

**Concepts, Cases, and Compellingness: Exploring the Role of
Intuitive Analysis in Philosophical Inquiry**

JILL N. CUMBY

**A DISSERTATION SUBMITTED TO THE FACULTY OF
GRADUATE STUDIES IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF DOCTR
OF PHILOSOPHY**

**GRADUATE PROGRAM IN PHILOSOPHY
YORK UNIVERSITY
TORONTO, ONTARIO**

MAY 2015

© JILL N. CUMBY, 2015

Abstract

This dissertation provides a better understanding of the method of cases, a method widely used in philosophical theorizing. Using this method involves relying on one's intuitive judgments about cases to guide theorizing. Recently, such judgments have been experimentally examined, and it has been argued that the results of these studies encourage skepticism about the trustworthiness of this method. Responding to this skepticism involves developing a better understanding of the method of cases and the reliance on intuitive judgments in theory construction. I contribute to this project by arguing for a constraint on the kinds of hypothetical cases that can function as compelling counterexamples in conceptual analysis.

Dedication

This is dedicated to Michael for the most precious gifts of love and time, to my children for filling my days with cuddles and smiles, and to my parents.

Acknowledgements

I would like to thank Professor Henry Jackman for asking questions that forced me to think more carefully throughout this project. I would like to thank Professor Claudine Verheggen for careful comments and suggestions on each chapter draft, and whose knowledge of grammar (among other things!) I will always envy. I would like to thank Professor Kristin Andrews for her encouragement and direction. One could not ask for better professional role models. I would also like to thank all those people who make the York University Philosophy Department a supportive and positive environment.

Table of Contents

Abstract.....	ii
Dedication.....	iii
Acknowledgements.....	iv
Table of Contents.....	v
Introduction.....	1
Chapter 1.....	11
1.1 Extra-mentalism and Conceptual Analysis	11
1.2 Intuitive Judgments: An Empirical Analysis.....	12
1.3 Calibration.....	19
1.4 Mitigation.....	21
1.5 Functional vs. Natural Kinds.....	26
1.6 Exhaustive vs. Extrapolative Calibration.....	30
1.7 Theoretical Bridges and the Need for Knowing One's Instrument...	34
Chapter 2	39
2.1 Mentalism.....	39
2.2 Reliability.....	40
2.3 Remoteness.....	42
2.4 Inter-personal Variation and Verbal Disputes.....	45
2.5 Intra-subjective Variation.....	51
2.6 Concepts and Cognitive Science.....	57
2.7 Beyond Reliability?.....	63
Chapter 3	70
3.1 Good Old Armchair Conceptual Analysis.....	70
3.2 Wide and Narrow Content.....	73
3.3 Motivating Narrow Content.....	74
3.4 Motivating Wide Content.....	76
3.5 The Foundations of Two-Dimensionalism.....	80
3.6 Intentions and the Targets of Intuitive Analysis.....	83
3.7 Metaphysical Necessity and A priori Knowledge.....	85
3.8 A Closer look at Intentions.....	87
3.9 Epistemic Necessity and A priori Knowledge.....	93
3.10 Epistemic Intentions and Narrow Content.....	101
3.11 Remarks.....	106

Chapter 4	108
4.1 Successful and Unsuccessful Appeals to Hypothetical Cases.....	108
4.2 Hypothetical Cases, Counterfactual Worlds, and Thought Experiments..	111
4.3 Hypothetical Cases, Centered Worlds, and Perspective.....	115
4.4 Criterion of Compellingness.....	121
4.5 Objections.....	125
4.6 Chinese Nation, Chinese Room.....	135
4.7 Remarks.....	139
Conclusion.....	141
Bibliography.....	143

Introduction

Suppose you are tasked with answering a question of the form “What is x?” You might start by identifying instances of x. Once you have a collection of these instances, you might begin theorizing by contrasting these instances with instances of non-x’s asking yourself what it is that all of the instances of x have in common with each other but not with non-x’s. Here, your initial pre-theoretic judgments about what counts as an x will provide a constraint on your theorizing about x. A good theory will be one that is consistent with your pre-theoretic judgments. If you come up with a theory that is inconsistent with your basic pre-theoretic judgments about what counts as x, then this will provide a reason for thinking that something is wrong with the theory. If, on the other hand, you manage to come up with a theory of x that is consistent with your pre-theoretic judgments about what counts as an x and what does not, if your theory gets this distinction right, that fact speaks in favor of the theory.

Suppose that you have managed to come up with a theory of x that gets your pre-theoretic judgments right, how might someone convince you that your theory is wrong? They could present you with a case that you haven’t thought of before that elicits a judgment that runs counter to your theory. If your theory fails to account for this judgment then it will function as a counterexample to your theory, providing you with a good reason to think that your theory has gone wrong. You may try to amend your theory to account for this judgment, or you may try to explain away the judgment. Either way, a counterexample will show that further work needs to be done.

This method sketched above is known as the method of cases. The method of cases is widely used in philosophy. Indeed, if there is a distinctive philosophical method, then the method of cases is probably it. This is not to say that there aren't other methods that philosophers use. Of course, there are. But this method has been central to a certain way of doing philosophy, and has characterized a great deal of work across various historical epochs and in many different areas.¹ Achieving a better understanding of this method, the method of cases, is the central aim of this dissertation.

It has recently been argued that while philosophers have been busy using the method of cases, they have been blind to its limitations. Being charitable, one could say that this has not been a willful blindness, but rather that previous philosophers were not in possession of the relevant evidence showing their preferred method to be untrustworthy. Today though, things are different. Philosophers who identify with the “negative program” in experimental philosophy argue that we now have evidence for the claim that philosophers cannot legitimately rely on intuitive judgments in inquiry.² Confronted with this evidence, those who rely on the method of cases must 1) respond to this evidence by showing that it does not

¹ For a skeptical view on the importance of intuitions about cases in philosophy see Herman Cappelen, *Philosophy Without Intuitions* (2012).

² It is common to distinguish between the negative program in experimental philosophy that uses the results of empirical studies into people's intuitive judgments to challenge the legitimacy of relying on these judgments in inquiry, and the positive program in experimental philosophy that uses empirical studies of people's intuitive judgments to make positive claims about how the mind works. For more on this see Alexander, Mallon and Weinberg (2010), Alexander and Weinberg (2007), Kauppinen, (2007), Nadlehofer and Nahamias (2007) and Weinberg (2007).

actually threaten the legitimacy of the method of cases, 2) give up on the method of cases, or 3) face charges of intellectual irresponsibility.

As early as 1998, DePaul and Ramsey warn “western analytic philosophy is, in many respects, undergoing a crisis where there is considerable urgency and anxiety regarding the status of intuitive analysis” (1998, x). This crisis has intensified since the turn of the millennium as philosophers have become increasingly interested in studying people’s intuitive judgments about cases through use of empirical methods of inquiry. Some have suggested that the results of these studies support the claim that there should be a widespread moratorium on the use of intuitive judgments about cases in inquiry. Weinberg et al. conclude that “much of what has been done in epistemology in the analytic tradition” is “seriously undermined” by the results of empirical studies into people’s intuitive judgments (2001, 429). Machery et al. claim to have produced experimental results suggesting that certain assumptions underlying the reliance on intuitive judgments in traditional philosophical theorizing are “spectacularly misguided” (2004, B8). Alexander and Weinberg write “experimental evidence seems to point to the unsuitability of intuitions to serve as evidence at all” (2007, 63). Swain, Alexander and Weinberg report experimental results that “strengthen the empirical case against intuitions, such that philosophers who deal in intuitions can no longer rest comfortably in their armchairs” (2008, 138). There is a present and prevalent skeptical attitude toward the method of cases owing to its reliance on intuitive judgments that is encapsulated in the quotes presented above, or even better, in the popularized image of a burning armchair.

Such critiques push philosophers who rely on judgments about cases when theorizing to better explain their preferred method. Those who rely on the method of cases and who (understandably) have been busy using the method to make progress in theorizing have left these same theories open to skeptical attack by not taking time to fully think through the methodological foundations their analyses rest on. In the twentieth century, philosophers such as Quine (1960), Putnam (1973/75) and Kripke (1980) exploited these gaps in understanding to develop powerful critiques of conceptual analysis, a project that relies heavily on judgments about hypothetical cases. Philosophers such as Jackson (1998) and Chalmers (1996, 2002a, 2004) pushed back to develop equally powerful defenses of conceptual analysis. However, skeptical attack on the method of cases has not abated. The focus of these attacks has shifted from debates concerning the notions of analyticity and the possibility of *a priori* knowledge to questions concerning the evidentiary status of the intuitive judgments themselves.

In a sense, the new skepticism is broader in scope than previous concerns, threatening to bring down not only conceptual analysis but also any method of inquiry that relies on intuitive judgments along with it. If it is true that when it comes to relying on intuitive judgments “most philosophers do it openly and unapologetically, and the rest arguably do it too, although they would deny it” (Kornblith, 1998, 129) then not only conceptual analysts but philosophers more generally must respond to the new skepticism concerning the reliance on intuitive judgments in inquiry. Indeed, according to Williamson, if the skeptic is right, then the situation is far worse. According to Williamson “so-called intuitions are simply judgments (or dispositions to judge); neither their content nor the

cognitive basis on which they are made need be distinctively philosophical” (2007, 3). If so, the new skepticism will target an even broader set of practices, including those used in natural sciences and people’s everyday lives. This can be used to build a *reductio* argument against the new skepticism by showing that it leads to the implausible conclusion that all judgments are untrustworthy. According to Williamson, the recent attention on intuitions is useful for reminding us that when it comes to making judgments we could all be more careful. But there is nothing in the critique which warrants the view that these same judgments are generally untrustworthy or deserving of our epistemic scorn.³

Weinberg (2009, 2007) has argued that it is possible to differentiate between philosopher’s use of intuitive judgments in inquiry and other uses of intuitive judgments, claiming that the former but not the latter are problematic. Successful epistemic practices, such as those used in the natural sciences, typically have built in constraints that limit the probability of untrustworthy outputs and allow those who are engaged in the practice to identify when these are likely to occur. Philosophers have failed to articulate analogous constraints in their own practice of appealing to intuitive judgments.

Whether the new skepticism is best understood as a reminder that we should all be more careful when relying on intuitive judgments, as Williamson suggests, or specifically targets the philosophers’ use of intuitive judgments as the Weinberg suggests, I take it that an increased understanding of the practice of appealing to intuitive judgments in inquiry is

³ It is worth noting that Williamson’s defense of the reliance on intuitive judgments in inquiry is not to be taken as part of a larger project that defends conceptual analysis. Rather, Williamson’s own view is that “Our understanding of philosophical methodology must be rid of internalist preconceptions” 2007, 5.

valuable. This is so whether we think this is a necessary condition of legitimately relying on intuitive judgments when theorizing, or not. It would be nice to have a deeper understanding of the method of cases either way. I take it that the new skepticism is best understood as a call to acknowledge that “there is a gulf between our understanding of intuitions and their importance to us” (Weinberg, 2007, 318) and to commit to working toward a more detailed articulation and defense of the reliance on intuitive judgments in inquiry.

One way in which this general demand has been precisified is in relation to the kind of the cases that are used to invoke the intuitive judgments that are relied on in theorizing. Weinberg writes

“the practice [of appealing to judgments about hypothetical cases] appears to set no constraints to how esoteric, unusual, far-fetched or generally outlandish any given case can be. Everyone is familiar with Davidson’s Swampman, or Searle’s Chinese Room, but one can look at the recent literature and find the likes of double-lesioned testifiers, new evil demons, and fissioning/fusioning/teleporting pairs (or are they?) of persons. So this anything goes aspect of the practice is what makes it particularly ripe for the opponent’s challenge.” (2007, 321)

This dissertation takes up this challenge of outlining a constraint on the type of cases that are profitably utilized in the project of conceptual analysis.

There is at least one generally accepted constraint on the use of cases in conceptual analysis. It is often said that in order for a given case to invoke an intuition that functions as a successful counterexample, that case must be found to be compelling. Successful counterexamples present cases that compel others to either outright reject their favored theory because it is inconsistent with the judgment invoked by consideration of the case, make substantial revisions to their theory to make it consistent with the judgment about the

case, or to find some way of explaining away the inconsistent judgment. One general constraint faced by those who use hypothetical cases to make progress in analyzing concepts is that they design cases that are compelling.

As it stands, this constraint is woefully under-described. Without some positive explanation of what it is that makes a case compelling there is no way to ensure that cases are designed so that they tend towards compellingness. Furthermore, reflecting on a naïve understanding of compellingness might make skepticism about conceptual analysis more attractive rather than less. According to this naïve understanding, whether a case is compelling depends on one's being already inclined, on the basis of background theoretical commitments, to reject the theory to which the scenario is supposed to provide a counterexample. Fortunately, the naïve understanding of compellingness does not withstand scrutiny. Many epistemologists who held the justified, true belief account of knowledge found Gettier's cases compelling, as evidenced by their proceeding to either reject or make substantial revisions to their views based on consideration of Gettier's cases. If the suggestion that hypothetical cases are compelling only to those who hold a background theory consistent with rejection of the theory in question were right, then one would not expect this to happen. If the explanation of compellingness on offer cannot explain reactions to one of the most well known and indeed prototypical instances of a successful counterexample, then it is time to look for a better explanation.

In Chapter 4, I'll present and defend a general criterion of compellingness: a given hypothetical case will function as a compelling counterexample to the extent that the

audience considering the case and the agent at the center of the case being considered share a perspective. This general criterion of compellingness provides a constraint on the use of hypothetical scenarios in conceptual analysis; to the extent that we want to design cases that function as compelling counterexamples, we are constrained by the requirement that the agent at the center of the case possess a perspective that we can reasonably expect our audience to share. I'll show how this criterion of compellingness follows from a proper understanding of the nature of the cases that are appealed to in conceptual analysis, I'll provide an account of what constitutes perspective such that it can be shared and I'll support this criterion by showing that it correctly predicts reactions to some well-known cases in the history of philosophy.

The general criterion of compellingness that is developed herein relies on a particular view of what the target of conceptual analysis is, to be presented in Chapter 3. Here, the target of conceptual analysis is understood from within a two-dimensional framework of conceptual content most carefully developed by Chalmers (2002a, 2002b, 2004, 2006a, 2006b). Here, each concept can be thought of as having two distinct kinds of content. Subjunctive content, the more familiar type of externalist content, is constituted by that concept's referent in the concept bearer's actual environment. Epistemic content, on the other hand, is constituted solely in virtue of that concept's role in reasoning and thought. This latter type of content is knowable *a priori*, and is the proper target of conceptual analysis.

In Chapters 1 and 2, I will examine two alternate theories of what the target of intuitive analysis is. In Chapter 1, I'll examine extra-mentalist views on the target of intuitive analysis according to which analysis targets “outside-the-head non-psychological entities”. In Chapter 2, I'll examine mentalist views on the target of inquiry according to which intuitive analysis targets “in-the-head psychological entities”.⁴ Each account of the target of intuitive analysis succeeds where the other fails. Extra mentalist accounts of the targets of inquiry are very good at explaining how the targets of intuitive analysis provide truth conditions for intuitive judgments, but are not good at explaining how these same judgments reliably indicate their targets. Mentalist accounts of the target of inquiry are very good at explaining how these judgments reliably indicate their targets, but not good at explaining how the targets of inquiry can provide truth conditions for these same judgments. Considering the shortcomings of each view will clear the way for a two-dimensional account of conceptual content, according to which intuitive analysis targets a concept's epistemic content. A concept's epistemic content provides truth conditions for the judgments made about cases that are relied on in conceptual analysis, and is reflected in our intuitive judgments when these same judgments are based on rational reflection.

In Chapter 4, I'll motivate the view that intuitive analysis targets a concepts epistemic content. Accepting this view will provide resources for developing an explanation of compellingness, which is an integral part of the method used by conceptual analysts. The new skepticism about intuitive judgments has shown that there are reasons for doubting the legitimacy of conceptual analysis, over and above the critiques of analyticity and a priori

⁴ The Extra-mentalist/Mentalist terminology appears in A. Goldman and J. Pust, 1998, 183. It also appears in Alexander, Mallon and Weinberg, 2010, 299.

knowledge that plagued conceptual analysts in the past. It challenges philosophers who use that method to provide a deeper understanding of how the method works. Chapter 4 shows that the defenses of conceptual analysis brought forward in response to the worries about analyticity and a priori knowledge can be developed to answer the new skepticism as well as the old.

Chapter 1

Extra-mentalism, Calibration, Mitigation, and Knowing One's Instrument

1.1 Extra-mentalism and Conceptual Analysis

Goldman and Pust differentiate between two stances one might take toward the target of intuitive analysis. Extra-mentalists hold that philosophy is continuous with the natural sciences in that both are directed toward gaining knowledge of “outside the head – non-psychological entities”. Mentalists deny this, holding instead that philosophy is, at least in part, an armchair friendly activity, directed toward gaining knowledge of “inside-the-head psychological entities”, or concepts (1998, 183).

Extra-mentalists typically do not consider the decision over which project to pursue to be a matter of taste or preference. When Kornblith writes that “epistemologists ought to be concerned with the nature of knowledge, not the concept of knowledge, the proper subject matter of ethics is the right and the good, not the concepts of the right and the good” he is not expressing his preference for one kind of inquiry over another, but rather expressing a normative judgment about what it is that philosophers should do (1998, 133). Among extra-mentalists, this normative judgment about what philosophers should do is typically accompanied by an outright disdain for those philosophers who choose to pursue the study of concepts. Kornblith writes, “in the case of philosophy of mind, for example, I believe the armchair approach has been disastrous... I would make similar suggestions about

epistemology” (2009,112) Even Williamson, who defends armchair methodologies, considers the study of concepts to be a “vain program” (2007, 280, fn1).

