Georgia State University
ScholarWorks @ Georgia State University

Philosophy Theses | Department of Philosophy

8-13-2012

The Abstract/Concrete Paradox in Moral Psychology

Shane Reuter
Georgia State University

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

Recommended Citation
Reuter, Shane, "The Abstract/Concrete Paradox in Moral Psychology." Thesis, Georgia State University, 2012.
http://scholarworks.gsu.edu/philosophy_theses/120

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

THE ABSTRACT/CONCRETE PARADOX IN MORAL PSYCHOLOGY

by

SHANE R. REUTER

Under the Direction of Eddy Nahmias

ABSTRACT

The epistemology of intuitions has become popular recently with philosophers' increasing use of experimental methods to study intuitions. Philosophers have focused on the reliability of intuitions, as empirical studies seem to suggest that conflicting intuitions are common. One set of studies, concerning what Sinnott-Armstrong (2008) calls the abstract/concrete paradox, suggests that conflicting intuitions are common and, hence, that mistaken intuitions are common. As Goldman (2007) notes, if mistaken intuitions are sufficiently prevalent, then we might have reason to think intuitions are unreliable. I argue that mistaken intuitions are not common, since studies concerning the abstract/concrete paradox have unknowingly studied several distinct phenomena instead of *the* abstract/concrete paradox and, hence, that they present merely *apparently* conflicting intuitions. I then discuss the import of empirical studies for debates about reliability, noting that those studies can inform us about the *un*reliability of intuitions but we are still unclear about the conditions for reliability.

INDEX WORDS: Philosophy, Reliability, Intuitions

THE ABSTRACT/CONCRETE PARADOX IN MORAL PSYCHOLOGY

by

SHANE R. REUTER

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

Master of Arts

in the College of Arts and Sciences

Georgia State University

2012

Copyright by
Shane Robert Reuter
2012

THE ABSTRACT/CONCRETE PARADOX IN MORAL PSYCHOLOGY

by

SHANE R. REUTER

Committee Chair: Eddy Nahmias

Committee: Andrea Scarantino

Dan Weiskopf

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2012

ACKNOWLEDGEMENTS

I would like to thank the members of my committee, Eddy Nahmias, Dan Weiskopf, and Andrea Scarantino, for many helpful (and some brutal) comments and discussions. Many thanks to Eddy, in particular, for continued guidance and feedback over the past year. Shout out also to Vince Abruzzo and Billy Brady. Y'all were awesome last year. Thanks for tolerating all my weirdness, my vehement opposition of most of your dietary practices, and my criticism of your squat and deadlift technique. Green House will always live on in spirit. (And I'll smoke you both in a push-up competition next time we meet.) Finally, I would like to thank the Philosophy Department for all the support and resources, and for creating and maintaining a great place for doing philosophy.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

LIST OF FIGURES vi

1 INTUITIONS AND RELIABILITY 1

2 THE ABSTRACT/CONCRETE PARADOX 5

3 POSSIBLE EXPLANATIONS OF THE ABSTRACT/CONCRETE PARADOX 13

4 EVALUATION OF EXPLANATIONS OF THE PARADOX 16

5 GENERAL PROBLEMS WITH THE ABSTRACT/CONCRETE LITERATURE 21

6 EXPERIMENTS AND RELIABILITY OF INTUITION 24

7 CONCLUSION 28

REFERENCES 29

LIST OF FIGURES

Figure 1.....23

1 INTUITIONS AND RELIABILITY

Intuitions have received a fair amount of attention in the philosophical literature recently. An increasing interest in intuitions should not be surprising, since philosophers have long appealed to what seems intuitively correct as a source of evidence. More specifically, intuitions are often thought of as a source of foundational evidence—as Alexander and Weinberg note, "intuitions provide a noninferential, defeasible justificatory foundation in at least the following way: a person may appeal to an intuition as evidence *without having to provide further evidence for the intuition*" (2012; p. 2; my emphasis). Thus, a claim that is intuitive has *prima facie* support, while we have *prima facie* support against counterintuitive claims (see, e.g., Goldman 1986, p. 66). For examples, appeals to intuitions can be found in debates about utilitarianism and consequentialism (Thompson 1971; Singer 1972), epistemology (Gettier 1963), personal identity (Parfit 1986), the relation of the mind and brain (Chalmers 1996; Jackson 1982; 1986), and free will and moral responsibility (Frankfurt 1969; van Inwagen 1983). Some philosophers even consider appeal to intuition to be distinctive of philosophical methodology (Jackson 1998; Goldman 2007; 2010).

It is clear that, as Alvin Goldman says, intuitions play a "critical role" in philosophical practice (2007, p. 1). Recently, however, interest in the *epistemology* of intuition itself has grown. The increasing popularity of the epistemology of intuition can be loosely traced through early work of Stich (1988; 1990) and the papers in the 1998 *Rethinking Intuition* anthology. Most recently the epistemology of intuitions has been widely discussed within some areas of the experimental philosophy program, wherein philosophers conduct empirical studies surveying people's intuitions about various cases and thought experiments concerned with philosophical issues.

Many issues concerning the epistemology of intuition have arisen throughout these works. For example, some theorists have focused on the nature of intuitions: What is the content of intuitions (Sosa 2006; Lynch 2006)? Are there different kinds of intuitions (Bealer 1998; Jackson 1998)? Other theorists have asked about the source and status of intuitions: Is there a distinct intuitive faculty (e.g., Hales 2012)? What is the epistemic status of intuitions (Pust 2000; Alexander and Weinberg 2007; Goldman 2007)? Are intuitions reliable (Goldman 2007; 2010; Liao 2008)? This latter group of questions has been addressed in detail lately by proponents of the "negative" branch of the experimental philosophy program. Proponents of the negative branch[1] believe that recent experimental findings undermine the epistemic status of intuitions by showing that they are unstable and overly sensitive to seemingly irrelevant factors such as the socioeconomic status of the intuiter (Weinberg, Nichols, and Stich 2001), presentation order of cases (Swain, Alexander, and Weinberg 2008), and the affective content of the intuition-eliciting case (Nichols and Knobe 2007). Based upon recent experimental findings, proponents of the negative branch conclude that we have reason to doubt the epistemic status of intuitions (Weinberg, Nichols, and Stich 2001; Alexander and Weinberg 2007; Swain, Alexander, and Weinberg 2008).

An important issue that has received much attention recently within the negative branch of experimental philosophy and in the epistemology of intuition more generally is *reliability*. We need to know if our intuitions are reliable or not. This is important for philosophers in many different areas of philosophy, since intuitions are commonly appealed to in many different areas. Furthermore, debates about reliability have been reinvigorated recently as more and more experimental studies of intuitions about philosophical issues have been conducted. In later sec-

[1] The negative branch of experimental philosophy can be contrasted with the positive branch, wherein experimental philosophy is viewed as a complement to armchair philosophical methodology, not as a replacement of armchair methodology. See Alexander and Weinberg (2007) and Alexander, Mallon, and Weinberg (2010) for a fuller discussion of the differences between the positive and negative programs of experimental philosophy. See also Nadelhoffer and Nahmias (2007) for discussion of three distinct projects within experimental philosophy.

tions of the paper I will discuss some of those studies and what they tell us about the reliability of intuitions. It will be helpful to briefly review challenges to reliability first.

