk Theories and Epistemic Access

This paper's first task, then, is to provide a more detailed introduction to Jackson's views. Specifically, we need to understand the story that Jackson tells in order to support his view that moral intuitions should occupy an epistemically privileged evidential role in moral inquiry.

The story begins with Jackson's philosophy of language. Jackson is often portrayed as a neo-descriptivist. At the heart of his views about language is the idea that to be able to competently use a term, a speaker must possess a fairly rich understanding of the referent of the term, where this understanding consists of some kind of representation of the properties commonly thought to be instantiated by the referent of the term. Jackson also believes that communities of language users will, to the extent to which that are able to engage in communication, share the same understanding. That is, for the term "K" used by speakers in a particular language community, all speakers who are competent users of "K" will share basically the same understanding about what properties Ks commonly have. This shared descriptive guide is what Jackson calls a folk theory.

The folk theory associated with a particular term contributes to determining

the reference of the term in the following way. The folk theory mediating uses of a term such as "water" will hold that water is whatever (if there is such a kind) is found in lakes, flows from the tap, fills our drinking glasses, falls from the sky when it rains, and so on. The term "water" refers to whatever actual kind, if there is any, best satisfies the folk theory. Speakers competent with the term "water" will accept the folk theory associated with "water" and rely on it when using the term. And when it comes to trying to figure out the nature or the essence of water, the folk theory will provide a constraint on this line of inquiry. We know, as the result of a couple of hundred years of research in chemistry, research which was guided by our folk theory of water, that at the actual world, "water" refers to H_2O. This is because we have discovered that H_2O best satisfies the folk theory governing uses of "water". Jackson holds approximately this theory for both philosophically interesting general terms and general terms generally.

According to Jackson, the folk theories that competent speakers use to mediate their uses of words will be largely implicit or tacit. This is why Jackson believes that intuitions are useful, for both philosophical inquiry generally and ethical inquiry more specifically. For intuitions provide the philosopher with a tool that can be used to generate an explicit description of the tacit folk theory guiding someone linguistic community's uses of a philosophically interesting term, and Jackson thinks that there are various philosophical benefits that accrue from explicit knowledge of various different folk theories.[4] Jackson writes, "my intuitions about which possible cases to describe as cases of *K*-hood, to describe using the term "K", reveal my theory of *K*-hood... In as much as my intuitions are shared by the folk, they reveal the folk theory."[5] So, in order to

[4]See Jackson 1998, especially chapter 2.
[5]Jackson 1998 p.37

characterize a folk theory, we are, following Jackson's remarks, to proceed as follows: first we consult our intuitions about a variety of hypothetical cases, then we look for patterns in the intuitions, and finally, we attempt to write down generalizations or principles that capture these patterns. These generalizations and principles will amount to an explication of the implicit folk theory guiding our intuitions.

So, by Jackson's lights, folk theories have three different functions. They provide (a) epistemic access to the phenomena covered by the folk theory, and so they serve as a constraint on inquiry into this phenomena; (b) they ground linguistic competence, insofar as someone who is competent with some term "K" will understand the folk theory governing uses of "K"; and (c) they determine the referent of a term, for according to Jackson, a term governed by a folk theory refers to whatever best satisfies the folk theory.

The three functions of folk theories explain why the method of conceptual analysis will be a fruitful method with which to pursue moral inquiry. By getting clear about the folk theory governing the use of a term like "good", we are some way along to figuring out the essence of the good, since the good will be whatever best satisfies our folk theory of the good, at least once we have completed the process of moral inquiry. Jackson sees this end as being realized in two stages. First, we use the method of conceptual analysis to identify the role that moral terms like "goodness" and "rightness" play in our current linguistic practices. That is, we elucidate the folk theory regulating our use of terms like "rightness" by way of appeal to our moral intuitions. After we have done this, we can then attempt to provide an analysis of the properties which best satisfy the relevant folk theory. Jackson holds that philosophical ethics usually con-

sists of attempts to carry out both of these projects. For example, he writes that "classical utilitarians can be understood as doing two things together. The first is giving an account of the role played by, for example, rightness, and the second is offering us an argument that the property which plays that role is maximizing expected happiness."[6]

But there is an important complication. The folk theory governing our uses of various bits of moral language is evolving. We may learn that it is not right to punish people for certain forms of artistic expression, for instance, or that it is appropriate to prohibit a particular kind of speech act. New moral insights may, over time, alter the folk moral theory that mediates people's use of "right" and other moral terms. So while some present folk moral theories may be able to guide moral inquiry by providing it with a place to start, since further inquiry can result in revisions to some components of our folk moral theory, it seems implausible that folk morality in its current form should provide the descriptive resources that are ultimately used to fix the referents for terms like "good" and "right".

Jackson is sensitive to this point. In view of it, he draws a distinction between folk morality – which he says is "the encapsulation of opinion on moral matters at some point in time" – and mature folk morality – which he describes as "the moral code towards which we are (hopefully) converging (which is not to say that we will ever get there)."[7] When it comes to determining the referents of moral terms, Jackson's view is that the relevant moral folk theory to use is mature folk morality. He writes that, "in my view, being right is having the property that plays the rightness role in mature folk morality, and that rightness

[6]Jackson 2001a p.622
[7]Jackson 2001a p.623

itself is the property that plays that role."[8] So, according to Jackson, the word "right" refers to what ever best satisfies the role that the term plays in mature folk moral theory.[9]