I find this either-or approach baffling. There is room in philosophical inquiry for the pursuit of many types of projects. One of these is a viable program of conceptual analysis. This chapter is not the place to argue for the need or legitimacy of such a practice, although these arguments will be presented in later chapters. Here, I want to present a charitable interpretation of the motivation for rejecting conceptual analysis and see what kinds of methodologies extra-mentalists take up and what can and can’t be achieved through use of these methods.

1.2 Intuitive Judgments: An Experimental Analysis

We can begin to understand the extra-mentalists disdain for conceptual analysis by noting that while conceptual analysts take intuitive judgments as reflecting underlying concepts, extra-mentalists see the deliverances of intuitions very differently. It is fair to take the following quote from Devitt as being representative “intuitions are parts of an empirical, fallible, and certainly inadequate set of folk opinions or, more pretentiously, “folk theory,” the linguistic wisdom of the ages” (1994, 547). Stitch holds a similar view about intuitive judgments, holding that they result, in large part, from the unreflective internalization of cultural norms that have been handed down from generation to generation. Because of this, theorizing that relies on them is “profoundly conservative” and should not be trusted (1990, 119).

Such a view may be motivated by the failures of the definitional account of concepts. Indeed, one might reasonably doubt that we are capable of providing a complete set of necessary and sufficient conditions for any ordinary concept. Further motivation for such a view comes from empirical data showing widespread inter-subjective variation in intuitive judgments. Below, we'll look at two such studies. These studies indicate that intuitive judgments that have played a central role in shaping the development of theorizing in philosophy of language and epistemology are problematically sensitive to cultural influences.

In an influential study, Machery, Mallon, Nichols and Stich (2004) experimentally tested people's intuitive judgments about certain hypothetical cases that have played a significant role in debates in the philosophy of language. In particular, these debates focus on which theory of reference is the correct one. There have been two main historical contenders for this title: the descriptivist theory of reference and the causal-historical theory of reference. Machery, Mallon, Nichols, and Stich present the descriptivist theory of reference as encompassing two major commitments as follows:

D1. Competent speakers associate a description with every proper name. This description specifies a set of properties.

D2. An object is the referent of a proper name if and only if it uniquely or best satisfies the description associated with it. An object uniquely satisfies a description when the description is true of it and only it. If no object entirely satisfies the description, many philosophers claim that the proper name refers to the unique individual that satisfies most of the description (Searle 1958, Lewis 1970). If the description is not satisfied at all or if many individuals satisfy it, the name does not refer. (2004, B2-3)

Likewise the authors understand the causal-historical theory of reference as encompassing two major commitments as follows:

C1.A name is introduced into a linguistic community for the purpose of referring to an individual. It continues to refer to that individual as long as its uses are linked to the individual via a causal chain of successive users: every user of the name acquired it from another user, who acquired it in turn from someone else, and so on, up to the first user who introduced the name to refer to a specific individual.

C2.Speakers may associate descriptions with names. After a name is introduced, the associated description does not play any role in the fixation of the referent. The referent may entirely fail to satisfy the description. (2004, B3)

Among philosophers of language there is a general consensus that Kripke, who presented novel cases invoking intuitive judgments that run counter to those that would be predicted by the descriptivist theory, has refuted the descriptivist theory of reference. Machery, Mallon, Nichols and Stitch, present two such cases from Kripke (1980):

The Gödel Case

Suppose that Gödel was not in fact the author of [Gödel's] theorem. A man called 'Schmidt' (...) actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the [descriptivist] view in question, then, when our ordinary man uses the name 'Gödel', he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description 'the man who discovered the incompleteness of arithmetic'. (...) But it seems we are not. We simply are not. (2004, B3-4)

The Jonah case

Suppose that someone says that no prophet ever was swallowed by a big fish or a whale. Does it follow, on that basis, that Jonah did not exist? There still seems to be the question whether the Biblical account is a legendary account of no person or a legendary account built on a real person. In the latter case, it's only natural to say that, though Jonah did exist, no one did the things commonly related to him. (2004, B4-5)

Kripke's claim here is that ordinary intuitive judgments invoked by consideration of these cases support the causal-historical theory of reference rather than the descriptivist theory of reference.

Machery, Mallon, Nichols, and Stitch design an experiment to test for cultural influences on judgments about these types of cases. The authors design four cases, half

modeled on Kripke's Jonah case, and half modeled on Kripke's Gödel case. They then give these cases, counterbalanced for order, to 31 (Western) undergraduates at Rutgers University and 40 (Chinese) undergraduates at Hong Kong University. In response to each case, participants were asked to choose one answer from two potential answers, one potential answer supporting the causal-historical theory of reference and the other supporting the descriptivist theory of reference. The experimenters coded responses by assigning a score of 1 to answers consistent with the causal-historical theory of reference and a score of 0 to answers consistent with the descriptivist theory of reference. Scores were then summed so the cumulative score could range from 0-2.

Westerners mean score for Gödel type cases was 1.13 with a standard deviation of .88, and 1.23 with a standard deviation of .84 for the Jonah type cases. Chinese participants mean score for the Gödel type cases was .63 with a standard deviation of .96, and 1.32 with a standard deviation of .76 for the Jonah type cases. An independent samples t-test was used to confirm that there was a significant difference between Western and Chinese respondents in the Gödel type cases, but not in the Jonah type cases. The authors consider various explanations for this latter result in their paper.

"Setting out the Jonah cases precisely requires a lengthy presentation ... so it's possible that our probes were simply too long and complex to generate interpretable data. Another, more interesting possibility hinges on the fact that in the Jonah cases, the descriptivist response is that the speaker's term fails to refer. It might be that for pragmatic reasons, both the Westerners and the Chinese reject the uncharitable interpretation that the speaker is not talking about anyone" (2004, B8)

Despite these concerns about the Jonah case, the results strongly suggest that intuitive judgments invoked by one of the most influential cases in the philosophy of language, the Gödel case, are influenced by one's cultural background. The authors summarize their data

as follows: “These results constitute *prima facie* evidence that semantic intuitions vary from culture to culture”. (Machery et. al 2004, B1) The preliminary data thus support Stitch’s earlier speculations: intuitive judgments do not originate in processes that can be trusted to track external fact but rather originate in processes that are sensitive to epistemically irrelevant factors, such as cultural background.⁵

Intuitive judgments invoked by influential cases in epistemology have also been empirically investigated for sensitivity to cultural background. Weinberg, Nichols and Stitch (2001) ran a study, that has become foundational in this area and experimental philosophy more generally, to test whether “epistemic intuitions vary from culture to culture” (437). They are interested in people’s intuitive judgments concerning Gettier style cases, the most well-known and influential cases in epistemological theorizing. In keeping with Gettier style cases, the authors design a case that is centered on an agent who forms a true belief that is supported by an evidentiary source that is generally trustworthy, though it happens that in this particular case, the evidence that appears to the agent to support her belief, actually does not support that belief, and her having a true belief is instead a matter of luck. In such a

⁵ While this study supports the view that intuitive judgments vary relative to cultural background, it was also found that intuitive judgments exhibit intra-cultural variation. Discussing the results of their study, Mallon, Machery, Nichols and Stitch (2009) note “While for each vignette a majority of Americans gave causal-historical responses, in each case a sizable minority of the population (as high as 45% in one case) gave descriptivist responses. Similarly for the Chinese population for each vignette, a majority of Chinese participants gave descriptivist responses, but in each case a sizable minority (in some cases over 30%) gave causal-historical responses.” This suggests that not only are philosopher’s mistaken in thinking that their own intuitions about reference are universal, they are also mistaken in thinking that their own intuitions are representative of their cultural group. For an argument of why we should nevertheless trust philosopher’s intuitive judgments in conceptual analysis see Jackman, “Semantic Intuitions, Conceptual Analysis and Cross-Cultural Variation” (2009).

situation, most epistemologists agree, that the agent lacks knowledge. Despite the convergence in judgments among epistemologists who consider this case, Weinberg, Nichols and Stitch suspect that such judgments are sensitive to epistemically irrelevant factors such as cultural background.

To confirm their suspicions, the authors present the following case to 112 participants at Rutgers University from three cultural backgrounds: 66 Westerners, 23 East Asians, and 23 people from the Indian Subcontinent of Indian, Pakistani, and Bangladeshi descent.

Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it? (2001, 443)

Participants were asked to indicate which alternative they agreed with by circling “Really Knows” or “Only believes”. The results of the study were as follows:

	Really Knows	Only believes
Westerners	17	49
East Asians	13	10
Indian Subcontinentels	14	9

While the majority of Western participants indicated that Bob only believes, the majority of participants in the other two groups indicated that Bob really knows. The initial pattern of responses in Western participants was “actually reversed” in the other two groups (2001, 443).⁶

⁶ There is some worry that the experimental findings reported here are not robust. Recent attempts to replicate these results have failed, see Nagel, San Juan and Mar, (2013), and Turri (2013). For further discussion of these results see Boyd and Nagel (2014). On the other hand, results of the previous study on semantic intuitions do seem to be robust, see Beebe and Undercoffer (2015).

These foundational papers barely scratch the surface of what has become a body of experimental work indicating that intuitive judgments are problematically sensitive to non-epistemically relevant factors. Cultural background has been found to effect intuitive judgments about phenomenal consciousness (Huebner, Bruno, and Sarkissian, 2010), epistemic intuitions have been found to be sensitive to socio-economic background, (Weinberg, Nichols and Stitch, 2001) and order effects (Swain, Alexander and Weinberg, 2008). Ethical intuitions have been found to be sensitive to gender (Zamzow and Nichols 2009). Intuitive judgments about free will have been found to vary relative to personality traits (Feltz and Cokley, 2009) even among those who have expertise in the area (Schulz, Cokely and Feltz, 2011). Moral intuitions have been found to be sensitive to order effects (Petrinovitch and O'Neil, 1996, Liao et al., 2012) and such order effects were found to persist in intuitive judgments amongst those who have an expertise in moral philosophy (Schwitzgebel and Cushman, 2012).

This list of studies is not meant to be exhaustive of the studies that the negative program in experimental philosophy draws on to cast doubt on the reliance on intuitive judgments in inquiry. Even less so, should it be thought of as representative of the positive program in experimental philosophy wherein researchers use empirical studies of intuitive judgments to make positive claims about how the mind works.⁷ Indeed, Knobe has recently reported that only 1.3% of published papers in experimental philosophy in the last five years aim to show that intuitive judgments are unreliable (forthcoming). Still, taken together, these studies provide motivation for rejecting the legitimacy of relying on intuitive judgments in

⁷ For more on the negative and positive programs in experimental philosophy see Alexander, Mallon and Weinberg (2010), Alexander and Weinberg (2007), Kauppinen (2007), Naddlehoffer and Nahamias (2007), and Weinberg (2007).

inquiry and can help to explain the disdain felt by many extra-mentalists about the project of conceptual analysis. In what follows we'll look at how extra-mentalists have tried to overcome these concerns about the reliance on intuitive judgments in inquiry.

1.3 Calibration

The studies canvassed above suggest that intuitive judgments do not originate in processes that can be trusted to track external facts, but rather originate in processes that are sensitive to epistemically irrelevant factors. If so, then, it becomes difficult to defend the reliance on intuitive judgments in inquiry given the extra-mentalists aim of uncovering truths about outside-the-head, non-psychological entities. This type of worry is discussed by Cummins who holds that “philosophical intuitions are artifacts” (1998, 119). The word “artifact” comes from science and refers to an “observation that carries information about the observational apparatus or process rather than about the target” (1998, 116). Cummins holds that intuitive judgments are artifacts that are products of tacit theory. Suppose we took these intuitive judgments and used them to guide the construction of an explicit theory about a traditional topic of philosophical interest such as ‘Knowledge’ or ‘the Good’, the resultant explicit theory would be no more of a reliable guide to the truth about these targets than the implicit theory. According to Cummins, implicit theories are either evolutionary strategies designed to promote successful action or are acquired and biased by the expressed views of those who are closest to you or they are innate. In either case, there is no good reason for thinking that they accurately represent external facts. Cummins concludes, “the most plausible account of the origins of philosophical intuitions is that they derive from tacit theories that are very likely to be inaccurate” (1998, 125).

Cummins suggests that in order to rule out this possibility and show that the intuitive judgments relied on in philosophical inquiry do reliably track external fact, and are therefore trustworthy sources of evidence, we need to adopt a methodological technique used in the sciences; calibration. Calibration requires “that there be, at least in some cases, access to the target that is independent of the instrument or procedure to be calibrated” (1998, 117). In this case, to ascertain the trustworthiness of intuitive judgments we should use an alternative method to access the targets of intuitive analysis to see if the results obtained by using that method are similar to the results we get when we rely on intuitive judgments. If so, then we can rest assured that the intuitive judgments typically relied on in philosophical inquiry are actually truth tracking, and that we are justified in using these in inquiry. Cummins is optimistic about the possibility of calibrating intuitive judgments.

“Personally I am inclined to think that there are, at least in some cases, non intuitive routes to the targets of philosophical intuition. We can give up on intuitions about the nature of space and time and ask instead what sort of beasts space and time must be if current physical theory is to be true and explanatory. We can give up on intuitions about representational content and ask instead what representations must be if current cognitive theory is to be true and explanatory” (1998, 117-18).

Physical theory and cognitive theory are here considered to not rely heavily on intuitive judgments as evidence, but rather on more traditional sources of evidence, such as observations. If we construct theories by utilizing these alternative tools and the theories agree with those that are based on intuitive judgments, then according to Cummins, we will have validated the thesis that intuitive judgments are accurately tracking external fact.

This approach to confirming the trustworthiness of intuitive judgments comes at a significant cost. As Cummins points out, if it is possible to assess the targets of philosophical analysis in a non-intuitive manner, then we can go ahead and dispense with the intuitive

judgments themselves and focus on developing these other methods of inquiry which rely on more generally trusted sources of evidence. Cummins concludes his discussion of calibration as a method for validating the reliance on intuition with the following observation:

“Philosophical intuition, therefore, is epistemologically useless, since it can be calibrated only when it is not needed. Once we are in a position to identify artifacts and errors in intuition, philosophy no longer has any use for it. But if we are not in a position to do this, philosophy should not have any faith in it” (1998, 118).

Those who agree with Cummins about the value of armchair methodologies might be happy to agree that intuitive judgments are epistemologically useless, were it not for the fact that intuitive judgments, as untrustworthy as they are, cannot easily be banned from inquiry. Many extra-mentalists are less sanguine about the prospect of finding a method of inquiry that does not rely on intuitive judgments. According to many of these philosophers, the thing to do when confronted with the untrustworthiness of intuitive judgments is not to ban the use of intuitive judgments in inquiry, but rather to take steps to mitigate this untrustworthiness when theorizing. Below, we will look at a two-step method for mitigating the untrustworthiness of intuitive judgments found in writings by Kornblith and Devitt.

1.4 Mitigation

Kornblith recognizes with Cummins and Stitch, that “background knowledge will play a substantial role in the first pass categorization of samples” and that these judgments are likely to reflect the tacit, naïve theory of the person who is having them rather than accurately reflecting external fact (1998, 134). Unlike Cummins and Stitch who advocate that we give up on intuitive judgments in inquiry due to their untrustworthiness, Kornblith

argues that we must find a place for them in inquiry. It may be true that intuitive judgments originate in bad theory and that this throws doubt on the claim that they are trustworthy guides to external fact. However for Kornblith, “the solution is not to try to return to some pure state of theory independent judgment, before the fall, as it were; rather the solution is to get a better theory” (1998, 135).

Kornblith recommends that we mitigate the general untrustworthiness of our intuitive judgments by using a two-step method. The goal is to use agreement and empirical methods of investigation to progress from our original ill-defined tacit or folk theory, to one which is empirically informed and therefore, a more reliable guide to the truth.

Step 1: Agreement

Suppose we have a set of intuitive judgments which we suspect are products of bad theory and therefore untrustworthy. How do we move toward a better theory? Kornblith suggests that we begin to mitigate the general untrustworthiness of intuitive judgments by focusing only on those that are obvious. What does it mean to say that we must begin inquiry by appealing to obvious cases? What exactly counts as an obvious case? Obvious cases of the phenomenon under study are those that “prompt our intuitions”. They cause us to intuitively judge that example e is an instance of X. Whose intuitions? Clearly, Kornblith is not saying that one should begin with the cases that are obvious to him. “[I]diosyncratic” intuitions should play no role in theorizing. Rather, obvious cases are recognized by the “wide agreement these examples command.” (1998, 134)

The claim that we should limit our focus when theorizing about a particular topic to those intuitive judgments that are widely shared while placing those which are not widely shared in escrow seems consistent with some aspects of standard philosophical practice. For example, Gettier cases have been influential precisely because philosopher's who considered the cases typically agreed in their response to them. Most philosophers who considered the cases, even those coming from theoretical backgrounds that would be consistent with contrary intuitive judgments, agreed that the subject in the hypothetical scenario lacked knowledge. One could argue that it was this widespread agreement that put the cases at the center of epistemological debate for decades to come. Furthermore, when philosophers appeal to intuitive judgments to motivate their preferred view they often speak as if they expect that the intuitive judgment in question will be widely shared. In practice, the fact that an intuitive judgment is widely shared does seem to be an important factor contributing to its trustworthiness. Such considerations support Kornblith's suggestion that "we must be assuming that disagreement with the majority is some evidence of error" (1998, 133)

The flip side of this observation would be to assume that cases on which there is widespread agreement are not in error and that these are trustworthy and do accurately represent external fact. One must be careful here though for this is not what Kornblith is suggesting. While Kornblith recognizes that intuitive judgments must be relied on in the beginning stages of inquiry and that these "must give us some purchase on the phenomenon under investigation" he does not hold, even with respect to those intuitive judgments on which there is widespread agreement, that these are ultimately trustworthy. Despite being widely agreed to, these first pass categorization judgments are "substantially inferior,

epistemically speaking, to those at later stages when theoretical understanding is further advanced” (1998, 134). To advance theoretical understanding we must move on to the second step of the two-step method.