According to Goldman (2007), there are two general types of challenges to the reliability of intuition. First, skeptics about intuition have argued that since people frequently have conflicting intuitions—and one of the two conflicting intuitions must be wrong—the percentage of intuitions that are wrong is unreasonably high such that we cannot claim that intuition is reliable. Second, intuitions cannot be used as evidence unless they have been calibrated. According to Weinberg, Crowley, Gonnerman, and Vandewalker (2012), "Calibration is a process of regulating a putative source of evidence, by inspecting it and, if needed, adjusting it to render it accurate" (260). The process of calibration involves testing the putative source of evidence and corroborating it against a trusted, independent procedure or source of evidence. However, according to Cummins (1998), we have no such independent procedure with which to corroborate intuitions.[2] Thus, it is impossible to calibrate intuitions and, hence, we cannot know whether intuitions are reliable or not (Cummins 1998).

The calibration challenge to the reliability of intuition can be blocked—that is, shown to be only an apparent challenge while not in fact challenging reliability—by appealing to the analogous case of memory as a putative source of evidence (Alston 1993; Goldman 2007). Memory is our basic way of forming true beliefs about the past, and all other ways of accessing facts about the past depend upon memory. Thus, we have no independent procedure for calibrating memory. But claiming that memory is unreliable risks general skepticism, so a lack of independent procedure for calibration cannot be used to challenge the reliability of intuition. Some sources of evidence may not require calibration, and intuition may be one of those sources (Goldman 2007, p. 5). Instead, Goldman suggests that we substitute independent calibration for

[2] As Cummins notes, if such an independent procedure did exist, there would be no use for intuitions. For if we had a trusted, non-intuitive procedure to calibrate intuitions against, we would already have access to the relevant facts of the domain for which intuitions are intended to be used (1998, p. 117-18).

a weaker condition on something to serve as evidence, namely, "that we *not* be justified in believing that the putative source is *un*reliable" (2007, p. 5; emphasis original).

With the calibration challenge blocked, the best challenge available left to skeptics of the reliability of intuitions concerns conflicting intuitions. If there are conflicting intuitions about some phenomenon—i.e., two individuals[3] each having a single intuition that conflicts with the other person's intuition about the same phenomenon—then one of those intuitions must be wrong (Goldman 2007).[4] Thus, if conflicting intuitions are common, then wrong intuitions are common as well. This challenge to the reliability of intuition therefore turns on an empirical question—are conflicting intuitions prevalent?[5] A positive answer favors the skeptic about the reliability of intuitions. A negative answer favors the proponent of appeals to intuition as evidence. To address this challenge, we must look at empirical studies of intuitions about hypothetical cases concerning philosophical issues in order to answer whether conflicting intuitions are prevalent or not.

In the next section I will review a series of empirical studies looking at people's judgments of moral responsibility, free will, and related concepts. This particular set of studies will be helpful for discussions about reliability since these studies have uncovered what appears to be a large set of conflicting intuitions. If conflicting intuitions are as prevalent as these studies seem to suggest, then the challenge to reliability of intuitions from conflicting intuitions seems to be

---

[3] We can distinguish conflicting intuitions between two individuals and conflicting intuitions within the same individual. In the latter case, for example, someone might think both that abortion is permissible and that abortion is not permissible (Liao 2008). Since the experimental studies I discuss later in the paper all concern comparisons of intuitions between different individuals, my discussion hereafter will not be targeted at conflicting intuitions within the same individual.

[4] Here "conflicting intuitions" refer to what Liao calls *Genuine Conflicts*—a conflict that "seem to remain after possible verbal disagreements and considerations of partiality, clouded emotions, mistaken heuristics, and so on, have been taken into account" (2008; p. 259). In the case of a Genuine Conflict, two individuals possess the same facts about some phenomenon but appear to interpret those facts in different ways; there are no additional facts to consider. Intuitions that seemingly conflict but are the result of factors such as partiality or verbal disagreements do not constitute a Genuine Conflict but only an *Apparent Conflict* (Liao 2008; p. 259). Only in the case of a Genuine Conflict *must* one of the conflicting intuitions be mistaken. Throughout the rest of the paper, I will use "conflicting intuitions" in the sense of Liao's Genuine Conflicts.

[5] It is unclear how prevalent conflicting intuitions need to be in order for this challenge to be a serious threat to the reliability of intuitions. Ultimately, my discussion will suggest that conflicting intuitions are not very prevalent at all in the recent experimental philosophy literature, and I will not try to say how prevalent conflicting intuitions would been to be for this challenge to reliability to succeed.

correct. However, I will argue that the studies have not uncovered conflicting intuitions, since these studies have involved several different phenomena. Thus, the conflicting intuitions of these studies are only *apparent* conflicts and, therefore, these studies have not shown that intuitions are unreliable.[6] The upshot here is important for all philosophers interested in issues concerning the reliability of intuitions. For to show that a set of intuitions are unreliable, we must show that those intuitions constitute genuine conflicts, where two individuals possess the same facts of the matter but interpret them in different ways. But if two or more studies purporting to study intuitions about the same phenomenon actually look at intuitions about different phenomena, then there can be no genuine conflicts. Since experimental studies play an increasingly larger role in the epistemology of intuitions, philosophers must be increasingly careful in constructing survey materials and interpreting the results of those surveys to ensure that the intuitions studied concern the targeted phenomenon.

2 THE ABSTRACT/CONCRETE PARADOX

In this section I will review of handful of recent studies concerning non-philosophers' intuitions about free will, moral responsibility, and related concepts. Collectively, these studies have found that people's intuitions about free will, moral responsibility, knowledge, and other concepts apparently vary with respect to how "abstractly" or "concretely" a hypothetical case is described (Nichols and Knobe 2007; Nahmias, Coates, and Kvaran 2007; Sinnott-Armstrong 2008; Roskies and Nichols 2008; de Brigard, Mandelbaum, and Ripley 2009; Freiman and Nichols 2011; Mandelbaum and Ripley ms). In this section I will follow Sinnott-Armstrong (2008) in referring to this collection of results as the *abstract/concrete paradox*, though below I will ar-

[6] I do not want to claim that these studies show, or even suggest, that intuitions are reliable. Debates about the reliability of intuitions are often framed in terms of *un*reliability, with theorists arguing either that intuitions are unreliable or that intuitions are not unreliable (see, e.g., Cummins 1998; Goldman 2007; 2010; Weinberg et al. 2012), and I will do the same in this paper.

gue that the various studies ostensibly concerning *the* abstract/concrete paradox actually concern several different phenomena (and, hence, that the handful of proposed unifying explanations of the abstract/concrete paradox are all mistaken). Thus, to be clear, I do not think the abstract/concrete paradox tracks a single phenomenon, nor do I believe there is a good way to distinguish between abstract cases and concrete cases. However, to make my case for those claims I must first review the studies and results (this section), proposed explanations of the paradox (section 3), and then critically examine those explanations (section 4). Sections 2-4 pave the way for my general critique of the abstract/concrete literature (section 5), and I then discuss the implications of my critique for arguments about the reliability of intuition in section 6.