The upshot, then, is that while the method of conceptual analysis applied to moral language cannot yield a description of the folk theory that some kind must satisfy in order to be properly called "the good" or "the right", all the same, the method of conceptual analysis can still reveal the folk theory that currently both governs our uses of "good" and "right" and constrains our inquiry into the nature of the good and the right. In this way, the method of conceptual analysis is united with the method of reflective equilibrium. Jackson writes that the first step in ethical inquiry is to try to come up with a description of the "general principles that we share as the commonplaces or platitudes or constitutive principles that make up the core we need to share in order to count as speaking a common moral language",[10] which is accomplished by conceptual analysis. Then, once we have this description of our folk moral theory, "we modify folk morality under the constraint of reconciling the most compelling general principles with particular judgments,"[11] and we do this with the aim of refining the theory mediating our uses of moral language towards closer and closer approximations of mature folk morality. As I noted in section 4.1, Jackson believes that this process will preserve many of our initial moral intuitions. This is because he thinks that these moral intuitions already provide fairly good insight into the world's moral properties, insofar as they are reflections of some present folk morality. Moral inquiry is therefore unlikely to change many of our moral intuitions. Instead, it will provide us with, as previously noted, "some kind

[8] Jackson 2001a p.623
[9] See Schroeter & Schroeter 2009 p.7-10 for an interesting criticism of this move.
[10] Jackson 1998 p.132
[11] Jackson 1998 p.133

of consensus about the nature and frequency of the exceptions to the general principles that we share"[12], i.e. the general principles that capture the majority of our moral intuitions. So, like many other philosophers, Jackson sees ethical inquiry as the pursuit of reflective equilibrium, starting from the results of conceptual analysis, and in particular, he sees moral intuitions as occupying an epistemically privileged evidential role in the pursuit of reflective equilibrium.

4.3 Empirical Problems for Folk Theories

Jackson's philosophy of language has been criticized on empirical grounds before. In this section, I'll describe a common criticism of Jackson's views that I believe fails, and then argue that a related criticism succeeds.

The failed criticism was first articulated by Stephen Stich and Jonathan Weinberg. They argued that Jackson's views are incompatible with all of theories of concepts considered to be plausible by cognitive scientists. Stich, Weinberg, and Jackson all agree that, in a superficial sense, intuitions derive from an intuitor's ordinary linguistic competence with a particular term, which they in turn understand as being constituted by the intuitor's grasp of some particular concept. But, as Stich and Weinberg write, "Jackson appears to be making [the] assumption that – at least for terms or categories likely to be of interest to philosophers – these intuitions derive from something that can plausibly be called a *theory*."[13] Stich and Weinberg then argue that many researchers in cognitive science hold that intuitions derive from "cognitive structures that are very different from

[12]Jackson 1998 p.133
[13]Stich and Weinberg 2001 p.638

folk theories."[14] The objection, then, is that Jackson presupposes an implausible theory of concepts, by way of his view that ordinary linguistic competence is knowledge of a folk theory.

However, Jackson has an easy reply to this objection. He can refine the superficial sense in which intuitions derive from an intuitor's understanding of a particular concept in the following way. Let "concept" mean something like the prevailing conception in some linguistic community. For example, the concept of a cow is the prevailing conception of cows in a particular linguistic community. There may be, of course, some psychological mechanism(s) by which members of the linguistic community represent their shared concept of cowhood. And of course most cognitive psychologists refer to the psychological representation of the shared concept of cowhood as a speaker's cow concept. But Jackson need not follow this practice. He is not required to hold that the psychological representation of the shared cow concept is itself the concept of a cow, and he can instead identify the prevailing conception of cowhood as the cow concept. This way of applying the notion of a concept seems to be Jackson's preferred way of replying to Stich and Weinberg's objection. For Jackson writes, referring to the explanatory role that the notion of a folk theory plays in his philosophy of language, that "the theory I am talking about concerns what is in common to [for instance] all the cows, not what is in common to all my head states when I use the word "cow" to describe something... No doubt there is something in common in the head as well as in the cows, but what it is is not revealed by intuitions about possible cases."[15] The upshot, then, is that according to Jackson, intuitions derive from an intuitor's understanding of a concept in an indirect way: some shared concept, or folk theory, is encoded somehow in

[14]Stich and Weinberg 2001 p.638
[15]Jackson 2001b p.660

an intuitor's mind, and it is the cognitive scientist's business to figure out what the structure of this representation is. The philosopher's business, by contrast, is to try to learn about the folk theory that governs people intuitions. It does not matter to the philosopher how this theory is encoded psychologically, for just so long as people somehow or other encode the relevant folk theory, or concept, then appeals to intuitions can be useful data about the details of some specific folk theory.[16]

Thus, Jackson has room enough to avoid the charge that his views are incompatible with the various scientifically respectable theories of concepts that Stich and Weinberg have in mind. But there is a related worry that I think poses a real difficulty for his view. The problem, put basically, is this: some recent work in social psychology has shown that people's intuitive judgments do not reflect the kind of stability relative to a common linguistic community that you would expect them to display if they were derived from a folk theory shared by all competent users of certain terms. More exactly: there is empirical evidence that, at least for a variety of philosophically interesting terms, ordinary linguistic competence does *not* consist in knowledge of a folk theory.

Before discussing this research in detail, I want to make it clear that Jackson considers the research relevant to an assessment of his views. He writes,

> I am sometime asked... why, if conceptual analysis is concerned to

[16]Here's a partial analogy to help illustrate this point. Suppose that the one true theory in cognitive science of concepts says that concepts are stereotypes. Now, the federal laws of many countries are conventions that have been approved up by the legislative and executive branch. The body of (in this case legal) conventions is just that – a body of conventions. But someone who understands some of her country's laws will encode some of these conventions in a psychological entity that has the structure of a stereotype, and not the structure of a set of conventions. So too with the folk theory: it is, basically, a body of conventions, and for Jackson's purposes, it doesn't matter what the structure of the psychological entity is that encodes the folk theory.