Step two: Empirical Investigation

To further mitigate the unwanted sensitivities to which our intuitive judgments are initially subject, we must undertake a second practice. Once we have our collection of agreed upon judgments we must empirically examine the identified examples to discover their underlying unity. Once we have empirically investigated the examples and come up with a theory of what the underlying unity is, then we can use this theory to identify any errors in our original judgments and to guide our future judgments, making these more trustworthy. It is expected that certain members of the set of agreed upon intuitive judgments that we begin with will be overturned by future empirical work. Intuitive judgments at the beginning stages of inquiry “will change with the progress of theory. What seemed to be a clear case of a given kind in the absence of theoretical understanding may come to be a paradigm case of a different kind once the phenomena are better understood” (1998, 134). Once our judgments are guided by this empirically informed theory then we can trust them to track external fact.

Is this general account of how intuitive judgments function in inquiry satisfactory? Goldman (2007) and Goldman and Pust (1998) are critical of Kornblith’s extra-mentalism and raise a number of questions intended to show its limitations. First, Goldman and Pust point out that the assumption that intuitive analysis actually does target natural kinds is controversial. “Presumably something qualifies as a natural kind only if it has a prior essence, nature or character independent of anybody’s thought or conception of it. It is questionable,

however, whether such analysanda as knowledge, justification, and justice have natures independent of our conception of them" (1998, 186-7). Suppose though, that we took Kornblith's approach here, and suppose further that we isolated a set of widely shared intuitive judgments concerning typical subjects of philosophical interest, such as "the Good" or justice or knowledge etc. The two-step method of inquiry suggests that we then take these examples, and using the best tools of empirical science available, investigate whether or not they share a common underlying nature. It might turn out that our investigations reveal that these examples do not share an underlying nature. Then, according to the natural kinds approach, we must conclude that we are not really theorizing about anything of substance. Goldman and Pust object to this understanding of things. "In our opinion, the lack of natural kind status would not place the topics of knowledge, justification or justice outside the scope of philosophical analysis. Nor do we think that the corresponding predicates should be abandoned if they fail to pick out natural kinds" (1998, 187). To have such topics excluded from serious inquiry seems like an unhappy development, but it is a possibility that the natural kinds proponent must take seriously.

The same authors point to a further worry about the natural kinds approach. While this approach explains the reliance on intuitive judgments, it does not do well at explaining the reliance on intuitive judgments about hypothetical cases and yet the latter are most relied on in philosophical inquiry.

"A ubiquitous feature of philosophical practice is to consult intuitions about merely conceivable cases. Imaginary examples are treated with the same respect and importance as real examples. Cases from the actual world do not have superior evidential power as compared with hypothetical cases. How is this compatible with the notion that the target of philosophical inquiry is the composition of natural phenomenon?" (Goldman 2007, 8)

In response, Kornblith suggests that the philosophical method must be “modified” to get rid of the reliance on hypothetical cases (2007, 9).

It is likely that the loss of this methodological tool will be deeply felt by many who engage in the standard philosophical practice of relying on intuitive judgments in inquiry. However, there are other versions of extra-mentalism that have much in common with Kornblith’s but are able to provide more satisfying responses to the concerns raised by Goldman and Pust.

1.5 Functional vs. Natural Kinds

Devitt’s account of the target of philosophical analysis goes some way in addressing the concerns that Goldman and Pust raise with respect to Kornblith’s view. Like Kornblith, the “ultimate method” of inquiry that Devitt endorses consists of a two-step process, similar to that discussed above. If we want to answer the question “What is F?” then, according to Devitt, we proceed as follows “First we identify some apparently uncontroversial examples of F’s and not F’s. Second we must examine the examples to determine the nature of being an F”. Once again, the second stage of this process is a “straightforwardly scientific one”, but intuitive judgments help in the initial identification of examples (1994, 562). Once again, judgments made at the first stage of inquiry are “seriously incomplete”, while those made at the second stage of inquiry are epistemically superior and can cause us to reject the less trustworthy judgments made at the first stage.

The main difference between Kornblith's and Devitt's views concerns the actual targets of inquiry. While both hold that the targets of philosophical analysis are outside the head non-psychological entities and therefore qualify as extra-mentalists on Goldman and Pust's definition, there are nevertheless significant differences in their views of what the actual targets of intuitive analysis are. For Kornblith, the ultimate target of inquiry is a natural kind: there are naturally existing kinds in the world and figuring out what these kinds are is the object of our study. Devitt on the other hand, holds that classes of individuals are unified functionally. The unifying element of a particular kind is not to be found independently existing in nature, but rather in our purposes. It is because each member of a set function to fulfill a specific purpose as defined by us, that it is a member of the set. Intuitive analysis and philosophical inquiry more generally is geared toward uncovering which property a thing must have in order to fulfill the purposes that we have identified. The properties are externally existing facts, and so this is a type of extra-mentalism.

For example, in the paper "The Methodology of Naturalistic Semantics" Devitt argues that in order to find out what meanings are we must first understand which purposes are served by ascribing them to thoughts and utterances. To figure this out we must consult our actual practices of ascribing meanings. But how do we know when we are ascribing meanings? We appeal to our intuitive judgments of the form "example x is an instance of F". That is, we appeal to uncontroversial examples of the kind we are interested in, here uncontroversial examples of meaning ascriptions. Once we have a set of these ascriptions we can describe the purposes that are served by ascribing meaning to thoughts and utterances. Reflecting on the uncontroversial cases, Devitt concludes that we ascribe meanings for two

main purposes: to explain and predict behavior and as guides to reality. Devitt recognizes that there may be more purposes for which we might ascribe meanings, but claims that these two are “fundamental” and that “we expect our ascriptions of meaning to serve them at least” (2004, 562: 552). Once we have described the fundamental purposes inherent in our practice of ascribing meanings, then we will have a theory of what a meaning is: a meaning will be anything which has the property of fulfilling these two purposes.

This is an important difference between the two accounts of the target of inquiry that shouldn’t be downplayed. The liberalism inherent in Devitt’s view of what can qualify as a target of inquiry makes it appealing because it helps to dampen the worries Goldman and Pust point to in Kornblith’s type of extra-mentalism.

For example, Devitt is upfront about the possibility that our investigations will reveal that there is no real underlying phenomenon unifying our examples that is worthy of study. “We may even conclude that none of the things identified as F’s are really F’s, for example there are no witches.” (1994, 562) It is therefore an open question whether or not inquiries into topics such as the Good, what is just, what is justification and what is morally right, etc. turn out to be legitimate. However, this worry is much less pressing on Devitt’s view. While it is plausible that topics of philosophical interest do not represent natural kinds, existing independently of anybody’s thought or conception of them, it is less clear that these do not represent properties as Devitt defines them. Because properties are defined in terms of our purposes there is less chance that we will find that our inquiries do not target anything of substance. Successful practices vouch for the existence of a functional kind.

Not only is there less of a worry that we will find that our inquiries have no targets, this account also has an easier time explaining the reliance on hypothetical cases. Reflecting on usage here identifies the targets of inquiry and responses to hypothetical cases do manifest our use of terms. “Instead of real experiments that confront the expert with the phenomenon and ask her whether they are F’s, we confront her with descriptions of the phenomenon and ask her whether she would say that they are F’s. These thought experiments provide valuable clues as to what the expert would identify as an F and Not F.” (1994, 563) Here “thought experiments have the same empirical status as the results of real experiments” (1994, 564). The capacity to find a place for thought experiments within the two-step method makes that method more attractive than the previous type of extramentalism.

When the two-step method of mitigation is put in the context of an inquiry that is directed toward gaining knowledge of functional kinds, rather than natural kinds, it overcomes many of the problems that Goldman and Pust discuss. In section 1.7, we will see reasons for doubting that even this refined version of the two-step method can provide us with a satisfactory theory of its target domain. In order to see why this is so, it will be helpful in the next section, to review Weinberg, Crowley, Gonnerman, Vanderwalker and Swain’s critical discussion of Cummins’ thesis that calibration leads to the epistemological uselessness of intuitive judgments. Weinberg et al. challenge this and develop a methodological technique for dealing with the original untrustworthiness of intuitive judgments, a technique known as “extrapolative calibration”. Extrapolative calibration, they

argue, does not lead to conclusion that intuitions are epistemologically useless. Once we get this view on the table, we can see that the same problems that Weinberg et al. claim to plague the method of extrapolative calibration also plague the two-step method that is found in Devitt's writings.

1.6 Exhaustive vs. Extrapolative Calibration

Weinberg et al. (2012) argue that it is possible to calibrate intuitive judgments without simultaneously making them epistemologically useless. According to Weinberg et al. while “exhaustive calibration”, the kind of calibration associated with Cummins, leads to the conclusion that intuitive judgments are epistemologically useless, “extrapolative calibration” does not.

To understand how extrapolative calibration does not lead to the conclusion that intuitive judgments are epistemologically useless, let's return to Cummins theory of “exhaustive calibration”. Suppose that we have come up with a theory about a domain based on consideration of intuitive judgments about cases. Let's call this theory IT. Because, according to Cummins, intuitive judgments about cases are likely “artifacts” and not trustworthy sources of evidence, we must find an alternative theory about our domain that does not rely on intuitive judgments to calibrate the first. Let's call this theory AT. Because AT is not based on intuitive judgments but on more generally trusted sources of evidence, we can trust AT to tell us about our target domain. We can then use our knowledge of the target domain to calibrate the intuitive judgments that ground IT. As noted above, even if IT

is found to be trustworthy, this is a hollow victory for those who wish to rely on intuitive judgments in inquiry. For now that we have an alternative method for accessing our target domain we no longer need to rely on intuitive judgments, calibrated or not.

Weinberg et al. hold that Cummins has slipped an un-argued for assumption into his argument for the epistemological uselessness of intuitive judgments, an assumption that should be rejected. On Cummins view, AT gives us independent access to our target domain. Following Weinberg et al. I'll refer to the range of the targeted domain that is accessible via AT as the “certified base”. Knowledge of the certified base will allow us to calibrate the intuitions that ground IT. According to Weinberg et al., Cummins mistakenly assumes that the certified base accessed by AT is exhaustive of the target domain. If this is the case, then once we have calibrated, intuitive judgments no longer serve any purpose because “there is no room left for an expansion of our epistemic reach” (2012, 262). Weinberg et al. suggest that we reject Cummins assumption and instead think of the certified base as only a subset of the target domain. If so, then once we calibrate we can use the newly calibrated intuitions to increase our understanding of the target. Future theorizing will “project out from the certified base to territory that is not otherwise epistemically reachable” (2012, 263). On this extrapolative account of calibration, calibration will not lead to the epistemological uselessness of intuitive judgments.

Not only do Weinberg et al. think that Cummins conclusion concerning the epistemological uselessness of intuitive judgments is mistaken, they do not share his optimism about finding an independent, non-intuitive theory of the target domain.

Cummins, you will remember, reports “I am inclined to think that there are, at least in some cases, non-intuitive routes to the targets of philosophical intuition” (1998, 117-18) But Weinberg et al. are more pessimistic.

“Where such theories or scientific results can be found, we should look to exploit them. Nonetheless, we fear that their contribution, albeit real, will be minimal. For most areas of philosophy, one simply does not, and probably will never find empirical theories that overlap the target domain. There does not look to be an empirical science of metaphysical necessity, or the moral good.” (2012, 266-67)

It follows that

“To the extent that we are going to calibrate intuitions in any wholesale fashion, we are going to need to do so in large part on the basis of our best philosophical instrument, which unfortunately is intuition itself. This, needless to say, makes things complicated. For we can use intuition as a successful calibrator for intuition only provided we have some reason to extend a heightened degree of trust to some subset of intuitions under consideration.” (2012, 267)

Weinberg et al. go on to identify “consensus intuitions” as a candidate subset of intuitions to which philosophers might extend their heightened trust (2012, 267).⁸

We are already familiar with this move. It is precisely this methodological technique appealed to in the first step of the two-step method for mitigating the original untrustworthiness of intuitive judgments presented by Kornblith and Devitt. Once we have limited our base to the intuitions that are characterized by widespread agreement, we can then undertake the second step, and appeal to empirical science to further refine our base. Empirical science here provides what Weinberg et al. refer to as a “welcome, further source of a certified basis” (2012, 282, fn 16).

⁸ Other potential candidates discussed by WCGVS include ordinary intuitions, reflective intuitions, expert intuitions, and clear and forceful intuitions. It is argued that all of these candidate subsets face problems.

I do not know if those who advocate the two-step method of mitigating the original untrustworthiness of intuitive judgments would be comfortable with the label extrapolative calibration. Besides the fact that both programs grant a level of *prima facie* trust to consensus intuitions, another reason for holding that these projects are homogeneous is that extrapolative calibration, unlike exhaustive calibration, does not commit one to the claim that intuitive judgments are epistemologically useless, a claim that those who advocate the two-step method of mitigation deny. On the other hand, a possible reason for thinking that these projects are not homogeneous concerns the extrapolation itself. While I think that this extrapolation is important, I do not know that everyone would agree. Of course, everyone wants their theories to apply to future cases, but one might think that this future directed extension of theory to cases is accomplished via inductive inference, without worrying much about extrapolative inference, or the extension of theory from clear to unclear cases. Nevertheless, it is common to recognize a need for extrapolative inference. Cummins, for example, holds that one main epistemic goal of theorizing is to “provide a bridge that transfers our relative certainty about the clear cases to an unclear case” (1998, 114). Especially in philosophical inquiry, where philosophers’ are typically interested in just what to say about these unclear cases, the usefulness of a theory that does not fulfill this bridging function is questionable.

The problems discussed in the next section arise, I think, for any attempt to use a theory grounded in consensus intuitions to bridge the gap between these cases and cases that invoke non-consensus intuitions. If one is not interested in bridging this gap, then they

needn't be worried about these problems, but, in this case, the theory in question will be unsatisfying to many who are.

1.7 Theoretical Bridges and the Need to Know One's Instrument.

I want now, to turn to problems confronting any project that relies heavily on consensus intuitions as a certified base for calibrating non-consensus intuitions. The main problem Weinberg et al. focus on concerning the prospect of extrapolating from consensus cases to non-consensus cases is as follows:

“Recall that calibration involves making the device’s reports fit with what is known through the certified basis. Such a comparison can only be done in cases covered by both the device’s reports and the certified basis. But with consensus, the certified basis consists of cases where the reports of everyone’s device agree. This means that any relevant device’s reports – in this case, any competitor set of intuitions under particular circumstances – will trivially fit with the certified basis, since that is part of what it is to be part of the certified base in this case (i.e. our certified basis is one that is, *prima facie*, equally easily generated by devices that stand, in other circumstances, in substantial disagreement). So we need some grounds for giving preference to one projection from the consensus data to another one” (2012, 269-70).

We can make this worry clearer by way of an example.

Suppose we are interested in developing a theory of moral obligation. According to the methodology we are considering, we first must identify obvious or uncontroversial cases where we consensually judge there to be a moral obligation. Second, according to the two-step method, we further refine the base by using tools from empirical science. Ideally, this will result in a theory of moral obligation. The question we need to consider is whether we now have the theoretical resources to build a bridge from these consensus cases, to the non-consensus cases. One well-known and uncontroversial case in which there is a moral

obligation can be found in Singer's writings. Singer asks us to consider the following case: "I am walking past a shallow pond and see a child drowning in it." In this case, Singer judges that "I ought to wade in and pull the child out" (1972, 231). Most people who consider this case agree with Singer that while saving the child "will mean getting my clothes muddy" this outcome is insignificant compared to the death of a child and that there exists a moral obligation to save the child (1972, 231). Now, suppose we examined this clear case of a moral obligation to construct a theory of moral obligation. Singer suggests that the case supports the following theory of moral obligation: "if it is in our power to prevent something very bad from happening, without thereby sacrificing anything morally significant, we ought morally to do it" (1972, 231). This theory of what counts as a moral obligation does correctly account for the case considered. However, there are other theories that are also consistent with the case just identified. Arthur, for example, examines the case and finds that it supports a different theory of what a moral obligation is. For Arthur, the case shows that a person is morally obliged to act "when there is no substantial cost to themselves, that is, when what they are sacrificing would not mean significant reduction in their own or their families' level of happiness" (2010, 145).

Both theories are arrived at by considerations involving the clear and agreed upon case of a moral obligation. And both theories are consistent with the case identified. However, each theory answers the question of what we ought to say in response to unclear cases differently. For example, suppose we want to extend the theory to answer the question of whether one is morally obligated to send one's extra-income (that is, income that is not used to meet any vital need) to victims of famine. If we use Singer's theory to answer the

question of whether there is a moral obligation in this case then we are lead to the answer that indeed, one ought to recognize donating one's extra-income as a moral obligation. That is because in this case, like the last, it is in our power to prevent something very bad from happening (i.e. death from starvation) without sacrificing anything morally significant (i.e. the material goods that would be affordable with the extra income). However, if we use Arthur's theory to answer the question of whether there is a moral obligation in this case we are lead to a different answer. According to this theory, a moral obligation to act exists when what one would sacrifice in acting will not significantly reduce the happiness of oneself or one's family. It is possible that sacrificing one's extra-income will significantly impact the happiness of oneself and one's family. If so, then there is no moral obligation in the case we are considering according to Arthur's theory.