Before reviewing the empirical studies, it will be helpful to clarify the relevant terminology. The *abstract/concrete paradox* refers to the finding that non-philosophers tend to attribute responsibility, free will, knowledge, and desert to agents in concrete cases, while they tend to not attribute responsibility, free will, knowledge, or desert to agents in abstract cases even though, according to Mandelbaum and Ripley, “both cases might describe an identical action in an identical situation—the only difference is that the cases aren’t identically described” (ms; p. 1). Philosophers have defined *concreteness* in several different ways. Sinnott-Armstrong (2008) claims that concreteness can refer to specific stimuli (as opposed to general stimuli) or particular stimuli (as opposed to universal stimuli). De Brigard, Mandelbaum, and Ripley (2009) say that a case can be abstract if the action performed by the agent in the case is unspecified. Freiman and Nichols follow Sinnott-Armstrong, claiming that a case is concrete if it mentions one or more specific agents and referring the reader to Sinnott-Armstrong for “further discussion on the significance of the abstract/concrete distinction” (2011; fn. 35). And Mandelbaum and Ripley state, “What concreteness is varies from case to case: sometimes it amounts to describing in more detail an action that had previously been left undescribed; sometimes asking about a particular case instead of a general case; sometimes asking about an action happening in our

world vs. one in an alternate universe; sometimes something else" (ms; p.1). We can now turn to the studies themselves.

Nichols and Knobe (2007)

To investigate folk judgments about the compatibility of determinism and moral responsibility, Nichols and Knobe (2007) gave non-philosophers the following materials:

> Imagine a universe (Universe A) in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it *had to happen* that John would decide to have French Fries.
>
> Now imagine a universe (Universe B) in which *almost* everything that happens is completely caused by whatever happened before it. The one exception is human decision making. For example, one day Mary decided to have French Fries at lunch. Since a person's decision in this universe is not completely caused by what happened before it, even if everything in the universe was exactly the same up until Mary made her decision, it *did **not** have to happen* that Mary would decide to have French Fries. She could have decided to have something different.
>
> The key difference, then, is that in Universe A every decision is completely caused by what happened before the decision – given the past, each decision *has to happen* the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does **not** have to happen* the way that it does.

Half the people in the study were randomly assigned to the *abstract* condition and received the following question:

> In Universe A, is it possible for a person to be fully morally responsible for their actions?
>
> YES NO

The other half of the subjects were in the *concrete* condition and read this question:

> In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family.

> Is Bill fully morally responsible for killing his wife and children?
>
> YES NO

In the *concrete* condition, 86% answered "Yes," while only 14% answered "Yes" in the *abstract* condition. According to Mandelbaum and Ripley (ms) and Freiman and Nichols (2011), the concrete condition is concrete because it mentions a specific action performed by the agent in the scenario. Similarly, Sinnott-Armstrong says that the *abstract* condition is abstract because it "does not mention any particular person or act or victim or time or place." In contrast, he claims that the *concrete* condition is concrete because it "does mention a particular agent, victims, and act" (2008; p. 214).

Nahmias, Coates, Kvaran (2007)

Nahmias, Coates, and Kvaran (2007) ran several experiments to test whether judgments about moral responsibility, free will, and related concepts shift according to the abstractness or concreteness of the case. Participants in the *Abstract* condition received the following case:

> Story: On Erta, the landscape and life are very similar to Earth, and there are advanced life forms called Ertans who look, talk, and behave very much like we do. For instance, the Ertans have families, schools, various jobs, parties, arguments, etc. However, the Ertans' science has advanced far beyond ours. Specifically, Ertan psychologists have discovered exactly how Ertans' minds work. For instance, they have discovered that whenever an Ertan is trying to decide what to do, the decision the Ertan ends up making is completely caused by the specific thoughts, desires, and plans occurring in his or her mind. The psychologists have also discovered that these thoughts, desires, and plans are completely caused by the Ertan's current situation and the earlier events in his or her life. These earlier events were also completely caused by even earlier events, eventually going all the way back to events that occurred before the Ertan was born.
>
> So, once specific earlier events have occurred in an Ertan's life, these events will definitely cause specific later events to occur. For instance, once specific thoughts, desires, and plans occur in the Ertan's mind, they will definitely cause the Ertan to make the specific decision he or she makes.

Other participants were assigned to either a *Good* condition or *Bad* condition. Both conditions were different from the *Abstract* condition in that a specific agent and specific action is mentioned. Mandelbaum and Ripley claim that the two scenario pairs, *Abstract*/*Good* and *Ab-*

stract/*Bad*, are both instances of the abstract/concrete paradox since *Good* and *Bad* mention specific agents and specific actions while *Abstract* does not (ms; p. 4). Moreover, Nahmias et al. note that they created and tested *Good* and *Bad* conditions because they "were . . . interested in replicating and examining Nichols and Knobe's finding that people's judgments are significantly influenced by the presentation of concrete agents performing morally salient actions" (2007; p. 223). Both *Good* and *Bad* groups received the same case as people in the *Abstract* condition, except that the last paragraph was replaced with one of the following (*Bad* condition shown; *Good* condition in brackets):

> So, once specific earlier events have occurred in an Ertan's life, these events will definitely cause specific later events to occur. For example, one day an Ertan named Smit decides to kill his wife so that he can marry his lover [donate a large sum of money to an orphanage in his community], and he does it. Once the specific thoughts, desires, and plans occur in Smit's mind, they will definitely cause his decision to kill his wife [donate a large sum of money to an orphanage in his community].

Compared to judgments of responsibility, free will, praise, and blame in the *Abstract* condition, those same judgments were significantly higher in the *Good* and *Bad* conditions, which both mention specific agents performing specific actions (Nahmias et al. 2007; p. 234).

Sinnott-Armstrong (2008)

Sinnott-Armstrong (2008) ran a study to investigate abstract/concrete differences with respect to judgments of knowledge. Half the survey participants received the following case and question:

> People sometimes believe things for no good reason. For example, people sometimes believe what a politician says about the economy when they have no good reason to trust what the politician says. Our question is about knowledge: If you cannot give any good reason to believe that the person whom you believe to be your mother really is your mother, is it possible that you *know* that she is your mother?

The other half of participants received the same case but with a different question:

> [All same up to the colon, then new question]: If a person cannot give any good reason to believe a claim, is it possible that the person *knows* that the claim is correct?

In the study, 88% of participants in the concrete condition said Yes, while only 52% of participants in the abstract condition said Yes. Sinnott-Armstrong refers to the second case as an “abstract question” (2008; p. 220) and contrasts it with “a more concrete version” in the first case (p. 221). Although he does not explicitly state what he thinks make the first case the “more concrete version” than the “abstract question” in the second case, he does note that the terms abstract and concrete can refer to general stimuli and specific stimuli, respectively (p. 226; fn. 1). Thus, presumably, the first case is concrete because it mentions a specific person—your mother—while the second case is abstract because it mentions only a general, unspecified person.

Roskies and Nichols (2008)

Roskies and Nichols (2008) were curious about the role of setting in the production of judgments about moral responsibility. Nichols and Knobe’s cases were set in an alternate universe, but other studies used scenarios set in our world (e.g., Nahmias et al. 2006). Roskies and Nichols found that judgments about moral responsibility, free choice, and desert in a deterministic universe can vary according to the setting of the scenario, either our world or an alternate world. Half of their survey participants were assigned to the *Actual* condition and read the following case:

> Many eminent scientists have become convinced that every decision a person makes is completely caused by what happened before the decision – given the past, each decision *has to happen* the way that it does. These scientists think that a person’s decision is always an inevitable result of their genetic makeup combined with environmental influences. So if a person decides to commit a crime, this can always be explained as a result of past influences. Any individual who had the same genetic makeup and the same environmental influences would have decided exactly the same thing. This is because a person’s decision is always completely caused by what happened in the past.