> elucidate what covers our classificatory practice, don't I advocate doing serious opinion polls on people's responses to various cases? My answer is that I do – when it is necessary. Everyone who presents the Gettier cases to a class of students is doing their own bit of field work, and we all know the answer they get in the vast majority of cases. But it is also true that often we know that our own case is typical and so can generalize from it to others. It was surely not a surprise to Gettier that so many people agreed about his cases.[17]

Jackson clearly thinks that it is important that the intuitions which are to count as evidence within the method of conceptual analysis be widely-shared. Presumably this is because those intuitions that are not widely-shared will not be reflective of the common folk theory understood by members of the relevant linguistic community, and they therefore will not be a source of epistemic access. Likewise, it seems to be Jackson's view that philosophers are typical speakers of the relevant linguistic community, and so their intuitions will be just as representative of the relevant folk theory as anyone else's intuitions.

So let me turn now to the relevant empirical findings that I believe are probably fatal for Jackson's claim that linguistic competence consists in knowledge of a folk theory. These results I have in mind have been produced by philosophers working in a nascent sub-field of social psychology that they call, somewhat oddly given its methods and ambitions,[18] "experimental philosophy". As Joshua Knobe, one of the most sophisticated and influential workers in this field, writes, "Claims about people's intuitions have long played an important role in

[17] Jackson 1998 p.36-37

[18] If it walks like a duck, quacks like a duck, swims like a duck, and it doesn't walk like a mule, bray like a mule, or kick like a mule, and there aren't a bunch of fake ducks around, then it's a duck, not a mule.

philosophical debates. The new field of *experimental philosophy* seeks to subject such claim to rigorous tests using the traditional methods of cognitive science – systematic experimentation and statistical analysis."[19] One standard method followed by some experimental philosophers is this: find a particular intuition that is widely accepted by many philosophers, such as the intuition that, in Gettier-cases, the subject of the example does not have knowledge, or in Gödel-Schidmt cases, that "Gödel" refers to Gödel. Then, conduct an experiment in which experimental participants are presented with various different hypothetical cases designed to resemble the hypothetical cases that philosophers have used to generate their own intuitions, and after considering each case for a few moments, to intuitively judge whether or not Smith knows that he'll get a job, "Gödel" refers to Gödel, and so on. These experiments typically show that non-philosophers normally do not have the same intuitions as philosophers, and that patterns in the experimental participant's intuitions can usually be explained by such factors as SES, ethnic background, educational background, and framing effects.[20]

In more detail, here are some of the more interesting results. In one of the seminal experiments in experimental philosophy, Weinberg, Stich, and Nichols asked their experimental participants to consider the hypothetical scenario in which a boy visiting the local zoo points to an animal in the zebra cage and says, "That's a Zebra". The twist is that, at that particular zoo, the zoo authorities sometimes dress their mules up to look like zebras, and place the disguised mules in the zebra cage. When they presented this scenario to undergraduates at Rutgers University, Weinberg, Stich, and Nichols found that the major-

[19] Knobe 2007 p.81

[20] See Knobe 2007 and Alexander and Weinberg 2007 for useful summaries of most of work done in experimental philosophy to date.

ity (though not all) white experimental participants said that the boy "only believed" that there was a Zebra in the cage, while the majority (but again, not all) of experimental participants of south asian ancestry said that the boy "really knows" that there is a Zebra in the cage.[21] Weinberg, Stich, and Nichols also report experiments in which their participant's epistemic intuitions varied according to the number of philosophy courses taken (typically, participant's sensitivity to sceptical scenarios increased with the amount of education in philosophy received) and their socio-economic status (the results suggest that high-ses individuals are more sensitive to epistemic defeaters that low-ses individuals). Similarly, when Machery, Mallon, Nichols, and Stich presented their experimental participants with a variation on Kripke's famous Gödel-Schidmt example, they found that most participants from Western cultures had the intuition that "Gödel" refers to Gödel, while most participants from East Asian countries had the intuition that "Gödel" refers to the person who, in the hypothetical scenario, really proved the incompleteness result, i.e. Schmidt.[22]

Another interesting result comes from the work of Swain, Alexander, and Weinberg.[23] In this trial, the experimenters wrote out a variation on Keith Lehrer's "truetemp" example, in which Charles unknowingly has his brain altered so that he is always right whenever he estimates the ambient temperature. So, when Charles estimates to himself that it is 71 degrees in the room, he is right. Swain, Alexander, and Weinberg asked their experimental participants to decide using a Likert scale to what extent they agreed with the claim that Charles knows that it is 71 degrees in the room. Basically, what the experimenters found was that, when compared to participants who first considered

[21] c.f. Weinberg, Stich, and Nichols 2001
[22] c.f. Machery, Mallon, Nichols, and Stich 2004
[23] c.f. Swain, Alexander, and Weinberg 2006

the truetemp example, experimental participants who considered an obvious example of knowledge before considering the truetemp case were less likely to say that Charles' estimation of the temperature is a case of knowledge. The experimenters also found that, again when compared to participants who first considered the truetemp example, experimental participants who considered an obvious example of non-knowledge before considering the truetemp case were more likely to say that Charles' estimation of the temperature is a case of knowledge.