This example illustrates the difficulties associated with extrapolating from clear, or consensus intuitions, to unclear cases.⁹ The consensus intuitions provide no guidance when choosing between competing theories. The theories that are based on consensus intuitions will not allow one to meet the goal that has been identified as important in theorizing; these will not function as a bridge, allowing for a transfer of confidence from the set of clear cases to an unclear case. Rather, the multiplicity of theories that are consistent with the set of clear cases has the opposite effect of undermining any confidence one might have in any of them.

⁹ It should be noted that no one is suggesting we start with a single case and attempt to use it ground a theory. Rather, we should start theorizing with a set of cases on which there is widespread agreement. Enlarging the set of data in this way does not help to circumvent the problem we are discussing. I'm simplifying here to illustrate the more general problem.

Despite pointing out this limitation on methodologies that attempt to calibrate or mitigate the untrustworthiness of intuitive judgments by appealing to consensus intuitions, Weinberg et al. hold that there is still hope for those who wish to rely on this method in inquiry. Initially, this hope came from the possibility of extrapolative calibration, which did not see the epistemological uselessness of intuitive judgments as following necessarily from attempts to calibrate them. Extrapolative calibration, however, comes with its own set of issues, which also threaten to infect the two-step method of mitigation that was presented above as an alternative to calibration. Still, Weinberg et al. argue that these considerations should not push us to abandon all hope, rather such considerations open up possible avenues to pursue. In particular, the proper way to react to these findings, according to Weinberg et al., is to gain a better understanding of the instrument that produces intuitive judgments.

“The real lesson here is not that consensus cases cannot be of use in calibration. It is that they cannot be of use in calibration *without the assistance of a theory of the instrument that will license particular extrapolations*. A theory of the instrument that does *not* arise naturally from a focus on consensus intuitions and that does not appear, at least to this point, to have been generated independently by those promoting the value of consensus” (2012, 270).

The idea here is that if we knew more about the instrument that is producing intuitive judgments, when it is likely to produce errors for example, than we would be better able to choose among the possible extrapolations from the consensus data. Admitting that there is a need for this this might be thought of as concessionary for extra-mentalists who are directed toward moving out from intuitive judgments to knowledge of the world. Based on what has been said above, in order to do so extra-mentalists will have to look inward as well, to investigations of the instrument that is producing intuitive judgments. Still, it would

be a mistake to see Weinberg et al.'s real dialectical opponent as the extra-mentalist. Weinberg et al. remark that their "cautious and limited optimism" about the ability to rely on intuitive judgments in inquiry is not one that "proponents of a pure and autonomous armchair for the philosopher will find satisfying" (2012, 259). This is because knowing one's instrument, on Weinberg et al.'s view is not something that can be accomplished from the armchair. To gain a better understanding of our instrument, "minimally, we need to come up with better folk theories of intuition. Failing that we will have to fall back on the full dress cognitive psychology approach". I have no doubt that there is a lot to learn from cognitive psychology and positive experimental philosophy about how intuition works, but I want to take up Weinberg et al.'s challenge from a different more armchair friendly perspective.

What, if anything, can we say about how the intuitive judgments work from the armchair? Weinberg et al. write

"We should expect that for any instrument – including intuition- that even at its best, it will have zones in which it will be error-prone, and bounds outside of which it cannot be expected even to be minimally reliable...much of the game is figuring out what the appropriate restrictions are" (2012, 263).

If this is the game, then in later chapters, I'll show that it can be played from the armchair. In Chapter 3, I'll develop an account of conceptual analysis, found in Chalmers' writings, according to which there is a component of conceptual content that can be known *a priori*. In Chapter 4, I'll derive from this view one constraint that conceptual analysts face when attempting to use intuitive judgments about cases when theorizing. Armchair reflections on the account of conceptual analysis presented in Chapter 3 will put us in a position to say something about the kinds of cases that can be most profitably relied on in conceptual analysis in Chapter 4.

Chapter 2

Mentalism, Reliability, and the Ubiquity of Variation

2.1 Mentalism

The last chapter looked at the role of intuitive analysis in inquiries directed toward uncovering knowledge of external facts. Many philosophers who rely on intuitive judgments when theorizing do not agree with the thesis advocated in the last chapter, that such inquiries should be directed toward gaining knowledge of “outside-the-head non-psychological entities”. This chapter will question whether intuitive judgments function as proper sources of evidence for inquiries directed toward uncovering knowledge of “in-the-head, psychological entities”, or concepts.

To forestall misunderstandings, it must be noted at the outset that the word ‘concept’ as it is used here, does not come with any built in assumptions about what the structure of a concept is. There is widespread disagreement about this, both between and within philosophy and psychology. However, one thing on which most people working in this area agree, is that the classical theory of concepts according to which, as Cappelin puts it, concepts are “neat little bundles of necessary and sufficient conditions inside the speaker’s heads” (2012, 209) is a non-starter. Keeping this in mind, we can begin this chapter with a fairly neutral understanding of concepts that is expressed in the following passages:

“The term concept refers to a psychological structure or state that underpins a cognizer’s deployment of a natural language predicate. Thus Jone’s concept of apple is the psychological structure that underlies her deployment of the predicate apple, and Jone’s concept of knowledge is the psychological structure that underlies her deployment of the predicate ‘knows’ (or ‘has knowledge’)” (Goldman and Pust, 2007, 188).

“A concept of X is a body of knowledge about x that is stored in long term memory and that is used by default in the processes underlying most, if not all, higher cognitive competences when they result in judgments about X... the knowledge that is stored in our concept of X is preferentially available when we think, reason etc. about X. So to speak, it spontaneously comes to mind.” (Machery, 2010, 602)

Because concepts are here thought of as psychological kinds, one might wonder what role the method of intuitive analysis plays in inquiries targeting them. After all, psychologists, and cognitive scientists more generally, have developed numerous experimental methods for studying psychological kinds that do not rely heavily on intuitive judgments about hypothetical cases. Still, Goldman (2007) and Goldman and Pust (1998) argue that taking a naturalistic view of concepts does not preclude one from relying on intuitive judgments as evidence when theorizing about them. In what follows, we will look at their argument for the claim that intuitions about cases provide a basic source of evidence when theorizing about concepts, as they understand them, and see how this argument stands up to recent skeptical challenges against the reliance on intuitive judgments in philosophical inquiry.

2.2 Reliabilism

The strategy employed by Goldman (2007) and Goldman and Pust (1998) for showing that intuitive judgments do indeed count as a proper source of evidence when theorizing about concepts is to first figure out what it is that makes a potential source of evidence count as a proper source of evidence and then to show that intuitive judgments possess this evidence making property.

Commonly, mental states are treated as evidence for the truth of singular claims, and in the construction of general theories. For example, my perception of there being a cup on the table in front of me is generally taken as evidence for the truth of the claim “There is a cup.” Furthermore, if one were interested in formulating a general theory of cups, then judgments, such as “This is a cup”, would play an important role in guiding the construction of this general theory. A general theory of cups should be comprehensive of most, if not all, of our “This is a cup” judgments. In this way, mental states can provide evidence for or against more general theories.

Why does the mental state of perceiving a cup provide evidence for the truth of the statement “There is a cup”? One plausible answer is that the mental state, in this case, the perception of the cup, is directly caused by the presence of the actual cup. Another plausible answer suggests that the evidential relationship between a given mental state with content p : $M(p)$, and the truth of “ p ” is best explained by appealing to a counterfactual dependence between the two, where this dependence is grounded in the truth of claims such as “If “ p^* ” were true then $M(p^*)$ ”. For example, “If the object in a person’s visual field is red, the person would seem to see that something is red” and “If the object in a person’s visual field is yellow, the person would seem to see that something is yellow” (Goldman and Pust 1998, 180).

Following Goldman and Pust (1998) and Goldman (2007) we can accommodate both views by focusing on a more minimal condition for specifying when a mental state qualifies as a legitimate source of evidence. A mental state of type M will qualify as a basic

evidential source if mental states of type M are reliable indicators of the truth of their contents. The reliable indicator constraint can be satisfied by assuming a causal relation and/or a counterfactual dependence between mental states and that which they are supposed to provide evidence for, but is not committed to the existence of either of these. “The reliable indicatorship requirement is simply the requirement that when M-states occur...their contents are generally true” (1998, 180).

The reliable indicator constraint, though minimal, can be used to explain why certain types of mental states are generally taken as basic evidential sources. Mental states, based on perception, memory and testimony all qualify as basic evidential sources according to the reliability constraint. To show that intuitive judgments are basic evidential sources then, we must ask whether they are reliable. Goldman suggests that we think of the reliability constraint as a negative constraint, advising us to treat a potential source of evidence as a basic source of evidence only when “there is no reason for strongly doubting that it is reliable” (2007, 5). To show that intuitive judgments reliably indicate their targets we have to show that there is no good reason for thinking that they do not.

2.3 Remoteness

Goldman considers two potential reasons for doubting the reliability of intuitions. The first potential reason for doubting the reliability of intuitions comes from reflections on occasions when other sources of evidence are unreliable. We are not typically willing to trust that a potential source of evidence is reliable when it is far removed from its target in space

or time. Remoteness of the target provides a reason to doubt an evidential sources' reliability. The second potential reason for doubting the reliability of intuitive judgments comes from experimental investigations into the nature of intuitive judgments suggesting that these are characterized by a high degree of inter-subjective disagreement. If judgments about cases reliably reflect underlying concepts, and as seems plausible, we typically share similar concepts, then we should expect that people's judgments about cases converge. But this expectation is inconsistent with experimental data showing that people's judgments about the same case often vary. Goldman (2007) and Goldman and Pust (1998) have discussed both potential reasons for doubting the reliability of intuitive judgments, arguing that they do not in fact give us a strong reason for doubting that intuitive judgments reliably reflect their targets. The third potential reason for doubting the general reliability of intuitive judgments discussed below is not explicitly discussed by these authors. Experimental studies into intuitive judgments suggest that these are characterized by a high degree of intra-subjective variation. If the same subject responds differently to the same case on different occasions than this provides a further reason for doubting the reliability thesis. Our discussion of this third potential source of doubt will show that given some modifications or refinements to our understanding of concepts, this potential reason for doubting the reliability thesis also fails as a strong reason for doubting the reliability thesis.

The first potential reason for doubting the reliability of a prospective source of evidence has to do with what the target, or entity that the evidential source is supposed to provide evidence for, is. If the target that a mental state is supposed to provide evidence for is remote, then we are typically less willing to trust that the mental state is reliable evidence.

For example, we tend to be less trusting of a perception, the farther removed we are from the object that is being perceived. We tend to treat memories of long ago events as more suspect than memories of events that occurred a day ago. And testimony tends to be considered less trustworthy the further removed from the original source it is. Considerations involving the remoteness of the target play an important role in determining whether we treat a given mental state as a legitimate source of evidence. In general, the more remote the target, the less likely we are to trust even generally agreed upon sources of evidence, such as perception, memory and testimony, as reliable indicators of the truth of that target. One way then, of easing doubts about whether intuitive judgments qualify as reliable indicators of the truth of their targets is to show that the target of these judgments is not remote.

Following Goldman (2007) and Goldman and Pust (1998) we can take as an extreme example the claim that intuitive judgments are meant to provide evidence in general theories about Platonic forms. Platonic forms exist outside the bounds of space and time, and are so remote from the mental states that are supposed to provide evidence for them, that it is hard to see how the latter could provide evidence for the former. Intuitive judgments here, really do feel more like guesses than a basic source of evidence. Indeed, if we treat any abstract object as the evidential target of intuitive judgments then we will have the same problem. As Goldman puts it “these accounts have too many mysteries, mysteries that undercut any putative reliability needed to support a reflective acceptance of an evidential relationship between intuitional episodes and their targets construed as abstract entities” (Goldman 2007, 7-8).

Goldman holds that there is a direct causal relationship between intuitive judgments and concepts, “although the psychological details remain to be filled in, there is nothing inherently mysterious in their being a causal pathway from personal psychological concepts to application intuitions pertaining to those concepts” (Goldman, 2007, 13). There is no reason, on this view, to strongly doubt that intuitive judgments do reliably indicate the truth of their contents and thereby qualify as a basic evidential source arising from considerations of remoteness.

2.4 Inter-personal Variation and Verbal Disputes

Suppose I have a perception with content P where P can be expressed by the statement “There is a cup on the table in front of me”. I generally take this perception as evidence for the truth of the claim that “There is a cup on the table in front of me”. I fully expect, all things being equal, that if you were to look at the same part of the table you would also perceive that there is a cup on the table in front of me. But suppose you denied this and said that there was no cup on the table in front of me. I would assume that one of us must be mistaken. If a substantial number of my perceptions were disputed in this way, then I might begin to suspect that the processes in which these originate are not reliable processes whose deliverances can be counted on as basic sources of evidence. When a potential source of evidence is characterized by widespread inter-personal disagreement, than that provides a reason for doubting its reliability.

Skeptics about the practice of relying on intuitive judgments as evidence argue that that this is exactly the situation we find ourselves in with respect to our intuitive judgments. We saw in the last chapter how speculations about inter-subjective variation in intuitive judgments have been supported by empirical evidence. We looked at two foundational studies in this area; intuitive judgments that have been at the center of debates in the philosophy of language, and epistemology, have been found to vary relative to cultural background. These foundational studies have been followed up by others suggesting that the finding of inter-subjective variation in intuitive judgments is robust. While further studies would be useful here, the studies undertaken so far show that the characterization of intuitive judgments as inter-subjectively variable is a live possibility. Skeptics of intuitive analysis claim that inter-subjective variation provides a good reason for thinking that intuitive judgments do not reliably indicate the truth of their targets. In response, those who hold that intuitive judgments are reliable indicators of the truth of their targets have argued that such skeptical arguments rest on an interpretation of the data that should be rejected.

According to one plausible interpretation, accepted by those who are skeptical about the reliance on intuitive judgments in inquiry, the data showing widespread inter-subjective variation in intuitive judgments indicates that there is widespread disagreement in intuitive judgments. If this is what the data indicates then skeptics of the evidential value of intuitive judgments can draw attention to the striking dis-analogy between intuitive judgments which display a high degree of inter-subjective disagreement and generally accepted sources of evidence (such as perception) where there is a high level of inter-subjective agreement and

use this dis-analogy in an argument against the reliability of intuitive judgments. Here's how Goldman puts the skeptical argument:

"People often have conflicting intuitions about philosophical cases. One person intuits that case x is an instance of property (or concept) F while another person intuits that case x isn't an instance of property (or concept) F. When such conflicts occur, one of the intuitions must be wrong. If the conflicts are frequent, the percentage of erroneous intuitions must be substantial and the percentage of correct intuitions not so high. Thus, the modest level of reliability of philosophical intuitions doesn't warrant assigning them significant evidential weight." (Goldman 2007, 3)

Sosa considers the negative experimentalists argument this way, writing

"When we rely on intuitions in philosophy, then, in my view we manifest a competence that enables us to get it right on a certain subject matter, by basing our beliefs on the sheer understanding of their contents. How might survey results create a problem for us? Suppose a subgroup clashes with another on some supposed truth, and suppose they all ostensibly affirm as they do based on the sheer understanding of the content affirmed. We then have a *prima facie* problem. Suppose half of them affirm $\langle p \rangle$ while half deny it, with everyone basing their respective attitudes on the sheer understanding of the representational content $\langle p \rangle$. Obviously, half of them are getting it right, and half wrong. Of those who get it right, now, how plausible can it be that their beliefs constitute or derive from rational intuition, from an attraction to assent that manifests a real competence? Not that it is logically incoherent to maintain exactly that. But how plausible can it be, absent some theory of error that will explain why so many are going wrong when we are getting it right?"

"So there will definitely be a *prima facie* problem for the appeal to intuitions in philosophy if surveys show that there is extensive enough disagreement on the subject matter supposedly open to intuitive access" (2007, 102).

Below, we will see how Sosa, Goldman and others have attempted to respond to this argument in such a way as to preserve the reliability of intuitive judgments. The response in question raises the possibility that the argument fails by challenging the idea that data showing widespread variation in intuitive judgments indicates that there is widespread disagreement in intuitive judgments. According to these philosophers, this is only one possible interpretation of the data. There is an alternative interpretation of the data, to which

we will now turn, that does not take it as indicating widespread disagreement in intuitive judgments.

The argument from disagreement is committed to cases of variation representing cases of disagreement. However, many authors have argued that these so called disagreements are actually better understood as verbal disputes. Chalmers has recently written on the nature of verbal disputes and while the details of his analysis are interesting, I want to let the following discussion be guided by what Chalmers refers to as “the most famous case of a verbal dispute in philosophy”. The case is found in writings by William James and is summarized by Chalmers as follows

“A man walks rapidly around a tree, while a squirrel moves on the tree trunk. Both face the tree at all times, but the tree trunk stays between them. A group of people are arguing over the question: Does the man go round the squirrel or not?” (2011).