In the *Alternate* condition, participants received the same scenario preceded by the following:

> Imagine an alternate universe, Universe A, that is much like earth. But in Universe A, many eminent scientists have become convinced . . .

Participants in both conditions were then asked about moral responsibility, blame, and free choice. In response to the scenario set in our world, most people said it is possible for people to be fully morally responsible for their actions and be morally blameworthy for committing crimes. In the *Alternate* condition, most people said it is not possible for people to be fully morally responsible or be morally blameworthy for committing crimes. Thus, judgments about moral responsibility and blame vary depending on whether the judgment-eliciting scenario is set in our world or an alternate world, even when all the facts about the two worlds are the same.[7] According to Mandelbaum and Ripley (ms), these cases are an instance of the abstract/concrete paradox because the *Alternate* condition, being set in an alternate world, is more abstract than the case set in our world. And for Sinnott-Armstrong, "the alternate condition is much less concrete than the actual condition" because "when we think about the actual world, we know much more about it" (2008; p. 216).

De Brigard, Mandelbaum, & Ripley (2009)

De Brigard, Mandelbaum, and Ripley (2009) also tested abstract/concrete differences in judgments of moral responsibility. They conducted a study with two conditions. In the *Concrete* condition participants read the following case and answered the question following it:

> *Dennis has recently found out from his doctor that he has a neurological condition that has, in the past, caused him to rape women. Were someone else to have this neurological condition then that person would have had to behave in the same ways as Dennis.*
>
> *On a scale of 1-7, 1 being not responsible, 7 being very responsible, how morally responsible is Dennis for raping women?*

In the *Abstract* condition participants read the following:

[7] Judgments of free will were also significantly higher in the *Actual* condition, though the means for both conditions were above the midline of responses, i.e., most people in *both* conditions said it was possible for people to make truly free choices.

> *Dennis has recently found out from his doctor that he has a neurological condition that has caused him to behave in certain ways. Were someone else to have this neurological condition then that person would have had to behave in the same ways as Dennis.*
>
> *On a scale of 1-7, 1 being not responsible, 7 being very responsible, how morally responsible is Dennis for the behaviors that are caused by his neurological condition?*

In the *Abstract* condition, participants held Dennis significantly less responsible than in the *Concrete* condition. De Brigard et al. claim that their second condition is abstract because the agent's action was not specified.

Freiman & Nichols (2011)

Freiman and Nichols (2011) investigated whether an abstract/concrete difference would be found with judgments of fairness and desert. In their *abstract* condition, people read the following statement:

> Suppose that some people make more money solely because they have genetic advantages.

In the *concrete* condition, people received the following scenario:

> Suppose that Amy and Beth both want to be professional jazz singers. They both practice singing equally hard. Although jazz singing is the greatest natural talent of both Amy and Beth, Beth's vocal range and articulation is naturally better than Amy's because of differences in their genetics. Solely as a result of this genetic advantage, Beth's singing is much more impressive. As a result, Beth attracts bigger audiences and hence gets more money than Amy.

In the *concrete* condition, most people said that Beth deserved the extra money and that it was fair that Beth got more money than Amy. However, in the *abstract* condition most people said that the people with genetic advantages did *not* deserve the extra money and also that it was *not* fair that those people got more money. Thus, changing only the concreteness—that is, whether or not one or more specific agents was mentioned—of the case resulted in a dramatic shift in people's judgments about fairness and desert.

Collectively, these studies seem to suggest that there is a shift in judgments about moral responsibility, free will, and related concepts across abstract and concrete stimuli, where “abstract” and “concrete” vary from case to case. A handful of explanations of the paradox have been proposed, which I will now review.

3 POSSIBLE EXPLANATIONS OF THE ABSTRACT/CONCRETE PARADOX

In section 2 I presented a collection of experimental studies that some theorists believe concern a single phenomenon, the abstract/concrete paradox. Those theorists have proposed several explanations of the abstract/concrete paradox, which I summarize in this section. I present these studies and explanations concerning the abstract/concrete paradox in order to provide a backdrop for my critique of the abstract/concrete literature (Sections 4 and 5), the implications of which are important for arguments about the reliability of intuitions (Section 6). Here are the explanations.

Knobe and Nichols

Nichols and Knobe (2008) offer what they call the affective performance error model. According to the affective performance error model, people normally use a tacit theory when making judgments of responsibility, but powerful affective reactions that arise in response to violations of moral norms leave people unable to correctly apply their tacit theory. Nichols and Knobe distinguish people’s representations of the criteria for moral responsibility and the performance systems people use to apply those criteria to specific scenarios (671). Affective reactions lead to performance errors in that they prevent the normal functioning of people’s performance systems that enable them to apply their criteria of moral responsibility to particular cases.

Sinnott-Armstrong

Sinnott-Armstrong (2008) provides an explanation of the abstract/concrete paradox based on different memory systems. Following Mandelbaum and Ripley (ms), I will call Sinnott-Armstrong's proposal the *separate capacities hypothesis*. This hypothesis claims that we have independent mental mechanisms for encoding concrete cases and abstract cases, which lead to different intuitions about particular cases. Specifically, Sinnott-Armstrong invokes a well-known distinction between two types of memory—episodic and semantic—to explain the abstract/concrete paradox. He states, "Episodic memory represents particular or specific events, including actions by a person. Semantic memory represents more abstract properties or general traits of a person" (2008; p. 222). For instance, my *semantic* memory of 9/11 includes a collection of facts such as "happened on September 11th, 2001" and "two planes crashed into the twin towers." But my *episodic* memory of 9/11 includes information about my actual experience that day of 7th-grade history class with Mr. Lowe being moved to the library, the school being locked down, and my general state of confusion and unease. For Sinnott-Armstrong, concrete cases are represented in episodic memory which tends to produce increased attributions of responsibility, while abstract cases are represented in semantic memory which tends to produce decreased attributions of responsibility.[8] The abstract/concrete paradox arises since "these systems naturally conflict with each other, because the concrete system stores counter-instances to the generalizations in the abstract system" (2008; p. 222). According to Sinnott-Armstrong, the separate capacities hypothesis is supported by an inference to the best explanation:

> [The separate capacities hypothesis] would explain why philosophical intuitions are so persistent. Philosophers who deny responsibility often admit that they have to fight the appearance of responsibility in horrible criminals . . . The same pattern recurs for paradoxes in epistemology and ethics. Both sides of the debate usually feel the force of intuitions on the other side. These appearances . . . are

[8] Sinnott-Armstrong, however, notes that abstract cases do not always lead to abstract representations, nor do concrete cases always lead to concrete representations. For this reason, along with individual variation and the fact that other factors such as the amount of affect-invoking content of cases will interact with the concreteness of the case, he says that "surveys responses will vary a lot, and only some of this variance will be explained by the dichotomoy between abstract and concrete" (p. 224).