Of note, all of these experiments were conducted in English, using experimental participants who were fluent in English, although not all experimental participants were native speakers of English. Most importantly, since the experimenters report definitive responses from all of their experimental participants, it is safe to assume that all of the experimental participants understood the various components of the experiment, which in turn suggests that all experimental participants were competent with the terms they used to express their intuitions. However, the experimental philosophers report a number of quite different systematic patterns in the intuitions of their experimental participants, and that is what is important. For, if intuition's were derived from a widely-shared folk theory that constituted ordinary linguistic competency, then these patterns should not have been observed.

These results suggest that, at least for some philosophically interesting terms, ordinary linguistic competence is not constituted by an understanding of a folk theory. Of course, there is still room for Jackson to argue that linguistic competency with moral terms – unlike ordinary linguistic competence for epistemic and semantic terms – is constituted by knowledge of a folk theory.

But this maneuver looks like it would be ad hoc. There is, therefore, very little reason to think that there are what Jackson calls folk moral theories, and a fortiori very little reason for him to think that moral intuitions should occupy an epistemically privileged role in moral inquiry.

4.4 Other Sources of Epistemic Access to Moral Properties

But more generally, the experimental results do not show that a conception of moral inquiry like Jackson's is incorrect. It could be that, for instance, appeals to our moral intuitions can be used to produce something akin to a consensus moral theory, and that this theory can then be used to tells us what our most fundamental ethical duties are. We can in turn make use of other epistemic factors in order to figure out how to apply the theory. Of course, such an approach to moral inquiry is in need of a new theory of reference, and any potential theory of reference needs to be able to explain why our moral intuitions are a fundamental source of insight into, say, our primary ethical obligations.

Looking past this issue, however, there is a deeper reason to think that our moral intuitions by themselves are unable to ground anything like a general moral theory. For as Jackson acknowledges, our initial moral intuitions are rather quite a mess. And this implies that, amongst various other kinds of judgments, judgments about the sources of the mess – meaning, judgments about the reliability of the various processes by which our moral beliefs are formed – are relevant to moral inquiry. Consider, for instance, trying to figure out whether or not to subscribe to a form of Kantian ethical theory. In order to make an informed decision, someone must first determine (amongst a variety of other

things) whether or not the moral judgments friendly to a Kantian ethical perspective have been produced by reliable processes. The same goes for the decision to adopt or reject a consequentialist ethical theory. The point here is just this: the set of considered judgments from which we begin moral inquiry needs to include, amongst other kinds of judgments, our judgments about the reliability of the various different moral-belief-producing mechanisms.

In fact, there is another reason why we should hold that the factors that should play a role in reflective equilibrium ought to include estimates of the strengths and weaknesses of the various processes and strategies for forming moral beliefs. For, morality is an extremely contentious discipline that can easily be influenced by social ideology. It is therefore important that, to the extent to which it is possible to do so, we accept only those moral beliefs that are formed by the processes we know to be reliable.

It is this section's purpose, therefore, to provide a survey the sources of accurate moral beliefs. Let us say that a mechanism that is able to establish and sustain epistemic access to moral properties is a *morally reliable mechanism*, and in order to calibrate our survey of various potential morally reliable mechanisms, I want to use epistemic access to properties closely associated with human flourishing as the 'test' for whether or not a pattern of inductively fruitful inquiry counts as an instantiation of a morally reliable mechanism. There are a couple of reasons why this particular calibration strategy is attractive. First of all, Jackson is himself a consequentialist.[24] He writes that he "inclines toward the view that [mature folk morality] will turn out to be something along the lines championed by consequentialists."[25] But furthermore, calculations concerning

[24] And for the purposes of this section, I will understand consequentialism to be the view, basically, that our morals should promote human flourishing

[25] Jackson 2001a p.623

the promotion of the various aspects of human flourishing are of central importance to just about any plausible ethical theory. So, even though it is not perfect, epistemic access to human-flourishing-related properties looks to be a pretty fair way of assessing whether or not some particular belief forming process or another counts as a morally reliable mechanism. And we soon see that a number of examples of past morally reliable mechanisms collectively demonstrate that there is a tension between holding the view that calculations dealing with human flourishing are an essential part of moral inquiry, and holding both that the right method for moral inquiry is the method of reflective equilibrium and that the pursuit of reflective equilibrium should be largely guided by trying to make sense of our moral intuitions about possible cases.

So, here is the first (and rather straightforward) example of a morally reliable mechanism. Suppose that we learn of a great tragedy overseas – in a far away country, a natural disaster has wiped out a country's food supply. We think about what it must be like for the thousands of families faced with the prospect of starvation, and we feel sympathy for them. Here, we have recognized an instance of an impending moral disaster, but the recognition is mediated by feelings of sympathy, not our ordinary linguistic competency. Or consider particularly grievous affronts to human dignity. Many kinds of sexual assault are most straightforwardly detected by strong feelings of indignation and outrage. Calm reflection on our moral intuitions is neither an appropriate nor an especially reliable way to detect these kinds of tragedy. In fact, to rely only on ordinary linguistic competence to represent these situations would be to miss many of the morally relevant features of the tragedy. The point: there are many phenomena intimately connected with human flourishing, such that the morally reliable mechanisms that bring moral inquiry into contact with these phenomena

is a suite of affective dispositions and capacities, not mainly ordinary linguistic competence.