If we understand this case as one of disagreement then it follows that one of the two possible answers to this question is right, and the other wrong. Settling the dispute will involve finding out which one is which. However, James offers a different approach to settling the dispute that does not treat it as an instance of disagreement. James argues that once we get clear on what both parties mean by the term “go round” we recognize that both parties might be right relative to what they each mean and the apparent disagreement disappears. James distinguishes between two possible senses of the term “go round”. In one sense “go round” means “passing from the north of him to the east, then to the south then to the west, then to the north of him again”, in another sense “go round” means “being first in front of him, then behind him, then on his left, and finally in front again”. If we are concentrating on the first sense of the term, then “obviously the man does go round him” if

on the other hand, we are concentrating on the second sense of the term, then “it is quite as obvious that the man fails to go round him.” James notes that once we “make the distinction... there is no occasion for any further dispute” (1968, 141).

The main upshot of this discussion for present purposes is that sometimes, what at first glance seems to be a disagreement is, in fact, a mere verbal dispute. Furthermore, James’ strategy for diagnosing a verbal dispute is to attend to the various senses, or meanings of the term involved; if we distinguish between different senses of a term, then the conflicting views may both be correct and there is no longer any disagreement.

Certain authors have suggested that purported cases of disagreement that the negative experimentalists point to as undermining the reliability of intuitive judgments are not really disagreements but rather verbal disputes. For example, Sosa, considers the argument from disagreement noting that on reflection, the so called disagreement which is supposed to pose a problem for intuitive analysis, can be regarded as merely a verbal dispute:

“Within the movement itself, one finds a growing recognition that the supposed "intuitive disagreements" may be only verbal.”

“To the extent that experimental philosophy adopts this way of accounting for diversity of verbal intuitive responses, it will avoid substantive clashes in favor of merely verbal disagreement. But once such disagreements are seen to be verbal, the supposed problem for philosophical intuition evaporates” (2007, 103).

Goldman explains:

“If the targets are construed as concepts in the personal, psychological sense, then Bernard’s intuition that F applies to x is evidence only for his personal concept of F, and Elke’s intuition that F doesn’t apply to x is evidence only for her personal concept of F.... Under

this construal of evidential targets, interpersonal variation in intuitions doesn't pose a problem for intuitional reliability because each person's intuition may correctly indicate something about his or her own concept, viz., whether the concept does or doesn't apply to the chosen example." (2007, 13)

On this view, the experimental data showing inter-subjective variation fails to show that intuitive judgments do not reliably indicate the truth of the contents. Rather, the variation in intuitive judgments is here held to reflect the fact that the parties in question have different concepts. If so, then each party's intuitive judgments can still provide evidence for their personal concept. The lesson drawn from the empirical work on intuitive judgments here is that philosophers who deal in intuitive judgments need to revise their assumption that intuitive judgments will be widely shared, or that we all share similar concepts, but the main thesis that intuitive judgments function as reliable indicators of the truth of their contents so that these same judgments can be taken as evidence in theories outlining the contours of those concepts remains untouched by the data showing inter-subjective variation in intuitive judgments.

Negative experimental philosophers have tried to resist the verbal dispute response to the argument from disagreement by noting that it leads to the "absurd conclusion" that we do not really know, of any of our discourse, whether it agrees or disagrees with the discourse of anyone else (Mallon et. al., 2009, 351). According to the verbal dispute response, one can only disagree with another if they belong to the same "intuition group" or if they share a common background concept. But plausibly, we cannot easily tell to which intuition group a person belongs, so we have no way of telling whether we are really disagreeing with someone or simply talking at cross purposes, much like the group of people who were arguing about whether the squirrel goes around the tree prior to James's stepping

in, in the example discussed above. On other occasions, negative experimental philosophers seem willing to accept the verbal dispute response to the argument from disagreement. For example Alexander, Mallon and Weinberg allow that incorporating “a modicum of relativization... may go some way towards defusing the threat of cross-group differences in intuitions – if Asians and Western subjects have different intuitions, then perhaps they just have different concepts” (2010, 302). In any case, a more worrisome threat to the thesis that intuitive judgments reliably indicate the truth of their targets comes from a different type of data indicating intra-personal variation in intuitive judgments, to which we will now turn.

2.5 Intra-subjective Variation

Let's recap. We have been questioning the idea that intuitive judgments are a proper source of evidence for theorizing about psychological concepts. It has been argued that intuitive judgments are a proper source of evidence because they are reliable indicators of the truth of their targets. We have looked at two potential reasons for doubting the reliability thesis and seen that neither of them provides a strong reason for doubting that thesis. We will now discuss a third potential reason for doubting the reliability thesis. The proposed source of doubt is again experimental studies of people's intuitive judgments. We will now turn to a sample of these.

Swain, Alexander and Weinberg (2006) provide an experimental analysis of people's intuitive judgments concerning Lehrer's Truetemp thought experiment. This thought experiment has been commonly taken as evidence against reliabilism because reliabilism

predicts that a subject in a Truetemp scenario has knowledge, while Lehrer and many others who consider the case, intuitively judge that there is something missing in the subjects epistemic situation that would qualify him as having knowledge. 220 subjects were given a series of cases and were asked to indicate whether they agreed or disagreed with the statement that the subject in each case had knowledge. These cases included a clear case of knowledge, a clear case of non-knowledge and a Truetemp case. If judgments about whether the subject ‘knows’ reliably track the subject’s underlying concept, then the order in which these cases are presented to the subject shouldn’t matter, but it turns out that the order of presentation does matter. In particular, the authors found that when subjects were presented with a clear case of knowledge, before considering a Truetemp case, they were less likely to attribute knowledge in a Truetemp case, and when subjects were first presented with a case in which a subject clearly lacks knowledge, they were more likely to attribute knowledge in a Truetemp case. Swain, Alexander and Weinberg take these results as threatening the thesis that intuitive judgments properly function as evidence in philosophical theorizing, noting that “evidence so unstable risks being discounted as not truly evidence at all” (2006, 141).

A more rigorous analysis of people’s intuitive judgments concerning the influential Trolley Problem in moral philosophy supports the thesis that intuitive judgments exhibit a high degree of instability. Schwitzgebel and Cushman (2012) show that not only are non-philosopher’s intuitive judgments subject to order effects, but that professional philosophers’ intuitive judgments are also subject to order effects and that this is so, even when the cases that are being considered are central cases in those philosophers area of expertise. Schwitzgebel and Cushman’s study includes responses from 2466 subjects, including 324

philosophers, 68% of whom specialize in ethics, 753 academics who were not philosophers and 1389 non-academics. Subjects were asked to respond to 17 cases as well as to 5 questions about which moral principles they endorsed. Schwitzgebel and Cushman are primarily interested in how these studies impact the thesis that philosopher's possess a certain kind of expertise that shields their own intuitive judgments from the biases which have been found to inflict non philosopher's intuitive judgments. This topic is not my main concern here. As such, I will focus only on a sample of question/responses from this study indicating the widespread influence of order effects on intuitive judgments. It is worth noting though, with Schwitzgebel and Cushman, that "both philosophers and non-philosophers showed significant order effects for all ... types of scenarios" that were tested. (2012,148)

Perhaps the most well known instance of the reliance on intuitive judgments in moral philosophy concerns the infamous Trolley Problem. There are many types of scenarios in this area and Schwitzgebel and Cushman's research focuses on two. In the Switch scenario, subjects are asked whether it is morally permissible to flip a switch knowing that doing so will send a Trolley hurdling down a track, inevitably ending the life of one person on the track, but simultaneously saving the lives of five people who will inevitably be killed if the switch is not flipped. People tend to intuitively judge that it is morally permissible to kill the one person to save the five in the Switch version of the Trolley problem. In the Push version of the Trolley Problem, subjects are asked whether it is morally permissible to push an innocent bystander into the way of an oncoming trolley knowing that it will kill that person while again saving the lives of five people who will inevitably die if the

bystander is not pushed onto the track. Interestingly, it is typical to intuitively judge that in the Push case it is not morally permissible to kill one person to save five.

6 of the 23 questions in Schwitzgebel and Cushman's study presented subjects with variations on these types of cases. There were 4 versions of the Push type case and 4 versions of the Switch type case. In questions 1 and 2, subjects were given one randomly drawn version of each type of case, counterbalanced for order. They were asked to read the case and to indicate on a seven-point scale the degree of moral goodness/badness of the actor in the case (1 being extremely morally good, 4 being neither good nor bad, and 7 being extremely morally bad). In questions 14 - 17, subjects were given one randomly drawn version of each type of case, in addition to two other cases that were not test cases and were expected to produce intermediate responses, again cases were counterbalanced for order. As before, subjects were asked to rate the degree of moral goodness/badness of the actor as above. The researchers were looking to see if Switch style cases and Push style cases would produce "equivalent judgments" (the same numerical response), or "inequivalent judgments" (differing numerical responses). Based on preexisting knowledge of typical reactions to these types of cases, one might expect that subjects would rate the Push style cases more harshly than the Switch style cases, thus producing a high percentage of inequivalent judgments. But this is not what happened. The results for question 1 and 2 were that 70% of respondents gave equivalent responses to these types of cases when Push was presented before Switch, and 54% gave equivalent responses when Switch was presented before Push ($Z=8.1$, $p<.001$). Results were similar for questions 14-17. 70% of respondents gave equivalent responses when Push was presented before Switch, and 60% of respondents gave equivalent

responses when Switch was presented before Push ($Z=6.27$, $p<.001$). These results support the claim that “respondents tended to assimilate their responses to the second scenario to their responses to the first scenario.” (2012, 142-3) Again, empirical studies find that our intuitive judgments are subject to order effects, seemingly undermining the claim that those judgments reliably indicate underlying concepts.

While these studies do not directly demonstrate intra-subjective variation in intuitive judgments, they are highly suggestive of it. The claim that intuitive judgments are characterized by intra-subjective variation is the claim that the same subject’s intuitive judgments about the same case will diverge on different occasions. If people’s intuitive judgments show a sensitivity to factors such as which other cases they have recently considered, than one can reasonably expect that if a single person’s intuitive judgments about the same case were sampled on various occasions, then those judgments would also exhibit variation. As far as I know, diachronic studies on single subject’s intuitive judgments about the same case have not been done. Still, existing studies point toward a high level of instability in intuitive judgments and this lends support to the claim that intuitive judgments are subject to intra-subjective variation. Intra-subjective variation problematizes the claim that intuitive judgments reliably indicate an underlying concept, personal or no. A plausible explanation of intra-subjective variation in intuitive judgments holds that these are easily manipulated and are not reliable guides to use when theorizing about concepts. Skeptics of the legitimacy of relying on intuitive judgments in inquiry can then argue that because intuitive judgments are so unstable, they cannot reasonably be held to be reliable indicators of the truth of their targets, and do not qualify as a proper source of evidence, even if the

target of inquiry is understood to be a personal concept in the way that Goldman (2007) and Goldman and Pust (1998) suggest.

In what follows I'll take it that intra-subjective variation, wherein the same subject has differing intuitive judgments about the same case on varying occasions, is strongly suggested by experimental studies.¹⁰ If there is such variation between a subject's judgments, then we are in a situation similar to that above. Instead of there being purported disagreement between different subject's intuitive judgments, intra-subjective variation provides a reason for thinking that there is disagreement between the same subject's intuitive judgments. Where there is disagreement, plausibly, one judgment must be right and the other wrong. Because the intuitive judgments of a single subject are characterized by a high degree of false judgments, these cannot, as a class, be thought of as generally reliable evidence.

Because of the similarity between both challenges to the reliability thesis, it would not be surprising, if an analogous solution to the problem posed by inter-subjective disagreement were also available here. Just as inter-personal variation in intuitive judgments does not pose a problem for the reliability thesis once we recognize that different people's intuitive judgments reflect different "personal" concepts, intra-personal variation will not pose a problem for the reliability thesis if it can be shown that the same person's judgments reflect distinct concepts. The idea that there is a single, unified concept serving as the source of all of one's judgments about a particular topic may be wrong. Instead, there may be various representations that can be called upon to make these judgments.

¹⁰ Weinberg also takes such studies as support for the claim that intra-subjective variation in intuitive judgments is a live possibility. See Weinberg, 2007, 338.

Somewhat surprisingly, though Machery is often associated with the negative program in experimental philosophy, whose self-avowed goal is to cast doubt on the reliance on intuitive judgments in inquiry, Machery's work on concepts provides the needed resources to argue against the claim that intra-subjective variation in intuitive judgments provides a strong reason for doubting the reliability thesis.

2.6 Concepts and Cognitive Science

Machery has argued that a survey of research on concepts in cognitive science shows that in order to fully explain categorization processes involving a particular ‘concept’, we need to posit not a single mental representation, as the personal psychological theory of concepts suggests, but rather three different types of ‘concepts’ about which no further unifying generalizations can be made. These are:¹¹

- 1) Prototypes - “A prototype of a given class is a body of statistical knowledge about the properties deemed to be possessed by members of this class” (2009, 83).¹²

¹¹ The following characterizations are very rough and do not take into account the subtle ways in which researchers differ in their uses of these terms. For a more detailed analysis see Machery, 2009, Ch. 4.

¹² It is worth noting that “a prototype” is “simply a convenient grammatical fiction”, that the term prototype refers to “the redundancy structure of the class as a whole” and that to think otherwise is to commit a “gross misunderstanding” or to assume a “covert theory of mental representation” (Rosch, 1999, 200).

- 2) Exemplars – A concept is a set of exemplars, where an exemplar is understood to be “a body of knowledge about the properties believed to be possessed by a particular member of a class” (2009, 94).
- 3) Theories - “a theoretical concept is supposed to store some nomological, causal, functional, and/or some generic knowledge about members of its extension”,¹³ or a theoretical concept “is a concept that gets its content by being embedded in theories” (2009, 101-2).¹⁴

While prototype theorists, exemplar theorists and theory theorists argue amongst themselves about which account of concepts can best explain higher order cognitive processes, Machery argues that a developed explanatory theory of higher order cognitive processes must assume all three. Machery’s arguments for this claim are numerous and detailed, covering a wide array of experimental and neuro-scientific research on cognitive processes including, categorization, induction, concept learning, and concept combination. I’ll focus on the last of these here to represent Machery’s broader argument for the necessity of recognizing that there are three distinct types of ‘concepts’ that are used by default in our higher cognitive processes, rather than just one.

At any given time we have a finite set of concepts, yet it seems we have an infinite ability to combine concepts from this set to produce new, complex concepts. Psychologists

¹³ This is similar to what Smith and Medin (1999, 220) call summary information, for example “All birds lay eggs”. Smith and Medin note that representation of summary information poses a problem for a “pure exemplar view”.

¹⁴ Closely related is what Laurence and Margolis refer to as the “Theory-Theory” of concepts, according to which “concepts are representations whose structure consists in their relations to other concept as specified by a mental theory” (1999, 47).

studying the process of concept combination have found that it is characterized by a phenomenon called property inheritance. Studies have shown that when we form a complex concept, this complex concept will likely inherit properties that are deemed to be typical of the simple concepts that compose it. Because “prototypes are supposed to encode the knowledge about typical properties ... it is plausible that when we produce on the fly a complex concept we retrieve the prototypes of the original categories and we transmit some of the typical properties they represent to the complex concept” (2005, 458) We need to posit mental representations resembling prototypes to explain the phenomena of property inheritance. However, we do not transmit all of the typical properties of the original concepts to the complex concept. We need theoretical concepts to decide which properties of the original prototype concepts are transmitted. For example, suppose we are asked to produce the complex concept GRANDMOTHER SPY. Although the concept SPY might have the prototypical feature of “being a male” associated with it, because we think that it is impossible for a GRANDMOTHER to be a male, this typical property of the concept SPY will not get transmitted to the complex concept GRANDMOTHER SPY. This shows that “property inheritance is influenced by our *theoretical knowledge* about the categories that are combined” (2005, 460). It seems we need to posit prototypes and theories to explain the phenomena of property inheritance. The process of concept combination has also been found to sometimes exhibit a phenomenon known as property emergence. Property emergence occurs when the complex concept represents properties that are not represented by the original concepts. Machery gives the example, PRESIDENT FROM TEXAS. Suppose someone possesses the concepts PRESIDENT and TEXAS without the complex concept PRESIDENT FROM TEXAS, and that for some reason they need to produce the

complex concept. It is possible that the resultant concept might include a property such as “being right wing” though this property is not part of either the concept PRESIDENT or TEXAS. It is plausible that what happens here is that when asked to produce the concept PRESIDENT FROM TEXAS one searches through memory for an individual who matches the description, in this case, G. W. Bush, an individual to whom they attribute the property of being right wing. The latter property then becomes associated with the complex concept PRESIDENT FROM TEXAS even though the property of being right wing was not part of either of these original concepts. “Hence concept combination accesses some singular representations. This suggests that these may be poised to be used in our higher cognitive processes, that is, that they are concepts” (2005, 462). Exemplars are also necessary to a full explanation of concept combination.

We started this chapter with a neutral view of concepts according to which they were whatever mental representations are used by default in our higher order mental processes, such as categorization judgments. But looking more closely at these higher order mental processes shows there is not one single kind of mental representation or ‘concept’ used by default, there are at least three that are fundamental, and about which no further inductive generalizations can be made. If these three different kinds of concepts can be expected to produce inconsistent categorization judgments relative to the same case, then this provides a means for claiming that instability is not a reason to doubt the reliability thesis. Instability, or variation of intuitive judgments within the same subject, will not be correctly interpreted as showing that there is a high level of disagreement among the subject’s judgments such that a high number must be wrong and therefore as a general class, unreliable. If we see the varying

judgments as each reflecting a different underlying concept then again, the disagreement disappears. There will be a direct causal link between a particular judgment and one of the three mental representations, or concepts, which is its source. Instability in the same subject's intuitive judgments will not provide a strong reason for doubting the reliabilist thesis and judgments about cases can still qualify as a proper source of evidence when theorizing about concepts.