> just what one would expect if the philosophical intuitions come from distinct representational systems. (2008; pp. 223-224)

Mandelbaum and Ripley

Mandelbaum and Ripley (ms) propose what they call the Norm Broken, Agent Responsible (NBAR) theory of the abstract/concrete paradox. According to the NBAR theory, people hold the following unconscious belief: if a norm is broken, an agent is responsible.[9] People use this belief as a heuristic when evaluating particular cases. In most concrete cases, a norm is clearly broken, and people tend to attribute responsibility to the agent. In most abstract cases, however, it is less clear that a norm is broken; thus, people are unlikely to attribute responsibility. Mandelbaum and Ripley believe that concreteness can be different from case to case, including the mention of a specific agent, a specific action, being set in the actual world (ms; p. 1). So the NBAR theory appears to account for many[10] of the putative instances of the abstract/concrete paradox reviewed in Section 2.[11]

---

[9] Mandelbaum and Ripley use a wide conception of "norm" on which norms are, roughly, expectations, such as "whales don't fly" and "philosophy professors shouldn't beat their students for bad papers." Mandelbaum and Ripley note that on their account, every belief (moral and non-moral) that someone has about how the world ought to be is a norm (fn. 6, p. 8).

[10] Specifically, Mandelbaum and Ripley claim that NBAR theory can explain the results of Nichols and Knobe (2007), Nahmias et al. (2007), Roskies and Nichols (2008), and de Brigard et al. (2009). They do not mention Sinnott-Armstrong's (2008) study. Nor do they mention Freiman and Nichols (2011), possibly because they were unaware of that study at the time of the most recent draft of their paper.

[11] There are seemingly obvious counterexamples to the NBAR theory that Mandelbaum and Ripley consider. They note that, for example, bad happenings—such as a tree blowing over in a storm and crushing someone's baby—are situations in which a norm is broken, but an agent is not responsible. Bad happenings thus appear to undermine the NBAR theory. However, Mandelbaum and Ripley reject this conclusion by appealing to cognitive dissonance theory, which states that someone holding inconsistent beliefs creates a negative motivational state in that person. This negative motivational state leads the person to either reject the threatened belief or find a way to not view it as actually threatened, e.g., disregarding salient evidence. In cases of bad happenings, then, people can reject the belief that that happening is bad, reject NBAR, or decide that someone is responsible for the bad happening (10). Mandelbaum and Ripley claim that people will typically decide that someone is in fact responsible in this situation, rather than thinking their baby being crushed is not actually bad or rejecting NBAR (which they take to be a well-entrenched heuristic) (10).

4 EVALUATION OF EXPLANATIONS OF THE PARADOX

I have now reviewed a collection of experimental studies purportedly concerning what Sinnott-Armstrong has called the abstract/concrete paradox (Section 2) and presented the handful of proposed explanations of that paradox (Section 3). I will now evaluate the success of the candidate explanations of the abstract/concrete paradox and conclude that none of the existing explanations is successful. The purpose of this section is to set up my discussion in Section 5, where I will highlight some general problems with the abstract/concrete literature and argue that the paradox has yet to be explained because theorists have mistakenly thought that there is a single phenomenon to be explained. I then discuss the implications of my critique for arguments about the reliability of intuitions in Section 6.

Nichols and Knobe

Nichols and Knobe's (2008) affective performance error model makes the following prediction: when people experience strong affective (or, more specifically, *sufficiently* strong affective) reactions when evaluating a scenario and deciding whether or not to attribute moral responsibility, they will withhold attributions of moral responsibility—this description just *is* the thesis of the affective performance error model. However, it is important to fully draw out the implications of this prediction. In particular, the model posits no special link between the concreteness of the intuition-eliciting scenario and the level of responsibility attributed. Rather, the level of responsibility attributed to the actor in the scenario is mediated by the affective reaction experienced in response to the details of the scenario by the person making the attribution. In experimental studies, concrete scenarios typically elicit higher attributions of responsibility than abstract scenarios because the concrete scenarios are typically more affectively-laden than abstract scenarios. The important point here is that the crucial link posited by the affective performance error model runs between the person's *experience* of affect and the functioning of their performance systems used to generate responsibility judgments. Thus, the affective perfor-

mance error model makes the following prediction: affectively-primed subjects will attribute levels of responsibility in response to low-affect abstract scenarios that are comparable to the levels of responsibility they attribute in response to high-affect concrete scenarios.[12] In other words, inducing a strong affective reaction, even if it does not come from the details of the intuition-eliciting scenario, should increase responsibility attributions regardless of the concreteness or other details of the scenario.

Here is one way to test this prediction: Prime people to engage in emotional processing[13] and then give them a low-affect abstract scenario. Nichols and Knobe's affective performance error model predicts increased attributions of responsibility for affectively-primed subjects given a low-affect abstract scenario, compared to non-affectively-primed subjects given the same scenario. Likewise, affectively-primed subjects given a low-affect abstract scenario should respond similarly to non-affectively-primed subjects given high-affect concrete scenarios.

A more sophisticated way to test this prediction is to look at clinical populations with deficits in emotional processing. If someone's affective processing is disrupted, they should not ex-

---

[12] This prediction may be blocked depending on how Nichols and Knobe's presentation of the affective performance error model is read. They say that "people ordinarily make responsibility judgments by relying on a tacit theory, but when they are faced with a truly egregious violation of moral norms (as in our concrete cases), they experience a strong affective reaction which makes them unable to apply the theory correctly" (671). There are at least three ways to interpret the role of moral norms violations in the model. First, moral norms violations may be *one example* of the various ways in which the details of the intuition-eliciting scenarios elicit a strong affective reaction (in addition to, e.g., affectively-laden words or phrases that do not refer to moral norms violations, such as "rotting cat corpse being eaten by maggots and cockroaches"). Second, moral norms violations may be the *only* way the scenarios elicit a strong affective reaction. Third, moral norms violations may be the only way the scenarios elicit a strong affective reaction *which leads to a performance error* (as opposed to moral norms violations that elicit strong affective reactions but do not disrupt the normal functioning of the performance systems). The prediction in question is compatible with the first two interpretations but incompatible with the third interpretation. Thus, Nichols and Knobe may block my prediction here (which, ultimately, creates problems for their model) if the role of moral norms violations in their model is intended to be read in the third way. However, in support of the affective performance error model, Nichols and Knobe appeal to the several findings in the broader social psychological literature on affect and cognition; those findings employ several ways in which affect may lead to performance errors (not merely through moral norms violations), so it appears that Nichols and Knobe assume something like the first interpretation.

[13] Emotional processing could be induced using a variety of techniques including pictures, videos, or tactile stimuli, among others. The International Affective Picture System—a standardized set of emotionally-evocative, internationally accessible color photographs—is very common in studies of emotion and could be used here as well. See Lang, Bradley, & Cuthbert (2008) for more details. Thanks to Billy Brady for this information.

hibit the performance errors in response to high-affect concrete scenarios predicted by the affective performance error model. In other words, if affective reactions prevent the correct application of people's tacit theory of moral responsibility, then patients with affective-processing deficits should attribute roughly the same levels of responsibility in response to abstract and concrete scenarios. Cova, Bertoux, Bourgeois-Gironde, and Dubois (2012) tested this hypothesis using people with a behavioral variant of frontotemporal dementia (bvFTD), who have been shown to have reduced emotional processing. Nichols and Knobe's affective performance error model predicts that people with bvFTD will attribute *less* responsibility to agents in concrete scenarios than people without bvFTD, since the affective content of typical concrete scenarios cannot disrupt the application of the bvFTD subjects' tacit theory of moral responsibility.

Cova et al. (2012) found that people with bvFTD were no less likely to attribute responsibility to agents in concrete scenarios than people without bvFTD. The results of Cova et al.'s experiment are not compatible with Nichols and Knobe's affective performance error model of the abstract/concrete paradox. Until evidence suggesting otherwise is provided or Cova et al.'s results are explained away,[14] we have reason to reject the affective performance error model as an explanation of the abstract/concrete paradox in moral psychology.