Second example. If the operations of sympathy are among the ways of achieving some kinds of epistemic access to moral properties, then the adoption of different forms of cultural organization – such the establishment of, for instance, multicultural communities along with the entrenchment of, for instance, cosmopolitan social policies – can be an indirect source of the relevant kind of epistemic access, insofar as these various forms of cultural organization function to promote and stabilize certain sympathetic capacities. Plausibly, we are better able to care for each other when we live in a society that provides us with the opportunity to appreciate another culture's form of life. And maybe there would be fewer wars if we could better appreciate the profound material and human damage they cause to otherwise distant and anonymous societies, or if we could replace the enthusiasm for nationalism with a stronger enthusiasm for humanism. The point: the adoption of certain forms of cultural organization and their attendant social norms can be an indirect source of the relevant kind of epistemic access, and for this reason, it makes sense to think of some forms of cultural organization as morally reliable mechanisms.

Third and fourth examples. Consider the various cognitive abilities of an individual moral expert. Consider, that is, that various different cognitive skills and social attributes required of someone who is able to reliably identify ways of stabilizing and enhancing human flourishing, at least in a specific context. For example, think about someone whose area of expertise is education theory or contract law. These experts may be very good at identifying specific pedagogical programmes or kinds of contracts that are able to enable many features

of human flourishing, and they might be sensitive to otherwise hard to detect ways in which certain kinds of unfair contracts or commonly designed systems of education can harm people. Of course, the pedagogical expert will not possess the legal expert's ability to identify fair contracts, and the legal expert will not possess the pedagogical expert's ability to implement effective education programmes. Even though they are both participants in the moral project, these kinds of moral expertise are very context specific. But speaking generally, a moral expert will need to know a lot of history, sociology, economics, and psychology, which in turns means that the expert will need to be fluent with a variety of different technical vocabularies. And, to the extent to which her knowledge of, say, past historical and sociological experiments guides her judgments about how to promote flourishing within the specific context that she is focusing on, she will not rely mainly upon her ordinary linguistic competency with moral terms to do so. Rather, it is, amongst other cognitive accomplishments, her competency with a technical vocabulary that allows her to produce the appropriately reliable judgments.[26] Moreover, much of the knowledge that she relies on in order to produce her judgments will consist of explicitly accessible theoretical knowledge. So, here we have examples of two of alternative morally reliable mechanisms: mastery of various technical vocabularies, and expert knowledge. And when it comes the sort of difficult calculations related to human flourishing that would be within the purview of the moral expert, it is hard to see how ordinary linguistic competence with moral terms could function as a comparable morally reliable mechanism.

Fifth example. Certain moral problems pertaining to human flourishing are

[26]Of course, this may include appropriately reliable *intuitive* judgments along with more explicitly-mediated judgments. See Chapter 1 for a view about the conditions under which expert intuitions may be manifest.

too complicated for any one person to solve, and in these cases, we leave it to communities of experts (which are usually composed out of individuals who often work in various different intellectual fields) to work on the problem. Consider, for instance, debates about how to modify the various different kinds of welfare states that currently exist, with the aim of better stabilizing or realizing various important aspects of human flourishing, aspects such as individual liberty, access to quality education and health care, and so on. Here, the debate includes participants from very many different fields of inquiry, and to the extent to which insight is achieved into how to use the welfare state to promote human flourishing, this insight is realized by the collective efforts of the community of experts. And the same can be said about insight into appropriate medical policies covering such issues as blood transfusions and end of life decisions, amongst many others. So, in both of these cases, the morally reliable mechanisms seem to be the deliberative practices of particular communities, and once again, the various moral insights that have been achieved by the relevant communities seem to be insights that we would have been unable to achieve if we relied mainly upon common or ordinary linguistic competence with moral terms.

Sixth example. Sometimes we test moral hypotheses by experiments, broadly construed. For example, suppose we want to figure out if standard academic interviews are fair, or whether unbeknownst to us, there are subtle but common ways in which people's rights may be violated. One way that this hypothesis has been tested is to film a suitably large number of academic interviews and search for different kinds of unconscious sex bias.[27] Or, suppose we want to determine if patients at a local hospital who only speak a foreign

[27] c.f., e.g., Valian 1998

language are more likely to have their rights violated than patients who speak the community's native language. To test this, we might conduct a historical analysis of trends in hospital complaints or treatment outcomes. Alternatively, we might conduct a series of standardized interviews with both native and non-native speakers who have visited the hospital for care. In both of these instances, the morally reliable mechanisms are the experiments themselves. The results of the experiments are, of course, not results that could be provided by relying only on our ordinary linguistic competence.

So, we now have several examples of morally reliable mechanisms that have been inductively fruitful in service of moral inquiry in the past, all of which quite clearly outperform ordinary linguistic competence with moral language. This suggests that moral intuitions should not occupy an epistemically privileged evidential role in moral inquiry, presuming of course that an important part of moral inquiry involves figuring out how to promote human flourishing. But allow me to mention two more examples that put even more pressure on the notion that ordinary linguistic competence should be a primary or significant source of evidence for moral inquiry. I have in mind two examples in which relying on ordinary moral competence would not simply fail to provide the relevant epistemic access; rather, in these examples, such reliance would be profoundly misleading.

Historical and sociological studies often identify particular ideological functions for moral language. To pick perhaps the most straightforward example, it is arguably the case that terms like "justice" and "fairness" have acquired an array of meanings that function to serve the interests of powerful members in society. The effect of these meanings associated with uses of "justice", "fair-

ness" and related terms is to effectively anesthetize large numbers of people against the possibility of recognizing that many aspects of their social world – aspects that, in this community, are called fair and just by competent speakers – are organized in a way that has a profoundly negative impact on their own well-being and potential to flourish. Under these circumstances, simply being a competent speaker of the language of justice and fairness, and, by extension, conducting moral inquiry by relying primarily on moral intuitions derived from the prevailing concepts of justice and fairness, can actually function to prevent the achievement of epistemic access to a certain class of important moral properties. So, relying on ordinary linguistic competence in these circumstances would amount to something close to a profound moral mistake. In this situation, rather than analyzing various moral intuitions, it is more appropriate to instead push forward with an ideological critique that helps to, amongst other ends, jettison the older meanings associated with the terms in question.