Not only is the instability in intuitive judgments consistent with Machery's view, he has argued that his view predicts that intuitive judgments will display this type of instability. On Machery's view, distinct 'concepts' can be expected to give rise to irreducibly, apparently inconsistent, categorization judgments. Machery and Seppala explain that his view is consistent with the polysemous thesis: a given word has more than one distinct meaning, because it can express more than one distinct concept. Machery and Seppala consider what predictions this view makes:

"Suppose that "dog" expresses two concepts of dog, DOG1 and DOG2. Suppose further... that one of these two concepts (DOG1) is a prototype, while the other (DOG2) is a theory...There might be some objects that are sufficiently similar to the prototype of dogs to be judged to be dogs if we categorize by means of the prototype of dogs, but that are not judged to be dogs if we categorize by means of the theory of dogs. ...People should thus be disposed to hold conflicting judgments about whether these objects are dogs. Then if the polysemous hypothesis is true, they should be willing to assent to apparently contradictory sentences about dogs. That is, for some objects, they should be willing to assent to the sentences "This is a dog" and "This is not a dog" (2009, 109).

If intuitive judgments are expected to exhibit intra-subjective variation for the reason that is given here, then intra-subjective variation will not be a threat to the idea that intuitive judgments reliably indicate their targets, understood as in-the-head mental representations.

In order to defend the reliability thesis from the experimental studies showing ubiquity of variation in intuitive judgments, we have to further refine our account of what the target of that inquiry is. When we are targeting a particular concept, X, in inquiry, we are not only targeting something that is personal, we are also targeting three distinct bodies of knowledge that independently play an explanatory role in categorization judgments involving X. The fact that concepts are personal allows one to reject variation at the inter-personal level as providing a strong reason for doubting the reliability of intuitive judgments. Just so, the fact that the word ‘concept’ refers to three distinct mental structures, allows one to reject intra-personal variation as providing a strong reason to doubt the reliability of intuitive judgments as proper sources of evidence when theorizing about concepts.

In some ways, the recognition that intuitive analysis is targeting three distinct kinds of psychological structures might seem concessionary. Goldman, for example, does write as if intuitive analysis targets “the” concept associated with a single predicate. On his view, there is a single mental representation, or concept, that is explanatory of one’s categorization judgments on a particular topic. On the other hand, the rejection of this idea seems consistent with other work that relies on intuitive judgments to ground theories of mental representations. Positive experimental philosophers, for example, who also rely on intuitive judgments as evidence when theorizing, are not committed to the existence of a single mental representation or concept underlying intuitive judgments about a particular topic. Knobe notes that a survey of recent work in experimental philosophy shows that “no one ever proposes anything like an analysis of the concept as a whole” (forthcoming).

Let's recap. We started with the question of whether the method of intuitive analysis can usefully be pursued alongside other more commonly used methods in cognitive science to study mental representations, or psychological kinds, such as concepts. Goldman suggests that in order to show that the method of intuitive analysis is a legitimate method, we need to show that intuitive judgments are reliable, just like other basic sources of evidence. Recent studies of intuitive judgments provide two main reasons for doubting the reliability thesis: intra-subjective and intra-subjective variation. We have seen how it is possible to argue for the reliability of intuitive judgments in the face of both types of variation. While this goes some way toward diffusing the skeptical attitude toward the reliance on intuitive judgments in inquiries targeting in-the-head psychological entities, it does not respond to all of the arguments against the legitimacy of relying on intuitions as evidence. Below we will look at an argument against the trustworthiness of intuitive judgments that does not rely on the premise that these are unreliable.

2.7 Beyond Reliability

A skeptic might accept that intuitive judgments are generally reliable but that this fact alone is not enough to secure their status as a legitimate source of evidence. Weinberg, who questions whether reliability is the crucial factor that must be weighed in decisions over whether intuitive judgments should be trusted when theorizing, explores this position. Weinberg reminds us that when it comes to assessing the relative trustworthiness of a prospective source of evidence “epistemic factors other than reliability are in play” (2007, 325).

Whereas earlier versions of skepticism concerning the practice of relying on intuitive judgments in inquiry argue from the thesis that intuitive judgments are unreliable to the conclusion that they are untrustworthy and should be abandoned, the present challenge to philosopher's appeals to intuitions accepts, at least for the sake of argument, that intuitive judgments are generally reliable and attempts to find another route to the conclusion that intuitive judgment should not be trusted. To this end, Weinberg sets out to pinpoint the "epistemically deleterious characteristic" possessed by intuitive judgments which differentiate them from other generally accepted sources of evidence and makes them unworthy of our epistemic trust (2007, 323). He suggests that while generally accepted sources of evidence and intuitive judgments are both fallible, only intuitive judgments "are guilty of ... unmitigated fallibility". That is, even if both standard sources of evidence and intuitive judgments are generally reliable and produce an acceptably small number of errors, only in the former case, is this compensated for "by a decent capacity for detecting and correcting the errors." (2007, 323) Weinberg's view is that mitigated fallibility, rather than reliability, "can make the difference between a successful epistemic practice and an unsuccessful epistemic practice" (2007, 323).

Weinberg discusses four main sources of error detection and correction that are at work in successful epistemic practices (2007, 330). The first way in which we can check and correct for errors is through *external corroboration*: "each sense can at least somewhat be corroborated by the others, and all are corroborated by our more theory-mediated predictions about the world" (2007, 330). For example, suppose I judge that there is a cup

on the table in front of me on the basis of a visual experience wherein I seem to see a cup. I could be alerted to an error in my judgment by reaching for the cup and finding that it is not there: I can't touch it. Here, the visual seeming will not be corroborated and I have a reason to suspect that it may not be trustworthy. Though Weinberg recognizes a "handful" of areas in which we can appeal to external corroboration to check and correct for error in philosophical intuitions, he notes that "it is clear that philosophical intuitions has not had very much of this" (2007, 338-39).

The second way we can check for errors in more generally accepted evidential sources, such as sense perceptions is to assess whether there is "*internal coherence* in terms of agreement both within and across subjects" (2007, 330). We have already noted how disagreement with the majority provides a powerful reason for doubting one's sense experiences. However, internal coherence is also evident in larger scale practices such as laboratory testing. Here too "results [are checked] against other results in the expectation that the odd man out deliverances are mistaken" (2007, 330) Weinberg points out that the case is different when it comes to intuitive judgments. Here, Weinberg notes that recent empirical studies of intuitions about philosophically relevant cases "reveal that broad disagreements can arise in key discipline defining intuitions" and that philosophers tend to "overestimate the amount of intra-subjective agreement that one can expect" (2007, 337). It is not that philosophers couldn't use agreement as a test for error, but, in practice, they do not and therefore do not take advantage of a potential source of error detection and correction.

Third, when it comes to perception there is a reasonable “*detectability of margins*”. We often know for a given source of evidence, which conditions are likely to influence that source, causing it to produce misguided results. For example, “if you know you have been doused with a hallucinogen, you will know not to trust your senses” (2007, 330). When it comes to intuitive judgments, we simply do not know the conditions in which the cognitive processes that produce them are likely to malfunction. Furthermore, when using standard sources of evidence, we can typically use the outputs of the device itself to alert us to suboptimal performances. For example, scientists who work with electron microscopes can recognize that the device is not performing properly when the output of the device is degraded; “a poorly done scan will likely look like nothing at all”, the same is true, of perception, we know not to trust a perception whose richness is significantly degraded. Again, there is no comparable capacity to check for accurate intuitions. What would it mean for an intuitive judgment to be degraded? As Weinberg notes, the results of applying our judgment here results in a 1-bit signal. “Does the concept/predicate apply to a specific case?” - Yes or No?” One might try and assess the feeling of certainty that accompanies such a judgment but “this gradation is largely unexplored and unexploited in current practice” (2007, 335).¹⁵

Finally, when dealing with standard sources of evidence we often, at least to a degree, possess *theoretical illumination* about “how they work when they do – and more important why they fail to work when they don’t” (2007, 330). Such an understanding of the underlying processes by which standard sources of evidence are produced also functions as a mitigating

¹⁵ Though see Wright, 2010.

factor allowing for the possibility of error detection and correction. In general, there is little to no understanding of how intuitive judgments work, their origins are introspectively opaque. We just do not yet know much about the underlying psychological processes by which they are produced.

Following Weinberg we can use the term “hopeful” to refer to those practices wherein practitioners have the capacity to check and correct for errors and “hopeless” to refer to those practices wherein practitioners lack this ability to check and correct for errors (2007, 327). According to Weinberg, the standard practice of appealing to intuitive judgments in philosophical inquiry is unaccompanied by a decent capacity to check and correct for errors and is therefore hopeless. Weinberg is more agnostic about whether the standard practice of appealing to intuitive judgments in philosophical inquiry is “hopelessly hopeless”, but argues that at present, we lack the ability to detect and correct for errors in our intuitive judgments and that it is this feature of the practice of appealing to them that makes that practice untrustworthy and differentiates it from standard practices of appealing to generally accepted sources of evidence (2007, 341).

There are two questions/concerns that might be raised here: First, one might argue with the degree of hopelessness attributed to intuitive analysis. Second, one might question whether one who relies on intuitive judgments to outline underlying mental representations really needs more than reliability. I am going to leave these questions aside here. I take it that Weinberg’s general point about the need to be more careful when appealing to intuitive judgments in inquiry is a good one, and that we are currently at the beginning stages of

investigations that will eventually lead an increased understanding of intuitive judgments and the processes in which they are produced that will contribute to this ability to carefully rely on them.

Later in this dissertation, I'll return to the project of locating "intuitions dark margins" (2007,335). One way to increase the hopefulness of intuitive analysis that Weinberg draws attention to is by developing a constraint on the type of cases that are relied on in inquiry:

"One candidate way of trying to detect intuitions' margins would be to think about what sorts of cases we might be more likely or less likely to be able to reason well about. For example, we could, as a profession, decide to be particularly cautious about using intuitions under circumstances far removed from ordinary conditions—such as cases involving wildly unusual or even nomologically impossible situations, or that can be described only using fairly highfalutin lingo. Were we to do so, we might at least thereby restrict ourselves to intuitions that receive some of the same sources of hope as everyday intuitions" (2007, 336).

In Chapter 4, I'll propose and argue for one constraint on the type of cases that should be relied on in conceptual analysis.

This chapter began with a fairly neutral understanding of concepts as whatever psychological structures are used by default in our higher order cognitive processes. We saw that in order to safeguard the reliability constraint, according to which intuitive judgments provide a proper source of evidence for theorizing about concepts, we had to refine our account of what a concept is. We end the chapter with an account of concepts wherein concepts are personal, in the sense that they are fixed by what is in the concept bearer's head, and wherein our 'concept' of X, is actually a collection of distinct concepts, about which no further inductive generalizations can be made. For certain purposes this account of concepts is satisfactory. I, for one, think that there is nothing wrong with inquiries targeting

concepts, so understood. But one must be mindful of the fact that as Goldman puts it, “most philosophers prefer some target of inquiry that is not person-specific” (2010, 13). Furthermore, the account of concepts given above has moved very far from the classical theory of concepts. According to the classical theory, a concept has definitional structure; it encodes a set of necessary and sufficient conditions for application judgments involving that concept. In this way, concepts contribute to the truth-values of statements in which they appear. The statement “John is a bachelor” for example, will be true only if John satisfies the necessary and sufficient conditions that are encoded in the concept bachelor. Furthermore, on the classical view, concepts can also explain reference fixation. The set of necessary and sufficient conditions encoded by the concept bachelor jointly determine the class of individuals who fall within its extension. As Laurence and Margolis note, the classical account of concepts “has powerful explanatory resources, offering a unified account of concept acquisition, categorization, epistemic justification, analytic entailment, and reference determination” (1999, 10). It is far from clear that the account of concepts given here has resources to ground such explanatory power. In the next chapter, we’ll look at a program of conceptual analysis that is more closely aligned with these traditional goals of conceptual analysts.

Chapter 3

Two-Dimensionalism, Conceptual Analysis and Narrow Content

3.1 Good Old Armchair Conceptual Analysis

According to the classical view of concepts, each concept encodes a set of necessary and sufficient conditions that guide application judgments involving that concept. According to the classical view, concepts: 1) contribute to the truth-conditions of the statements that they occur in, and 2) play a role in reference fixation. Take a statement like “This is water”. On the classical view, the concept ‘water’ encodes conditions for its proper application that determine the truth-value of the judgment expressed in this statement, and these conditions jointly determine the extension of the concept “water”.

Together, these ideas support the view that one can have a priori knowledge of a concept’s extension. By relying on one’s intuitive judgments about how his/her concept applies across a range of real and hypothetical cases, one can work backward to knowledge of the necessary and sufficient conditions that form the basis for these judgments, at the same time, gaining knowledge of the essential features of that concept’s extension. Given that philosophers often partake in armchair inquiries into topics of interest, we can hardly disagree with Schroeter when she writes “It would be nice if good old armchair conceptual analysis were possible” (2004, 425)¹⁶.

¹⁶ It should be noted that Schroeter (2004) offers strong (though I don’t think fatal) criticisms of both good old armchair conceptual analysis and the modified version of conceptual analysis that I will be discussing in this chapter. I won’t be getting into these

Alas, it is generally recognized that good old armchair conceptual analysis is not possible. Most philosophers no longer equate concepts with sets of necessary and sufficient conditions. This leaves us with a question. What, if anything, are intuitive judgments good for in serious inquiry?

In the previous two chapters, we looked at two different answers to this question. The answer presented in the first chapter, associated with extra-mentalism, is, not much. Intuitive judgments are untrustworthy sources of evidence for uncovering knowledge of external facts. We might try and calibrate or mitigate to decrease the original untrustworthiness of intuitive judgments, but these still qualify as low-grade evidential sources. In the second chapter, we looked at a very different answer. Here, intuitive judgments are regarded as proper evidential sources, reliably indicating personal, psychological concepts, or mental structures that are used by default in higher order cognitive processes. This understanding of the target of inquiries relying on intuitive judgments is very different from that associated with good old armchair conceptual analysis. Finding out about personal, psychological concepts can help explain why we make the judgments that we do, and what our implicit or folk understanding of the referents of our concepts are, projects that are worth pursuing in their own right. But good old armchair conceptual analysis was targeting something different than concepts, as these were understood at the end of the last chapter. Good old armchair conceptual analysis took

debates here. My main interest in this dissertation is to see whether this, or a closely related account of conceptual analysis, can meet a specific kind of challenge posed by negative experimental philosophers to the reliance on intuitive judgments in inquiry.

concepts as functioning to actually determine the reference of our words, and as providing truth conditions for the use of our words. The idea that intuitive judgments reliably indicate their psychological sources, on the other hand, is consistent with an externalist view of what the truth conditions of these judgments are. It may be that the judgment “This is water”, reliably indicates the concept ‘water’, but that the truth value of that statement is still false, as truth conditions for such statements are settled not by the concepts themselves but rather externally. Furthermore, while we can understand what we take our concepts to be referring to by studying our personal psychological concepts, this is consistent with the idea that the actual referent of those concepts is determined externally, by contingent facts about the environment that we find ourselves in.

This chapter will discuss a promising recent version of conceptual analysis, associated most closely with Chalmers, which shares some explanatory goals with the old program, but is more modest at the same time. This version of conceptual analysis provides an alternative account of what the target of intuitive analysis is that is based in a two-dimensional account of conceptual content. Rather than a definitional view of concepts, where concepts are thought of as encoding necessary and sufficient conditions, this view adopts an intensional view of concepts. In particular, it will here be argued that intuitive analysis targets a concept’s epistemic intension. It is a familiar idea that each concept can be associated with an intension and an extension. It is a less familiar idea that each concept can be associated with two distinct types of intensions. According to the two-dimensional view of conceptual content developed by Chalmers, each concept has at least two dimensions of content corresponding to that concept’s “epistemic” and “subjunctive” intensions. While the

intuitive judgments that are relied on in philosophical inquiry are not well suited to guide theorizing about a concept's subjunctive content, they are well suited to guide theorizing about a concept's epistemic content. Furthermore, if we understand intuitive analysis as targeting a concept's epistemic content, then we are able to defend the thesis that the target of intuitive analysis is knowable *a priori*. Thus, conceptual analysis can be engaged in from the armchair as has been traditionally thought.

3.2 Wide and Narrow Content

We can begin with a rather well known distinction, that between wide and narrow content. Theories defending the thesis that conceptual content is wide hold that the contents of those concepts depend on facts about the subject's environment. Theories defending the thesis that conceptual content is narrow hold that the contents of those concepts depend on facts that are intrinsic to the thinker. On both views, concepts are held to be the constituents of thoughts. A thought, in virtue of the concepts that compose it, represents the world. The debate about whether conceptual content is wide or narrow can be thought of (for our purposes), as centering on the question of how concepts represent the world. The question is whether concepts are representational in virtue of reflecting aspects of the environment that one finds oneself in, or in virtue of factors that are internal to the subject and in some sense, independent of the actual environment that the subject finds herself in.

Both accounts of conceptual content have explanatory virtues. The thesis that concepts have wide content allows for an explanation of the widespread reluctance to assert

that persons in Twin Earth scenarios possess concepts with the same content. The thesis that concepts have narrow content allows for an explanation of the cognitive significance or inferential role of that concept for the person who possesses it. Though most accounts of conceptual content focus on one of these alternatives at the expense of the other, I believe a more satisfying view of conceptual content recognizes both. It is a virtue of two-dimensional accounts of content that they are able to accommodate both explanatory goals. The thesis here is that a concept has both a wide dimension of content and a narrow dimension of content; the former determined by facts about that subject's actual environment and the latter determined by facts that are, in some sense, independent of the subject's actual environment. The two-dimensional view of concepts developed by Chalmers utilizes standard possible world semantics to understand wide content and outlines a novel theory of narrow content. On the view developed here, it is this narrow dimension of conceptual content, (what Chalmers refers to as a concept's epistemic content) which provides the proper target of intuitive analysis.