Sinnott-Armstrong

Sinnott-Armstrong's separate encoding capacities hypothesis is implausible for two reasons. First, he provides little support for his claim that abstract and concrete scenarios[15] are represented by semantic memory and episodic memory, respectively. By the same token, I can simply claim that abstract scenarios are encoded by episodic memory and concrete scenarios are encoded by semantic memory. He does claim that the separate capacities hypothesis, if

[14] There are at least a few possibilities for someone wishing to pursue this line. For example, they may highlight the very small sample size (10 subjects). Or they may argue that people with bvFTD are in fact left with sufficient emotional processing abilities to disrupt application of their tacit theory of moral responsibility when given concrete cases.

[15] Recall that for Sinnott-Armstrong, the abstract/concrete distinction can refer to a distinction between general and specific stimuli or a distinction between universal and particular stimuli.

true, would explain why pairs of paradoxical intuitions are common in philosophy—nearly[16] every person has both episodic and semantic memory, so if the separate capacities hypothesis were true, it would be unsurprising that philosophers often feel the force of *both* the intuitions constituting a paradox. However, this inference is false. Though the abstract/concrete paradox involves intuitions about abstract cases and intuitions about concrete cases, not all paradoxes in philosophy are constituted by an intuition about an abstract case and an intuition about a concrete case. Thus, the intuitions of some paradoxes many both be represented by semantic memory or by episodic memory. If the separate capacities hypothesis were true, we would not expect to feel the force of both intuitions of a paradox if those intuitions are both represented by the same memory system. Yet we do feel the force of both intuitions of many paradoxes that do not involve an abstract case and a concrete case. So Sinnott-Armstrong's inference is false.

Second, even if concrete scenarios were encoded by episodic memory, we should not expect anything resembling uniform judgments about concrete cases since each person's episodic memories are specific to that person—episodic memories represent specific episodes that an individual herself has experienced. Thus, unless most people's episodic memories are sufficiently similar, we should expect scenarios encoded in episodic memory to be processed quite differently by each person and, hence, we would have no reason to predict much uniformity in judgments to hypothetical cases. In particular, we have no good reason to predict, as Sinnott-Armstrong's hypothesis claims, increased attributions of moral responsibility, free will, and related concepts in response to concrete cases.

Mandelbaum and Ripley

One problem for the NBAR theory concerns its limited scope. The NBAR theory offers an account of the abstract/concrete paradox in moral psychology based on the empirical assumption that most people have a well-entrenched heuristic which they employ when making

[16] Sinnott-Armstrong cites cases of people with brain damage who have episodic memory but not semantic memory and vise versa (Tulving 1989; Klein, Loftus, & Kihlstrom 1996).

evaluations of responsibility. However, the limited scope—the domain-*specificity*—of the NBAR heuristic is problematic for the NBAR theory. For the Norm Broken, Agent *Responsible* theory cannot account for a large part (possibly even the *majority*) of the existing relevant data, since abstract/concrete differences have been found in judgments of concepts other than responsibility, including fairness, free will, desert, and knowledge (Sinnott-Armstrong 2008). Minimally, the NBAR theory must be revised and its scope broadened to account for much of the relevant data.

A second problem for the NBAR theory concerns the wide conception of norm that it employs and the predictions that follow from that conception. On the NBAR theory, a norm is best conceived of as any kind of expectation, where "expectation" is construed as broadly as possible (16). The broad construal of expectation permits both moral and non-moral expectations to count as norms and leads to a bevy of predictions for the NBAR theory. For example, Mandelbaum and Ripley claim that "if traffic doesn't swell at five, we predict that someone who had been expecting it to swell will look for an agent responsible for the lack of swelling" (16). Though some form of empirical testing would be needed to actually test that prediction of the NBAR theory and I can only provide my own intuition here, I doubt the average commuter expecting traffic to swell at five would look for an *agent* responsible for the lack of traffic swell.[17]

[17] I set aside the commuter prediction here somewhat quickly, and Mandelbaum and Ripley themselves only mention it in passing at the end of their paper as an example of a norm. But it is not wholly uninteresting to explore the first-person phenomenology of that case. In my own analogous case, when the train comes at the end of the day and it is not overcrowded, I typically assume that another train not normally scheduled was in service that day and cleared most of the passengers from the station immediately preceding my station. Perhaps in some non-obvious way I am attributing responsibility to some agent in this case (or maybe the unscheduled train counts as an agent). But my experience does not clearly indicate deployment of something like the NBAR heuristic.

5 GENERAL PROBLEMS WITH THE ABSTRACT/CONCRETE LITERATURE

In Section 4 I argued that a satisfactory explanation of the abstract/concrete paradox has yet to emerge. In this section my goal is to highlight more general problems with some of the ways theorists have been thinking about the abstract/concrete literature. I will try to show that no satisfactory explanation is available because the experiments putatively looking at the abstract/concrete paradox have actually been looking at several distinct phenomena mistakenly labeled as a single, unified phenomenon. There are several sources of confusion that have led researchers to mistakenly attribute the results of their empirical studies to a single phenomenon.

First, there is no clear definition of abstract and concrete. Sinnott-Armstrong says that "the slippery terms 'abstract' and 'concrete' can refer either to the dichotomy between universal and particular or to the separate dichotomy between general and specific (as well as to other dichotomies)" (2008; fn. 1). Freiman and Nichols distinguish between questions that are phrased abstractly and questions concerning a concrete individual (2011, p. 8). And de Brigard, Mandelbaum, and Ripley (2009) claim that their abstract condition is abstract because the scenario does not specify the action that the agent commits. Creating more confusion and ambiguity, Mandelbaum and Ripley state, "What concreteness is varies from case to case: sometimes it amounts to describing in more detail an action that had previously been left undescribed; sometimes asking about a particular case instead of a general case; sometimes asking about an action happening in our world vs. one in an alternate universe; sometimes something else" (ms; p.1). Though these philosophers have treated the various studies all as instances of the abstract/concrete paradox, it is clear that there is no standard definition or abstract and concrete. The various notions of abstract and concrete—e.g., whether a case mentions a specific agent or not, or whether it is set in the actual world or an alternate world, or whether an action is described in more detail or less detail—pick out what seem to be very different variables. Thus, even if we tried to include all possible notions of abstract and concrete in one definition of ab-

stract and concrete (perhaps via a large, disjunctive definition), many different cases that would count then as abstract (or concrete) could appear significantly different from each other.

Another source of confusion surrounding the abstract/concrete paradox is the conflation of two distinctions: abstract/concrete and low-affect/high-affect. Whatever the preferred notions of abstract and concrete, the distinction between abstract cases and concrete cases will be different from the distinction between cases with little affect-invoking content and cases with lots of affect-invoking content (unless, of course, the abstract/concrete distinction is explicitly defined as the low-affect/high-affect distinction). Given that affective experiences can activate different psychological processes, experimental philosophers must control for the affect-invoking content of their cases in order to avoid getting spurious results. Nichols and Knobe's (2007) case conflated this distinction, though they controlled for affective content in later experiments. De Brigard, Mandelbaum, and Ripley (2009) also conflate these two distinctions in their experiment, since the specific action in their concrete case, a man raping women, is heavily affectively-laden. With such an action in the concrete case, it is difficult to separate effects from the concreteness of the scenario from effects due to the affect-invoking content of the scenario.