On last example. Suppose that a particular community holds a number of extremely prejudicial beliefs concerning the relative equality of men, women, and members of different races and ethnic groups. It is possible that in communities like these the prejudices may be so central to their everyday inferential practices that, over time, various bits of moral language have acquired connotations that reflect these prejudices. The language becomes, in certain contexts at least, too warped to be able to correspond to the relevant aspects of human flourishing. As before, in situations like these, it would be a profound mistake for the community to conduct its moral inquiry by relying upon its moral intuitions. Once again, epistemic access to the relevant moral properties would be achieved by some means other than relying upon ordinary linguistic competence with moral terms.

We have now considered a series of examples that put pressure on the idea that, as Michael DePaul writes and which Jackson endorses, moral "intuitions [should] play a special guiding role in moral inquiry."[28] Again, so long as working out how to promote human flourishing is an important part of moral inquiry, these examples show that there is no reason to think that moral intuitions should occupy a privileged evidential role in moral inquiry.

But before moving on, let me raise the following point. Some of the examples we have just considered also show that one of the more interesting properties of our epistemic access to the world's moral properties is that we borrow this epistemic access from a variety of experts, which suggests that inductively fruitful moral inquiry is especially dependent upon particular forms of social organization, co-ordination, and co-operation. This has at least one interest consequence for our choice of a theory of reference for moral terms. Specifically, this consequence means that we should accept as an alternative to Jackson's neo-descriptivism a theory reference for moral language that can make sense of the idea that, amongst other factors, a social network of experts can help establish, stabilize, and even increase our epistemic access to moral properties. And of course, we do not have to look far to find such a whole family of potential alternatives. The family of theories of reference that I have in mind is, naturally, the naturalistic tradition of causal and causal-descriptivist theories of reference, beginning with the work done in the 1960s and 1970s by Kripke and Putnam and running through the work of Dretske, Devitt, Boyd, Millikan, Psillos, amongst others. Plausibly, any member of this family will do as a suitable starting point for a theory of reference for moral language.

[28] DePaul 2004 p.597

4.5 Moral Intuitions and Reflective Equilibrium

Jackson's conception of reflective equilibrium sees its pursuit as largely concerned with attempting to make sense of our various moral intuitions. But the previous sections shows that this particular way of conducting moral inquiry will overlook many additional sources of evidence. So in this section I want to describe, as an alternative to Jackson's, a conception of reflective equilibrium that assigns moral intuitions a much smaller evidential role. The aim, basically, is to identify some of the ways of refocusing Jackson's conception of reflective equilibrium so that moral intuitions do not dominate its view.

I want to start by noting that there is a way of thinking about moral intuitions and their role in reflective equilibrium that is almost certainly derived from earlier philosophical interest in classical ethical intuitionism. For the method of reflective equilibrium requires us to start moral inquiry with at least some of the beliefs or judgments that we already accept, and there is a history of using "intuitions" to refer to whatever beliefs you start with. Perhaps this is because we cannot always provide an explicit account of why and how we came to accept certain beliefs in the first place; but it is of course a property of intellectual inquiry generally that it can proceed even if we can't provide a full justification of our starting beliefs, concepts, and inferential practices. However, the view that starting with some moral beliefs that we cannot completely justify is not a deep problem for our willingness when doing moral inquiry to rely upon what initial moral beliefs we have does bear some vague resemblance to an older position held by early 20^{th} century ethical intuitionists – namely that certain moral principles are self-justifying and therefore in no need of any explicit justification. But all the same, this use of "intuition" is too loose to carry any methodological

water, and it should be avoided in philosophical discussions of evidence and moral inquiry. For, it is a truism to say that moral inquiry must start with our intuitions, where the sense of "intuition" here means some of the beliefs, moral and otherwise, that we already accept. There is nowhere else for moral inquiry to start out from, after all. But I mention this because there is a more important problem that can arise when the informal use of "intuition" is confused with the more specific sense the term has, namely the sense "intuition" has when it is used to refer to a particular kind of categorization judgment about a possible case mediated by unconscious mental processes. While the loose use of "intuition" has no specific implications with respect to the issue of whether the method of conceptual analysis should or should not be a central part of moral inquiry, the more precise use does. Running together the two uses, therefore, can lead one to, on the basis of accepting a truism, over-estimate the evidential importance of moral intuitions within moral inquiry.

And when we turn our attention to moral intuitions – in the sense of more or less spontaneous judgments about possible cases – there are four similar but ultimately distinct ways in which we can see moral intuitions playing a role in reflective equilibrium without, at the same time, dominating its pursuit. The first two pertain to our understanding of the composition of the initial set of moral judgments we accept as our starting point for ethical inquiry, and the second two pertain to our conception of what sources of evidence to bring to bear on whatever moral theories capture our initial set of moral judgments. Let me explain.