3.3 Motivating Narrow Content

Most philosophers accept the thesis that conceptual content is wide, that is, that concepts represent the world by reflecting aspects of the environment that one finds oneself in. In this section I'll present key motivations for the contrary view; that conceptual content is narrow, or that content depends on facts that are in some sense independent of the environment that the subject finds herself in.

If you find the Cartesian view that it is conceivable that we would have the same beliefs and perceptions that we have now even if the external world were entirely different from how we now think it is at all tempting, then you should also find the thought that content is narrow tempting.¹⁷ The lasting impact which Descartes' work in the first meditation has had seems to suggest that the idea that content is narrow has a natural appeal. Even so, this is not persuasive, for despite being the natural view, the view could, of course, be wrong. Further motivating considerations are presented by Chalmers (2002, 137) who outlines three reasons typically given for adopting the view that content is narrow.

First, we want to be able to make sense of the idea that “empty concepts” such as ‘Santa Claus’ or ‘phlogiston’ which lack extension still have content. We can use these terms incorrectly, statements involving them can be true or false, so they should have truth conditions, and therefore content. It is tempting to say that the content of such concepts is determined internally.

Second, there are cases where the truth-value of a statement seems to be determined by factors that go beyond the extensions of the concepts that compose it. “For example, it is plausible that ‘John believes that Cary Grant is an actor’ could be true, while ‘John believes that Archie Leach is an actor’ is false even though Cary Grant and Archie Leach have the same extension” (2002, 137). If this is right and, as is traditionally thought, conceptual content determines the truth-value of statements, then again it seems there is more to conceptual content than extension.

¹⁷ For more on the relation between Descartes and narrow content see Lepore and Loewer, 2011, Ch. 9, “Solipsistic Semantics”.

Finally, there are cases where two concepts pick out the same extension but intuitively differ in content. Consider the complex concepts: ‘creature with a heart’ and “creature with a kidney”. Assuming that every creature with a heart is also a creature with a kidney, these complex concepts have the same extension. However, it seems obvious that these are different complex concepts. If so, there must be more to conceptual content than extension. Here one might take as a further example the concepts ‘Hesperus’ and ‘Phosphorus’ which both refer to the same planet (Venus) and have the same extension but nevertheless appear to be different concepts.

Cases such as these, point toward the recognition that conceptual content goes beyond a concept’s extension providing support for the thesis that content is narrow.

3.4 Motivating Wide Content

Despite the motivations for recognizing that concepts have narrow content outlined above, many philosophers find the view implausible. This is due, in no small part, to persuasive arguments against the view presented by Putnam. Following Putnam, we can outline two general commitments of the view that concepts have narrow content.

- 1) There is a narrow component of conceptual content that is determined by factors that are intrinsic to the thinker.

2) This component of conceptual content determines that concepts extension by providing a condition that any entity must meet in order to belong to that concepts extension.

Together, these two theses imply that the following statement is true:

3) For any thinker T, an intrinsic duplicate of T will have concepts with the same intensions and therefore extensions.

Putnam argues that 3 is false in his papers “Meaning and Reference” and “The Meaning of Meaning”. “We claim that it is possible for two speakers to be in exactly the same psychological state (in the narrow sense), even though the extension of the term A in the idiolect of the one is different from the meaning of the extension of the term A in the idiolect of the other” (1975, 222).

Considering the following influential thought experiment makes this point most forcefully. Putnam asks us to suppose that somewhere in the galaxy there is a planet called Twin Earth. Twin Earth is exactly like Earth except for a small difference in the chemical composition of the stuff we call water. Whereas on Earth the stuff that is called water is primarily composed of H₂O molecules, on Twin Earth, the stuff that is called water, which appears and behaves in exactly the same way that water appears and behaves on Earth, is actually composed of something different, call it XYZ. Twin Oscar lives on Twin Earth and is an exact Doppelganger, or intrinsic duplicate of Oscar who lives on Earth. When twin Oscar says ‘water’ on Twin Earth he is referring to XYZ, while when Oscar says ‘water’ on

Earth he is referring to H₂O. Putnam claims that because of this difference, and despite the fact that they are intrinsic duplicates, Oscar and twin Oscar have different concepts.

If we agree with Putnam (and many philosophers do), then it seems the statement expressed in 3 is false and we can no longer simultaneously maintain both commitments on which it is based. Either there is no narrow content that is determined by factors intrinsic to the thinker and content is wide in the sense that it is determined by factors in a thinker's environment *or* there is a component of narrow content which Oscar and Twin Oscar share but that does not determine the extension of our terms.

The intuitive judgments elicited by considering Twin Earth style thought experiments are sufficiently robust to present a challenge to the theory that thoughts have narrow content. As Jackson puts it:

“What is striking about the Twin Earth case is that it appears to be a case where clear intuitions...tell us that we must move to a position ...that sees content as a property that comes and goes ... with movements from one environment to another...in the absence of changes under the skin” (2003, 56-7).

Putnam's work shows that there is strong intuitive support for the thesis that content is wide, but, those intuitions are in tension with the cases discussed above which push us toward adopting the view that content is narrow. Chalmers argues that it is possible to accommodate the intuitive judgments elicited by Twin Earth style thought experiments which suggest that content is wide, and simultaneously hold that there is a separate dimension of content that is narrow in the sense that is shared by intrinsic duplicates. “The existence of wide content is compatible in principle with the hypothesis that every mental

state has narrow content: that is, that there is a (different) sort of content such that corresponding mental states of intrinsic duplicates have the same content of this sort" (2003, 46). According to Chalmers view, which is sometimes called "epistemic two-dimensionalism" we can recognize a component of conceptual content that is determined by facts about the thinkers actual environment and a distinct component of conceptual content that is determined by facts that are intrinsic to the thinker, wherein the contribution from the subject's actual environment falls away as inessential and which remains constant through environmental changes. This latter, narrow, dimension of conceptual content is here defined solely in terms of a concept's inferential role, or the role it plays in reasoning and thought.

In what follows, we will look at the foundations of the two dimensional view of conceptual content as presented by Chalmers. Two-dimensionalism, on this view, comes out of a confluence of ideas emanating from Frege, Carnap and Kant. We will see how strands from each writer's work can be woven together to create a picture wherein intuitive analysis targets a concept's intension that is knowable *a priori*. We will go on to see that Kripke challenges the foundational ideas that give rise to this position. To respond to Kripke's challenge and defend the idea that intuitive analysis is an *a priori* activity targeting a concept's intension, we will be pushed toward adopting a full-blown two-dimensionalism about conceptual content, wherein intuitive analysis targets a concept's epistemic intension.

3.5 The Foundations of Two-Dimensionalism

Chalmers traces the roots of two-dimensionalism back to three prominent figures in the history of philosophy: Frege, Carnap and Kant. To see how two-dimensionalism develops out of these positions it is not necessary to provide a detailed analysis of their thought, rather Chalmers suggests that if we “squint at the history from just the right angle, focusing on one strand of thought and setting aside others, we obtain a simplified rational reconstruction that brings out the key points” motivating a two dimensional account of conceptual content (2006a, 56).

It is useful to begin with Frege’s distinction between sense and reference. “Frege noted that the extension of an expression does not in general determine its cognitive significance: the role it plays in reasoning and in knowledge” (2006a, 56). This is evident when we consider the two expressions “Hesperus is Hesperus” and “Hesperus is Phosphorus”, where the term Hesperus refers to the evening star and Phosphorus refers to the morning star. In Frege’s view, the expression ‘Hesperus is Hesperus’ is cognitively trivial or insignificant while the expression ‘Hesperus is Phosphorus’ is cognitively non-trivial or significant. In the former case, the extension or truth value of the statement is known on the basis of reason alone, while in the latter case, the extension or truth value of the statement depends on the way the world is; it is an empirical discovery that “Hesperus is Phosphorous” is true. This difference in cognitive significance is not reflected in the referent of the terms that compose the statements; Hesperus and Phosphorus both have the same referent (Venus). Consequently, Frege thought it necessary to posit a second dimension of an

expression's meaning - an expression's sense, which reflects this difference in cognitive significance. Because an expression's sense is here held to be constitutive of meaning and because senses reflect cognitive (in)significance, which can be "known trivially by a rational being" it is possible to "see this characterization of sense as providing a first bridge between meaning and reason" (2006a, 56-7).¹⁸

Though Frege's general framework has seemed compelling to many philosophers, it is generally held that a better account of what a sense actually is, is required in order for the Fregean analysis of meaning to be acceptable. As Chalmers notes "Frege's own discussion leaves the matter somewhat unclear. He says that [senses] are not mental entities, such as the idea or image associated with an expression, and he holds that they are abstract objects of some sort (inhabiting the "third realm"). But this still leaves their nature open" (2002, 142). According to Chalmers we can get a clearer account of what a sense is by looking to Carnap.

Carnap's relevant "insight" was that "we can use the notions of possibility and necessity to help understand meaning and in particular to help understand sense" (2006a, 57). Carnap recognized that just as we have the ability to evaluate what a concept's extension is in the actual world, we likewise have the ability to use language to outline alternate possible worlds and evaluate a concept's extension in those worlds. This ability to evaluate the extensions of one's concepts across both the actual and possible worlds is explained by associating each concept with an intension, which is here understood as "a function from possibilities to extensions" (2006a, 57). For Carnap, judgments about how a

¹⁸ For present purposes, the terms "meaning" and "content" are synonymous.

concept applies across the set of actual and possible worlds provide evidence for a concept's intension. Here's Chalmers on Carnap:

"An expression can be applied to the actual state of the world, yielding an actual extension, or it can be applied to alternate possible states of the world, yielding alternate possible extensions. Take expressions such as 'renate' and 'cordate'. In the world as it actually is, all renates are cordates, so these terms have the same extension. But it is not *necessary* that all renates are cordates: if the world had been different, some renates might have failed to be cordates. Applied to such an alternate possibility, the two terms have a different extension. We can say: 'renate' and 'cordate' are co-extensive, but they are not necessarily co-extensive. Carnap suggested that we say two expressions have the same intension if and only if they are necessarily co-extensive. So 'renate' and 'cordate' have the same extension but different intensions" (2006a, 57).

Carnapian intensions behave like Fregean senses, that is, they are constitutively tied to meaning and are such that it is possible for two terms to have similar extensions (at least in the actual world) and different intensions, as shown in the quote above. Carnap's definition of intension "as a function from possibilities to extensions" is clearer than Frege's discussion of sense.

Carnapian intensions can also be contrasted with Fregean senses in so far as Frege held that senses are constitutively tied to reason while Carnap argues that intensions are constitutively tied to modal judgments about the extensions of our terms across a set of possible worlds. As Chalmers puts it "just as a sense can be seen as a kind of meaning constitutively tied to reason, intensions can be seen as a kind of meaning that is constitutively tied to modality" (2006a, 58).

At this point, it might be useful to borrow some imagery from Chalmers and to see the two positions as forming two sides of a triangle. At the top of the triangle is meaning, at the two bottom corners are reason and modality. Frege bridged the gap between meaning

and reason and Carnap bridged the gap between meaning and modality. Is there a way to bridge the gap between modality (judgments of possibility and necessity), and reason?

The easiest way to do this is by adopting the Kantian thesis that “A sentence S is necessary iff S is a priori” (Chalmers, 2006a, 58). It is easy to see how the notion of a priority (which says of a statement that it is knowable without any empirical evidence) links up with the Fregean notion of triviality, where a trivial statement (like “Hesperus is Hesperus”) is knowable on the basis of reason alone. One can rescue something like the Fregean idea that sense and therefore meaning is tied to reason if we tie the Carnapian insight that meaning is constitutively tied to judgments about possibility and necessity with the Kantian thesis that what is necessary is knowable a priori. If one accepts the Kantian thesis in combination with the Carnapian thesis then one has linked modality and reason, filling in the third side of the “golden triangle” pictured above. What we are left with, says Chalmers is a theory wherein “Modality serves as a bridge in explicating the tie between meaning and reason. One constructs a notion of meaning using modal notions, combines this with the claim that modality is constitutively tied to reason and ends with a link between all three. The central connection between meaning, reason and modality is captured within the Neo-Fregean thesis: intension is a notion of meaning, defined in terms of modality, that is constitutively tied to reason” (2006a, 58).

3.6 Intentions and the Targets of Intuitive Analysis

At this point it might be helpful to take a step back from the picture that is being developed here and ask how it relates to the present project as a whole. A broad question that this dissertation is concerned with answering is “Is the practice of relying on intuitive

judgments in philosophical inquiry a legitimate one?” To answer this question it is necessary to answer the more specific question “What is the target of intuitive analysis?”

In the last chapter we looked at two prominent answers to this question. Based on what has been said above we can begin to see a third; intuitive analysis targets a concept’s intension. Intensions are here distinguished on the basis of modal judgments: two concepts A and B will have the same intension iff A and B pick out the same extension in every possible world. In this case, we can say that the statement ‘A=B’ is necessary. On the other hand, two concepts A and B will have different intensions iff A and B pick out different extensions in some possible world. In this case, the statement ‘A =B’ will be contingent. In either case, by relying on modal judgments about what concept’s A and B pick out in possible worlds, we learn something about the intensions of those concepts. It is tempting to say that when philosophers make intuitive judgments about how concepts apply in hypothetical scenarios, they are making modal judgments about how a concept applies in various possible worlds and that these judgments can be utilized as evidence in a theory outlining a concept’s intension. Furthermore because we can link the modal judgments of possibility and necessity to reason by way of the Kantian thesis that what is necessarily true is knowable *a priori*, it seems possible to defend the thesis that conceptual content is knowable *a priori*. Intuitive analysis utilizes modal judgments to grasp an aspect of a concept’s content known as its intension. A concept’s intension is defined in terms of what is necessary across a set of possible worlds. What is necessary can be known *a priori*. Thus, intuitive analysis is an *a priori* activity that can be pursued from the armchair, in much the way that it is traditionally characterized. I think that something like this is correct. However, more work

needs to be done to show that the story told above, and in particular the Neo-Fregean thesis, “that intension is a notion of meaning, defined in terms of modality that is constitutively tied to reason”, is correct.

In the next section we will see how this way of thinking is challenged by Kripke. We will then move on to use a two dimensional conceptual framework to restore the thesis that intuitive analysis is an *a priori* activity that targets a certain kind of conceptual content identified with its (epistemic) intension.

3.7 Metaphysical Necessity and A priori Knowledge

The picture of intuitive analysis sketched above was challenged when Kripke “cut the connection between reason and modality” (2006a, 58). Kripke notes that while many philosophers treat the terms ‘necessary’ and ‘*a priori*’ as synonyms, “it’s not a matter of obvious definitional equivalence, either that everything *a priori* is necessary or that everything necessary is *a priori*” rather, the terms ‘*a priori*’ and ‘necessary’ belong in “two different domains, two different areas, the epistemic and the metaphysical” (1980, 36).

To see this, Kripke invites us to reflect on the “traditional characterization” of these terms. Something is considered *a priori* when it can be known independently of experience; to ask if something is *a priori* is to ask of a particular thinker whether that thinker can know something without helping herself to any empirical evidence. This is a different type of question than asking whether something is necessary. In the latter case

“We ask whether something might have been true or might have been false. Well, if something is false it’s obviously not necessarily true. If it is true might it have been otherwise? Is it possible that, in this respect, the world should have been different from the way it is? If the answer is yes, then this fact about the world is a contingent one. This in and of itself has nothing to do with anyone’s knowledge of anything” (1980, 36).

Because necessity is grounded in metaphysical facts about the world, while what is a priori is grounded in epistemological facts about what a subject can know, there is a gap between what is knowable a priori and what is necessary. As Chalmers puts it:

“Kripke argued that the Kantian thesis is false: there are many sentences that are necessarily true but whose truth is not knowable a priori. For example, Kripke argued that given that Hesperus is actually Phosphorus, it could not have been that Hesperus is not Phosphorus: Hesperus is necessarily the planet Venus, and so is Phosphorus. So although ‘Hesperus is Phosphorus’ is not knowable a priori, it is nevertheless necessary” (2006a, 58-9).

If Kripke is right and the Kantian thesis which links necessity with what is knowable a priori is false than the Neo-Fregean thesis which states that “intension is a notion of meaning, defined in terms of modality, that is constitutively tied to reason” is also false. And this makes problems for the general picture of intuitive analysis presented above.

To respond to Kripke’s challenge and to defend the Neo Fregean thesis and the picture of intuitive analysis that was sketched above, we will need to bridge the gap between necessity and what is knowable a priori. Chalmers proposes to do this by outlining a type of necessity that is cast purely in epistemic terms. If we can find a way of grounding necessity in epistemic facts rather than metaphysical facts, then Kripke’s challenge to the link between modality and reason will not hold. This is possible, according to Chalmers, if we adopt a two dimensional view of conceptual content. Below we will see that by tweaking some of the basic notions that we have already discussed, we can defend a version of the Neo-Fregean thesis and the picture of intuitive analysis that went with it.