A third source of confusion surrounding the paradox arises from the fact that philosophers have overlooked differences in the lengths of their cases. Abstract and concrete cases can *both* be either very long or very short (or some length in between). For example, Nichols and Knobe's abstract and concrete cases are both quite long, at over three full paragraphs, while Freiman and Nichols' abstract case consists of just one sentence. The length of cases is important because it may affect how people process the scenario and probe. If an abstract case is long enough, it may include enough details—even if not about specific agents or specific actions—to, e.g., elicit affective reactions or other potentially biasing cognitive processes. In sum, using scenarios of different lengths can lead people to use different cognitive processes, even if those different scenarios all satisfy the definition of abstract or concrete. Scenario length is an

important variable for empirical surveys, and so far it has been overlooked by philosophers studying the abstract/concrete paradox.

The preceding three concerns provide converging evidence of serious confusion surrounding the abstract/concrete paradox. The debate appears to be quite a mess. Given that there is no standard definition of abstract or concrete, scenarios have not been controlled for their amount of affect-invoking content or length or judgment type, and also that many of the relevant findings have not been replicated, the best explanation of the abstract/concrete paradox is that there is no single phenomenon to be explained. The term abstract/concrete paradox may not refer to any particular phenomenon at all. It therefore is unsurprising that there is no good unifying explanation currently available for the various results discussed. There exists no satisfactory explanation of the paradox because there is no one phenomenon to be explained! There are several dimensions along which the scenarios used in studies concerning the abstract/concrete paradox may vary. As shown in Figure 1, no single dimension captures all the studies; putative cases of the abstract/concrete paradox have varied across five dimensions: mention of a specific agent, mention of a specific action, a relatively high amount of affect-invoking content, setting (the actual world or an alternate world), and mention of a specific agent or a specific action. None of those five dimensions captures all putative instances of the paradox, and each putative instance varies across at least two or more of those dimensions.

Figure 1

	Agent	Action	Affect	Setting	Agent *or* Action
Nichols & Knobe	x	x	x		x
Nahmias, Coates, Kvaran	x	x			x
Sinnott-Armstrong	x				x
Roskies & Nichols				x	
De Brigard, Mandelbaum, Ripley		x	x		x
Freiman & Nichols	x	x			x

These concerns do not stop at the abstract/concrete paradox, however. Many of the problems surrounding the paradox are applicable to most, if not all, other projects within experimental philosophy. Most experimental philosophy projects need not worry about providing precise definitions of abstract and concrete, nor do they need to worry about conflating the abstract/concrete distinction with low-affect/high-affect or other distinctions. But they must ensure that they way in which they present their hypothetical cases elicits intuitions about the phenomenon in question. This is especially important for studies attempting to replicate or expand upon previous experiments. As shown by the collection of studies on the abstract/concrete paradox, even seemingly insignificant changes in the language or length of a case can easily make it difficult to determine whether survey participants' intuitions actually concern the relevant phenomenon trying to be replicated.

6 EXPERIMENTS AND RELIABILITY OF INTUITION

I began my discussion by noting several popular challenges to the reliability of intuitions and how they might be met. Appealing to conflicting intuitions and then claiming that conflicting intuitions are prevalent enough to think that a high percentage of intuitions are wrong seemed to be the best available route for the skeptic of the reliability of intuitions. And, conversely, making this same appeal and claiming that conflicting intuitions are not very common seemed to be a good way for proponents of appeals to intuitions to meet the skeptic's best available challenge.[18]

[18] Note that meeting the skeptic's challenge to the reliability of intuitions does not tell us whether or not intuitions are reliable, nor what reliability entails. If conflicting intuitions are not common, we would have evidence suggesting only that the skeptic's challenge to reliability from conflicting intuitions fails. To be clear, I do not think my discussion of the abstract/concrete paradox brings us any closer to understanding what reliability entails (though I discuss some alternative approaches to understanding reliability later in this section). I claim only that the abstract/concrete studies can be used to meet the skeptic's challenge to reliability from conflicting intuitions.

Above I showed that philosophers claiming to study the abstract/concrete paradox have actually been studying several distinct phenomena. That conclusion holds important implications for debates about the reliability of intuitions. For if the abstract/concrete paradox referred to a single phenomenon, then the results of those empirical studies would support the skeptic's challenge to reliability, since those studies would provide a large set of conflicting intuitions.[19] But if the abstract/concrete paradox did not refer to a single phenomenon, then the results of those studies would favor the proponent of the reliability of intuition, since those studies would provide many instances of intuitions that do not conflict.

Even if the results of empirical studies could be used to meet the skeptics' challenge to the reliability of intuitions, that alone suggests only that intuitions are not unreliable. Those results alone would not help us get clear on what it means for an intuition to be *reliable*. One strategy for honing in on reliability is to distinguish between different kinds of intuitions and define one kind of intuition in such a way that those kinds of intuitions are always (seemingly) reliable, at least to some extent. For example, Kauppinen (2007) distinguishes surface intuitions and robust intuitions. Robust intuitions are intuitions that a competent speaker of the language would give in ideal conditions—free from distorting influences and other biases—based upon semantic considerations and not pragmatic considerations. Surface intuitions, on the other hand, are quick, gut reactions to cases that do not involve careful reflection on and consideration of the details of the intuition-eliciting case and key concepts (2007, p. 103). This distinction between surface and robust intuitions might provide a useful first step in better understanding reliability, for if we can identify and isolate robust intuitions, we might be able to pull out the common properties of those intuitions in virtue of which they are reliable.

[19] As a reminder, by "conflicting intuitions" I mean conflicts that persist after possible verbal disagreements and potentially distorting factors such as clouded emotions, biases, etc. have been accounted for. In those cases, two people have the same facts about the phenomenon in question but interpret them differently. As a result, one of the conflicting intuitions must be wrong (Goldman 2007).

However, Kauppinen's distinction between surface intuitions and robust intuitions does not provide much help in getting clear on what reliability of intuitions entails. Robust intuitions may be more reliable than surface intuitions, but only because they are defined by Kauppinen as more reliable. Robust intuitions are intuitions that a competent speaker of the language might have under sufficiently ideal conditions. On this definition, robust intuitions are generally more reliable than surface intuitions since they are produced under ideal conditions—that is, 'conditions in which there are no perturbing, warping or distorting factors or limits of information, access or ability' (2007, p. 103). However, the problem with this account of intuitions, with respect to the debate about the reliability of intuitions, is that few, if any, intuitions studied by experimental philosophers will be robust intuitions. Moreover, few, if any, professional philosophers' intuition will be robust intuitions. For Kauppinen's "sufficiently ideal conditions" are unlikely to be realized in most situations where intuitions are produced. (Perhaps professional philosophers can attain sufficiently ideal conditions when thinking through a hypothetical case in their armchair, if they are aware of common biases that may affect intuitions and guard against them, but it is unlikely that the typical survey participant will be both aware of those common biases and careful enough to try to avoid them.) If Kauppinen's conditions for robust intuitions are never satisfied or if they are only very rarely satisfied in experimental studies, then his distinction between surface intuitions and robust intuitions cannot inform the debate about the reliability of intuitions, since we will almost always be dealing with surface intuitions which do not help us get a grip on what it means for an intuition to be reliable. Other accounts of intuitions that include reliability as a part of the definition of intuition will also not be helpful for getting clear on reliability.