We begin seeking reflective equilibrium with some of the beliefs or judgments that we have reason to accept. But changes in our estimation of the re-

liability of various methods of belief creation and maintenance can of course alter our prior views about what initial beliefs to accept. For example, both the failure of Jackson's neo-descriptivism and the examples in the previous section show that there is little reason to invest moral intuitions specifically with any particular initial evidential privilege, or to think that considered moral judgments about counterfactual cases should be privileged over considered moral judgments about ordinary cases when trying to form a compelling initial set. And more generally, our conception of the best possible starting basis for the pursuit of reflective equilibrium can also be influenced by what ethical theories we have a degree of confidence in. For, the examples in the previous section show that if consequentialism (or, at least something close enough) is the correct moral theory, then our initial starting base seems to involve a very wide array of initial beliefs, some of which will be patently theory-laden (i.e. developed by a history of more or less explicit reasoning) – such as, for instance, our beliefs about issues in political economy. So, the initial set of moral judgments need not be composed of primarily moral intuitions, and the moral intuitions that do make it into our initial set need not be accorded any special epistemic privileged over the other kinds of moral judgment in the set.

Along similar lines, it is important to emphasize the "wide" in "wide reflective equilibrium". Following Rawls,[29] we can distinguish between narrow reflective equilibrium and wide reflective equilibrium. When someone pursues narrow reflective equilibrium, she attempts to formulate a set of principles that are able to capture patterns in her some initial set of moral judgments that she believes to be basically reliable. These judgments are standardly called her considered judgments. But when she asks herself, Should I believe or respect

[29] c.f. Rawls 1971; see also Daniels 1984, Daniels 2003

the moral principles that capture my considered moral judgments?, and subsequently starts to consider the very large number of different arguments and considerations that can be brought to bear on this question, the moral thinker has begun the process of pursuing wide reflective equilibrium.

Now, it is common to describe the shift from narrow to wide reflective equilibrium as what it takes to start doing moral philosophy. Norman Daniels writes, for example, that after determining which moral judgments to treat as our initial base, we then

> propose alternative sets of moral principles which have varying degrees of "fit" with the moral judgments. Rather than settling immediately for the "best fit" of principles with judgments, which would give us only narrow equilibrium, we advance philosophical arguments that reveal the strengths and weaknesses of the competing sets of principles.[30]

But here are two ways of reading this passage. The first interprets this passage as saying that only philosophical arguments are appropriate for assessing the various strengths and weakness of whatever moral theory captures the initial judgments. When this reading is combined with the common view that a standard philosophical argument should start with appeals to intuitions as premises,[31] we arrive as a conception of reflective equilibrium that very closely resembles Jackson's – that is, a conception of wide reflective equilibrium as a method largely driven by appeals to intuition.

However, the second way of reading this passage (which, incidentally, seems

[30] Daniels 1984 p.121

[31] See, e.g., Bealer 1992, Bealer 1996, Bealer 1998, Alexander and Weinberg 2007

to be Daniels intended meaning)[32] does not place any particular importance on the idea that only literally *philosophical* arguments are to be used to reveal the strengths and weaknesses of the initial moral theory. As we have seen in the previous sections, there are more avenues by which we can achieve epistemic access to moral properties than by strictly philosophical reflection. This shows, then, that when we subject our initial set of considered judgments to rational scrutiny, we need not take moral intuitions about possible cases as the primary or most central source of evidence, and of the moral intuitions that we do rely upon when subjecting our considered moral judgments to rational scrutiny, we need not, as before, assign to these moral intuitions any particular epistemic privilege. Indeed, if the general thrust of the argument is section 4.4 is right, then it could very well turn out that moral inquiry properly done is very much like scientific inquiry properly done.[33] Instead of trying to fix up a common and largely a priori folk moral theory, moral theorists instead look as though they proceed from an evidential base composed of a fairly general collection of beliefs, and, using a variety of empirical methods, attempt to improve the epistemic virtues of the initial collection of beliefs.

Of course, this is not to deny that philosophers have an important role to play in moral inquiry. For, we have already seen that estimates of the (im)plausibility of theories of reference for moral terms can make a methodological difference to moral inquiry. But furthermore, to return to an example from the previous section, consider again what is involved in making an informed decision between consequentialist and deontological conceptions of normative

[32]He writes: "I construe these [philosophical] arguments as inferences from relevant background theories (I use the term loosely)." (Daniels 1984 p.121) Presumably the relevant background theories can include for Daniels the kinds of theories generated by the various morally reliable mechanisms canvassed in section 4.4.

[33]c.f. Quine and Ullian 1978

ethical theory. In addition to our judgments about the reliability of the processes from which moral beliefs are derived, we evaluate different conceptions of ethical theory according to their degree of 'fit' with the results of various kinds of philosophical theory. Metaphysical issues related to agency, the nature of human rationality, the plausibility and/or the implications of philosophical theism – all of these issues, amongst a host of others, are relevant to our assessment of the plausibility of various different ethical theories. Ruling in or out alternative conceptions of normative ethical theory goes well beyond moral inquiry itself, and it looks as though philosophers have an important epistemological role to play in this endeavor. So, the various philosophical resources and arguments that play a role in our assessment of ethical theories are an important resource for moral inquiry.

Thus, there are various ways in which someone can think that the method of reflective equilibrium is the appropriate method for moral inquiry and all the same, reject the view that moral intuitions ought to occupy a special or evidentially privileged evidential role in moral inquiry.[34] Specifically, we need not see the initial base of considered moral judgments as composed of primarily or only moral intuitions in the sense of intuitive judgments about the sorts of counterfactual cases important to both Jackson's and many other philosopher's conception of conceptual analysis, and we need not think that, compared to the other moral judgments in our initial base, such moral intuitions deserve an epistemically privileged status. Likewise, amongst the different kinds of evidence that we will rely upon in our attempts to bring our moral judgments into reflective

[34]Incidentally, my aim here is not to enter into the debate between, one on hand, philosophers who believe that moral intuitions should not be assigned any initial evidential weight (see, e.g., Brandt 1979) and those philosophers, like Francis Kamm, who believe that they should (see, e.g., Kamm 2006). I think that there are conditions under which moral intuitions can be, to use the language of section 4.4, morally reliable mechanisms. So I am not a sceptic about moral intuitions, but neither am I an enthusiast.

equilibrium, we need not think that appeals to moral intuitions about possible cases will be either an especially common source of evidence or an evidentially privileged source of evidence.

4.6 Conclusion

This concludes my assessment of Jackson's conception of moral inquiry and, in particular, his views about the role of moral intuitions within moral inquiry. I have argued that there is good reason to believe that ordinary linguistic competence for moral terms is not constituted by a folk morality. This undermines his view that moral intuitions should occupy a central evidential role in moral inquiry because they reveal the contours of the folk theory that, in turn, provides epistemic access to the world's moral properties. And I have also argued that consequentialists like Jackson and their allies both have reason to believe that ordinary linguistic competence with moral terms is not normally a very significant source of epistemic access to the world's moral properties, and so, again, our moral intuitions should not be thought of as a special source of evidence in moral inquiry.

Still, I suppose that it is ultimately a historical question whether or not moral inquiry is best conducted by relying mainly upon moral intuitions, or instead by relying on a much broader array of morally reliable mechanisms. My personal sense is that most of the progress that we have be able to achieve so far has resulted from our reliance on the later, and I see no reason to change.

But suppose that I'm right. Suppose, that is, that in moral inquiry ordinary linguistic competence is not an especially good way of establishing epistemic

access to moral properties. As I've noted, this implies that moral inquiry is quite similar in at least this respect to scientific inquiry. But this line of thought also raises the following issue. It becomes hard to see why, in other areas of philosophical inquiry, it makes sense to rely almost entirely on ordinary linguistic competence. For example, it is hard to see what about metaphysics but not morality makes it the case that, for the former but not the later, our intuitive judgements and their underlying linguistic competencies are reliable enough so as to be able to effectively ground successful inquiry.

BIBLIOGRAPHY

[1] Alexander, J. and Weinberg, J. (2007) Analytic Epistemology and Experimental Philosophy. *Philosophy Compass*, 2(1), 56-80

[2] Bealer, G. (1996) A priori knowledge and the scope of philosophy. *Philosophical Studies* 81, 121-42

[3] Bealer, G. (1992) The incoherence of empiricism. *Proceedings of the Aristotelian Society* 66, 99-138

[4] Bealer, G. (1998) Intuition and the autonomy of philosophy. In DePaul, M. and Ramsey, W. (eds.) *Rethinking Intuition*. Rowman and Littlefield Press.

[5] Brandt, R.B. (1979) *A theory of the good and the right*. Oxford University Press.

[6] Daniels, N. (2003) Reflective equilibrium. IIn Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*.

[7] Daniels, N. (1984) Two approaches to theory acceptance in ethics. In Copp, D. and Zimmerman, D. (eds.) *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*. Rowman and Allanheld Publishers.

[8] DePaul, M. (2004) Intuitions in moral inquiry. In Copp, D. (ed.) *The Oxford Handbook of Ethical Theory*. Oxford University Press.

[9] Huemer, M. (2005) *Ethical intuitionism*. Palgrave MacMillan.

[10] Jackson, F. (1998) *From metaphysics to ethics: a defense of conceptual analysis*. Oxford University Press.

[11] Jackson, F. (2001a) Précis of *From Metaphysics to Ethics*. *Philosophy and Phenomenological Research*, 62(3), 617-624

[12] Jackson, F. (2001b) Reponses. *Philosophy and Phenomenological Research*, 62(3), 653-664

[13] Kamm, F. (2006) In search of the deep structure of morality: an interview with Francis Kamm by Alex Voorhoeve. *Imprints*, 9(2), 93-117

[14] Knobe, J. (2007) Experimental Philosophy. *Philosophy Compass*, 2(1), 81-92

[15] Machery, E., Mallon, R. Nichols, S., and Stich, S., (2004) Semantics, cross-cultural style. *Cognition*, 92(3), B1-B12

[16] Quine, W.V. and Ullian, J.S. (1978) *The web of belief*. McGraw-Hill Humanities/Social Sciences/Languages.

[17] Rawls, J. (1971) *A theory of justice*. Harvard University Press.

[18] Schroeter, S. and Schroeter, F. (2009) A third way in metaethics. *Nous*, XLIII(1), 1-30

[19] Stich, S. and Weinberg, M. (2001) Jackson's Empirical Assumptions. *Philosophy and Phenomenological Research* 62(3), 637-643

[20] Stratton-Lake, P. (2002) *Ethical intuitionism: re-evaluations*. Oxford University Press.

[21] Sturgeon, N. (2002) Ethical intuitionism and ethical naturalism. In Stratton-Lake, P. (ed.) *Ethical intuitionism: re-evaluations*. Oxford University Press.

[22] Swain, S., Alexander, J., and Weinberg, J. (2006) The instability of philosophical intuitions: running hot and cold on truetemp. Presented at *The 32nd Annual Meeting of the Society for Philosophy and Psychology*, Washington University, St. Louis, June.

[23] Valian, V. (1998) *Why so slow? The advancement of women*. M.I.T. Press.

[24] Weinberg, J., Stich, S., and Nichols, S. (2001) Normativity and epistemic intuitions. *Philosophical Topics*, 29, 429-60