However, it may be possible to get a better grip on reliability by using something like Kauppinen's general strategy of distinguishing between different kinds of intuitions. Perhaps reliability of intuition can be characterized by proper functioning *under normal circumstances*. The problem with this strategy is that it will be difficult to distinguish normal circumstances from ab-

normal circumstances. It is unclear what the boundary is between normal and abnormal circumstances, and I suspect any attempts to define that boundary would be met with opposition. Consider, for example, the notions of abstract and concrete. Those two terms may seem at first to be reasonably easy to define in a satisfactory way. But as the set of studies on the abstract/concrete paradox show, that is not the case. It is actually quite difficult to define abstract and concrete in a satisfactory way—does affect-invoking content count, and if so, how much affect-invoking content counts? Does the mention of specific agents count as concrete, or must a case mention specific agents and specific actions? Demarcating normal and abnormal circumstances with respect to intuition-eliciting cases would be difficult as well. For the same questions concerning the demarcation of abstract and concrete must be answered along with many other questions. For example, does an intuition formed after discussing a case with a colleague count as an intuition formed under normal circumstances, or not? Reliability of intuition cannot be defined as proper functioning under normal circumstances unless we have an accepted notion of normal circumstances, which will be difficult to produce.

It is also possible to characterize reliability as proper functioning under a specific set of relevant possible cases. For example, we might characterize reliability as proper functioning in only cases set in the actual world or possible worlds some sufficiently close distance from the actual world. For perhaps we have reason to think that we only need intuitions about cases set in the actual world or close possible worlds, in which case intuition need only function properly in those cases. Even if intuition about cases set in distant possible worlds does not always function properly, this is not a strike against the reliability of intuition, since we only need intuition to function properly for cases set in the actual world or close possible worlds.

However, characterizing reliability in this way suffers from the same problem as characterizing proper functioning under normal circumstances. Characterizing reliability in this way does not solve the problem. It merely trades the problem of defining reliability for the problem of defining the set of relevant possible cases. And I see no easy way of satisfactorily defining the

set of relevant possible cases under which intuitions must function properly in order to be reliable. Intuitions are used in a variety of ways in many different areas of philosophy. Determining the set of relevant possible cases to ground reliability of intuitions *in general*—that is, across all the ways intuitions are used and across all the areas of philosophy where intuitions are appealed to as evidence—will be a difficult task.

7 CONCLUSION

When we attend to the overlooked differences in survey materials used to study the abstract/concrete paradox, we see that the paradox does not refer to a single phenomenon, contrary to what philosophers currently believe. It is therefore unsurprising that a satisfactory explanation of the paradox has yet to emerge, since existing explanations are attempting to explain a handful of different phenomena mistakenly unified under the label “abstract/concrete paradox.”

Consequently, the intuitions given in these studies do not actually conflict. They are simply different intuitions about different stimuli. These studies provide a large set of empirical data suggesting that conflicting intuitions, *contra* some skeptics about the reliability of intuitions, are not actually that common in the philosophical literature. If philosophers can overlook the important differences in materials used to study the abstract/concrete paradox, then philosophers may also overlook many other important differences in other areas of the philosophical literature. In that case, conflicting intuitions given in response to cases *seemingly* concerning the same phenomenon may only be *apparently* conflicting intuitions. And if many of the conflicting intuitions in the philosophical literature are only apparent conflicts, then the percentage or wrong intuitions may be quite low. This is an important empirical fact, since the best available way for skeptics about intuitions to challenge the reliability of intuitions turns on the empirical question of whether wrong intuitions are common. If philosophers attend more carefully to the details of their hypothetical cases used to elicit intuitions, they may find that their intuitions do not actually

conflict very often at all. Thus, the most promising challenge to the reliability of intuition may be able to be met by more careful consideration of what phenomena hypothetical cases actually refer to. But even if this challenge is met and it cannot be shown that intuitions are unreliable, future work will need to provide a careful analysis of the notion of reliability and a positive argument for the reliability of intuitions.

8 REFERENCES

Bealer, G. (1998). Intuition and the Autonomy of Philosophy. In Depaul M. & Ramsey W. (eds.) *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. New York: Rowman & Littlefield, 1998.

---. (2002). Modal Epistemology and the Rationalist Renaissance. In Gendler T. & Hawthorne J. (eds.) *Conceivability and Possibility*. New York: Oxford University Press.

Berker, S. (2009). The Normative Insignificance of Neuroscience. *Philosophy and Public Affairs* 37(4), 293-329

De Brigard F, Mandelbaum E, & Ripley D. (2009). Responsibility and the Brain Sciences. *Ethical Theory and Moral Practice* 12:5, 511-524.

Erickson, M., & Kruschke, J. (1998). Rules and Exemplars in Category Learning. *Journal of Experimental Psychology: General*, 127, 107-140.

Foley, R. (1998). Rationality and Intellectual Self-Trust. In Depaul M. & Ramsey W. (eds.) *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. New York: Rowman & Littlefield, 1998.

Freiman, C., & Nichols, S. (2011). Is Desert in the Details? *Philosophy and Phenomenological Research* 82, 121-133.

Gettier, E. (1963). Is Justified True Belief Knowledge?" *Analysis* 23, 121-123.

Goldman, A. (2010). Philosophical Naturalism and Intuitional Methodology. *Proceedings and Addresses of the American Philosophical Association*,115-150.

Gonnerman, C. Unpublished dissertation manuscript. Indiana University.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., & Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral Judgment. *Science* 293, 2105-2108.

Greene, J.D. (2007). The Secret Joke of Kant's Soul. *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*. (W Sinnott-Armstrong, ed.). Cambridge, MA: MIT Press.

Hahn, N., & Chater, U. (1998). Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable? *Cognition* 65, 197-320.

Kauppinen, A. (2007). The Rise and Fall of Experimental Philosophy. *Philosophical Explorations* 10(2), 95-118.

Mandelbaum, E., & Ripley, D. (unpublished manuscript). Explaining the Abstract/Concrete Paradoxes in Moral Psychology.

Nichols, S., & Knobe, J. (2008). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Nous* 41, 663-685.

Nosofsky, R., Palmeri, T., & McKinley, S. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101(1), 53-79.

Posner, M., & Keele, S. (1968). *Journal of Experimental Psychology*, 77(3), 353-363.

Roskies A, & Nichols S. (2008). Bringing Moral Responsibility down to Earth. *Journal of Philosophy* 105:7, 371-388.

Sinnott-Armstong W. (2008). Abstract+Concrete=Paradox. In Nichols S & Knobe J (eds.) *Experimental Philosophy*. New York: Oxford University Press.

Stich, S. & Weinberg, J. (2001). Jackson’s Empirical Assumptions. *Philosophy and*

Phenomenological Research 62, 637-643.

Swain, S., Alexander, J., & Weinberg J. (2008). The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research* 76(1), 138-155.

Weinberg J., Alexander J., Gonnerman C., & Reuter S. (2012). Restrictionism & Reflection: Challenge Deflected, or Simply Redirect? *Monist* 95(2), 201-222.

Weinberg, J., Nichols, S., & Stich, S. (2001). Normativity and Epistemic Intuitions. *Philosophical Topics* 29, 429-460.

Weiskopf, D. (2009). The Plurality of Concepts. *Synthese* 169, 145-173.

Williamson, T. (2004). Philosophical ‘Intuitions’ and Skepticism about Judgment. *Dialectica* 58(1), 109-153.