3.8 A Closer Look at Intensions

We will return to the two-dimensional response to Kripke's challenge below, but it might be worthwhile to further clarify the nature of concepts as here understood, before moving forward. We are dealing with an intensional view of concepts, where concepts are constitutively tied to intensions, or functions that map possibilities to extensions. It is here taken to be a mundane truth that once we grasp a concept, and are given enough information about the world, we are able to identify the extension of our concept in that world. Chalmers and Jackson write that it is a "general feature of our concepts" that "if a subject possess a concept and has unimpaired rational processes, then sufficient empirical information about the world puts a subject in a position to identify a concept's extension" (2001, 323). For example, if I possess the concept 'water' and am given enough information about the distribution of chemical compounds in my actual environment, that for example, the clear, drinkable liquid that fills the oceans, rivers and lakes, and that falls from the sky, is mainly composed of H₂O molecules, then I am in a position to know that my concept "water" picks out substances primarily composed of H₂O.

To explain this general feature of concepts, we should jettison the association between concepts and necessary and sufficient conditions and rather think of concepts purely in terms of the ability that possessing them affords. Here, possessing a concept will amount to grasping that concept's intension, where

"the grasp corresponds to a *conditional ability* to identify [that concept's] extension, given sufficient information about how the world turns out and sufficient reasoning. That is, a sufficiently rational subject using [concepts] such as bachelor, knowledge, and water will have the ability to evaluate certain conditionals of the form "If E, then C", where E contains

relevant information about the world (typically not involving the [concept] in question) and where C is a statement using the [concept] and saying whether a given case fall into its extension (e.g. ‘John is a bachelor’, ‘Sue knows that p’, Water is H₂O”’ (2006a, 16).

Here, a subject can grasp a concept’s content (or intension) and possess the conditional ability to apply that concept, without grasping a corresponding definition. This is plausibly what happens in Gettier cases. The intension of the concept ‘knowledge’ can be approximated by the definition, “justified true belief” but this definition does not perfectly correspond to our ability to apply the concept. It is precisely because we can apply the concept in Gettier cases absent a perfect definition of the concept that we should think that the conditional ability to apply our concepts does not depend on possessing a definition.

If we take intensions, rather than sets of necessary and sufficient conditions as the target of analysis, the goal will be to arrive at definitions that approximate a concept’s intension. For example, we can say that the definition “justified true belief” approximates the intension of the concept ‘knowledge’. This definition does capture many of our intuitive judgments about how the concept applies in a wide array of scenarios. Chalmers suggests a better definition might be “justified true belief not essentially grounded in a falsehood” (2012, 17). Philosophical inquiry will progress by utilizing counterexamples to build better and better definitions of a concept’s intension, while at the same time recognizing that a true definition, which faithfully reflects every rational judgment involving that concept, may not be possible. If we take intensions as the targets of analysis then we do not need to worry about the problem of counterexample: “Arguments from counterexample can make a case against definitions but they can’t make a case against the claim that expressions have

intensions” (2012, 17). The picture of conceptual analysis we are left with is expressed in the following quote from Chalmers and Jackson:

“Once an essential role for explicit definitions is eschewed, the model of conceptual analysis that emerges is something like the following. When given sufficient information about a hypothetical scenario, subjects are frequently in a position to identify the extension of a given concept, on reflection, under the hypothesis that the scenario in question obtains. Analysis of a concept proceeds at least in part, through consideration of a concept’s extension within hypothetical scenarios, and noting regularities that emerge... What emerges as a result of this process may or may not be an explicit definition, but it will at least give useful information about the features in virtue of which a concept applies to the world” (2001, 322).

Furthermore, it is important to note that on this view, intensions are not necessarily operationally defined. It might be tempting to say that one can simply read off the intension of one’s concept based on observing how one uses that concept. However, this is not quite right. The thesis that judgments involving a concept will reflect that concept’s intension is here subject to various idealizations. For example, it is here held that in order for a categorization judgment to reflect a concept’s intension, that judgment should be based on rational reflection. When it comes to analyzing a concept’s intension, “it is not the speaker’s snap judgment that matters, nor any actual judgments of the speaker. Rather it is the speaker’s (potential) rational judgment that matters. Here we idealize away from poor reasoning, and consider judgments on ideal rational reflection” (2002, 147). It is an empirical question just how many of our judgments are so based and thus provides evidence for a concept’s intension. It may be that many times we do not make judgments on the basis of the idealization here described. But this fact does not speak against the possibility of targeting a concept’s intension. To reflect the fact that intensions are here constituted by rational inferences rather than snap judgments we can refine the definition of intensions given above as follows: “the intension of [a concept] E maps a possible case c to the

extension that the subject would identify for E, if they were to be presented with c and were to reason ideally” (2012, 209). It follows that we do not need to exclusively rely on first person reports to determine the extension of a subject’s concept. Instead of asking a subject to identify instances of a concept C, and then taking these categorization judgments as providing evidence in a theory of a concept’s intension, we may need to invoke a normative idealization and ask “Given all of a subjects background beliefs about the world, what do those beliefs imply about whether a given instance belongs to the extension of her concept, C?”

The view that we have been discussing is committed to a thesis that Chalmers refers to as the scrutability thesis. In its most general and unrestricted form the scrutability thesis says that once we know how the world has turned out, we are in a position to determine the extensions of our concepts in that world. This thesis can seem completely unremarkable and obvious or highly controversial. Thought of in relation to the example above, wherein a subject infers on the basis of a complete knowledge of the distribution of chemical compounds that her ‘water’ concept picks out those entities primarily composed of H₂O molecules the scrutability thesis seems easily supportable. However, it is also possible to consider cases wherein the scrutability thesis seems to be in trouble.

For example, suppose I want to evaluate the extension of the complex concept ‘people in Toronto’. The extension of this concept is in a constant state of flux so the evaluation will have to be relativized to a particular time T. The scrutability thesis suggests that if I am given enough information about the physical positions of all individuals at time T along with information concerning the municipal boundaries of the city, then I should be

able to evaluate the extension of this complex concept. However, it is far from clear that I could do so. It is more likely, that for a certain group of people, who are right at the borderline of the boundary between Toronto and not Toronto, say those people crossing Steeles Avenue, that I will not be able to decide whether they are part of the extension of the concept. It seems that “borderline cases”, like the one just described, pose a significant threat to the scrutability thesis.

On one popular theory, the question of whether or not a person (let's call him John) crossing Steeles Avenue falls within the extension of the concept “people in Toronto” will have no definite answer; there is no determinate fact of the matter when it comes to John's being in or out of Toronto. An alternate theory claims that there is a fact of the matter about whether John is in or out of Toronto at any given time T; we just can't know this fact due to our epistemic limitations.

Prima facie, on either view, the existence of borderline cases, presents a challenge to the scrutability thesis, which states that given enough information about the world, a subject S is in a position to determine for every individual whether or not that individual belongs within the extension of C. In the first case, it seems that there is no answer to the question of whether John belongs within the extension of the concept C and so it seems that the case presents a counterexample to the scrutability thesis. On the other hand, if there is a fact of the matter about whether John belongs within the extension of C, a fact of which we must remain ignorant because of our epistemic limitations, than the scrutability thesis also fails; we

cannot know the extensions of our concepts in borderline cases. On either view, borderline cases, will present a challenge to the scrutability thesis.

There are various options open to one who wants to defend the scrutability thesis in light of the existence of borderline cases. If one takes the first view according to which there is no fact of the matter about whether John falls under the concept “people in Toronto” then one could restrict the scrutability principle so that it applies only to cases where there is a determinate truth about whether a case belongs to the extension of a concept. The scrutability thesis will have to be redefined so as to reflect the fact that it holds only for determinate truths. Conversely, if one accepts that there is a fact of the matter about whether each individual falls within the extension of a given concept, and then explains our ignorance of this fact as emanating from certain epistemic limitations, then it may still be possible to defend the scrutability thesis by arguing that it applies to idealized agents who are not subject to these limitations. On this view we can say that most of the time, truths concerning whether or not a given individual falls within a concept’s extension are scrutible for epistemic agents such as ourselves, but there are cases wherein such truths are only scrutible for idealized versions of ourselves. Nevertheless, the scrutability thesis, stating that truths concerning whether or not a given individual is part of a concept’s extension are knowable given sufficient information about the world and sufficient reasoning, will hold.

On Chalmers view, the former move makes more sense. That is, one should reject the claim that there is a fact of the matter in borderline cases about whether the individual in question belongs to the extension of the concept and instead describe such cases as one’s in

which it is indeterminate whether the individual belongs within the concept's extension. In favor of this move, Chalmers argues that the contrary theory, that there is a fact of the matter in borderline cases, "is extremely counterintuitive" (2012, 289). We needn't settle the issue here. The point of this section is to illustrate how the scrutability thesis can handle "hard cases" such as the case of John described above.¹⁹

3.9 Epistemic Necessity and A priori Knowledge

Now that we have a better understanding of the intensional view of concepts we can return to the problem that we left above. We began with a purported connection among intensions, modal judgments of possibility and necessity, and reason giving rise to the "Neo-Fregean thesis: "intension is a notion of meaning, defined in terms of modality, that is constitutively connected to reason" (2006, 58). But Kripke argued that the connection between reason and modality does not hold. To restore this connection, Chalmers suggests that we have to cast necessity purely in epistemic terms, a project to which we will now turn.

Two-dimensionalism argues that it is possible to associate each concept with two distinct types of intensions. These

"correspond to two different ways of thinking about possibilities. In the first case, one thinks of a possibility as representing a way the actual world might turn out to be: or as it is sometimes put, one *considers a possibility as actual*. In the second case, one acknowledges that the actual world is fixed, and thinks of a possibility as a way the world might have been but is not: or as it is sometimes put, one *considers a possibility as counterfactual*" (Chalmers 2006a, 59)

¹⁹ There are other "hard cases". For a discussion of these see Chalmers, 2012, 259-312.

Typically, the possibilities in the domain of a concept's intension are understood counterfactually, as possible ways the actual world might have turned out but did not. When we consider a possibility counterfactually, judgments about a concept's extension relative to that possibility are determined by facts about what the concept picks out in the actual world.

For example, suppose I consider the possibility in which the stuff in the oceans, rivers and lakes is primarily composed of XYZ molecules rather than H₂O molecules counterfactually, and ask what the concept 'water' picks out relative to that possibility. If Putnam is right, then the concept 'water' will pick out H₂O in that possibility, not the "watery stuff" that is composed of XYZ. Furthermore, if Kripke is right, that names and natural kind terms function as rigid designators, then the concept 'water' will pick out H₂O in every counterfactual world. Thus, the statement "Water is H₂O" is metaphysically necessary, or true in the actual and all counterfactual worlds.

We can call the intensions that map concepts to their extensions in counterfactual possibilities "second dimensional intensions". More specifically, Chalmers refers to these intensions as "subjunctive intensions".²⁰ Subjunctive intensions capture our judgments about what our terms refer to in counterfactual worlds. The type of conceptual content represented

²⁰ Chalmers argues for a two dimensional view of conceptual content wherein concepts have both a "subjunctive intension" and an "epistemic intension". The more generic terms "second dimension intension" and "first dimension intension" are used to refer to two-dimensional views more broadly. In particular, these terms refer to two-dimensional views that define the relevant intensions in ways that differ from Chalmers. For example, instead of understanding first dimensional intensions as "epistemic intensions", that are fully grounded in the epistemic domain one may understand first dimensional intensions as "contextual intensions" or "linguistic intensions". These latter approaches to intensions, according to Chalmers, do not have the resources to restore the connection between modality and reason. For more on why this is so, see Chalmers 2006(a), 62-75.

by a concept's subjunctive intension is naturally thought of as a variety of wide content that depends on the subject's actual environment. Furthermore, when a concept picks out its extension rigidly, (when it picks out the same extension across all counterfactual worlds) then that concept will have a necessary subjunctive intension. It is this kind of necessity that Kripke rightly claims cannot be known a priori. We can't know a priori that water refers to that substance primarily composed of H₂O, even though, the statement "Water is H₂O" is metaphysically necessary.

This way of thinking about conceptual content has certain familiar consequences. Notably, it allows for the possibility that I could be massively mistaken about the descriptions I associate with my concepts. For example, in the actual world, the concept Hesperus picks out the planet Venus and is roughly associated with the description "the brightest star in the night sky". Suppose I imagine a counterfactual possibility wherein the brightest light in the night sky is not a star at all but rather a satellite orbiting the earth and wherein Venus is visible only in the morning. If we consider this possibility counterfactually, recognizing that the concept "Hesperus" picks out the planet Venus in the actual world and that names refer rigidly, then Hesperus will pick out Venus in this possibility. Because Venus is visible only in the morning, the description "the brightest star in the night sky" will no longer be correctly associated with the concept Hesperus. We will not think that the concept Hesperus now picks out Mars, which is, in our scenario the brightest star in the night sky.

Most philosophers accept that this view of things is correct and a two-dimensionalist about conceptual content need not deny that there is a wide or subjunctive dimension of

conceptual content that is reflected in our judgments of how concepts behave in counterfactual possibilities. However, as illustrated in the quote above, two-dimensionalism takes a broader stance to the question of how possibilities are to be understood. Rather than thinking that the set of all possibilities is exhausted by counterfactual worlds, the two-dimensionalist recognizes a broader set of possibilities. Rather than considering a given possibility counterfactually, as a possible way the world could have been but is not, I can also consider a possibility as actual. To consider a possibility as actual rather than counterfactual is to hypothetically accept that the possibility accurately describes one's own world. A single concept may have a different extension depending on what type of possibility it is being evaluated in.

For example, instead of considering the XYZ world counterfactually, as a possible way that the actual world could have been but is not, I can also consider the XYZ world as actual. If I hypothetically accept that my actual world is as described in the XYZ scenario then my ‘water’ concept will have a different extension than it would have were I to consider the XYZ world counterfactually. If I consider the XYZ world counterfactually, then as we have seen, my ‘water’ concept will necessarily pick out H₂O. On the other hand, if I consider the XYZ world as actual and ask “If the world really were that way what would my ‘water’ concept pick out?” then I will be lead to conclude that ‘water’ picks out those entities primarily composed of XYZ rather than H₂O relative to that possibility.

To take a further example we can return to the case discussed above wherein the brightest object visible in the night sky is actually a satellite orbiting the earth and where

Venus is visible only in the morning. If we take this possibility as representing a counterfactual world than we will say that Hesperus (Venus) is not visible in the night sky. We will not say that Hesperus refers to Mars, which is now the brightest star in the night sky, rather we will respect the fact that Hesperus is a rigid designator that refers to Venus necessarily in each counterfactual world. On the other hand, if take the scenario described above as a world considered as actual, then judgments about the concepts extension in that scenario will be different. Instead of taking the scenario as a counterfactual world, we can take the possibility as representing a way the actual world might hypothetically be. In this case, if the world were the way described in this scenario, then Hesperus would pick out Mars, which is the brightest star in the night sky.²¹ Again, the extension of the concept Hesperus will change depending on the type of possibility it is evaluated in.

The two-dimensional view of conceptual content argues that this difference in extension reflects a difference in intension; to capture the difference in how concepts behave relative to the two kinds of possibilities outlined above, we have to associate each concept with two distinct types of intensions. A concept's subjunctive intension is a function that takes possibilities considered counterfactually as input and outputs extensions relative to those possibilities, while a concept's first dimensional or epistemic intension is a function that takes possibilities considered as actual as input and outputs extensions relative to those possibilities. Just as we can take modal judgments about what a concept picks out in counterfactual worlds as evidence for that concept's subjunctive intension, we can take

²¹ This example is discussed in Chalmers 2002, 60.

modal judgments about what a concept picks out in worlds considered as actual as evidence for that concept's epistemic intension.

While the subjunctive content of a concept is grounded in that concept's extension in a subject's actual environment, the epistemic content of that concept is "usually quite independent of the subject's actual environment, and of the way the actual world turns out more generally" (Chalmers, 2002b, 616). To evaluate a concept's extension at worlds considered as actual we rely solely on that concept's role in indicative conditionals.

"Let us take a singular concept C expressible by a term B. To evaluate C's epistemic intension in a [world considered as actual] W, one considers the hypothesis that W is actual, and asks, "What is B?" To answer this question, "one can appeal to the indicative conditional 'if W is actual, what is B?'" (2002, 612)

As long as the scrutability thesis holds then these kinds of inferences are unproblematic. Chalmers notes that it is "a *prima facie* datum that there is an epistemic dependence between ...possibilities and sentence tokens" of the sort that we are considering here (Chalmers 2006, 81).²² For any given hypothetical description of a world, we will be able to evaluate the extension of our concepts relative to that description.

We are now in a position to outline a kind of epistemic necessity; one thought epistemically necessitates another when the first is epistemically incompatible with the negation of the second. The thought that the actual world is as described in the XYZ scenario epistemically necessitates the thought that "water is XYZ", likewise, the thought

²² Of course, one could disagree. To show that Chalmers is wrong, one would need to provide an argument against the scrutability thesis that was discussed above. Chapters three and four in Chalmers' book "Constructing the World" (2012) "mount the core arguments for the scrutability thesis"(Chalmers 2012, xix). Indeed, the whole book can be thought of as a long defense and examination of the scrutability thesis, in its varied forms.

that the H₂O world is the actual world epistemically necessitates the thought that “water is H₂O”. The thought that “water is XYZ” is incompatible with the thought that the actual world is the H₂O world, while the thought that “water is H₂O” is incompatible with the thought that the actual world is the XYZ world. Another way to put the same point is to say that the XYZ world verifies the thought that water is XYZ, while the H₂O world verifies the thought that water is H₂O. Following Chalmers, we can say that a thought is “deeply epistemically necessary” when its negation is epistemically incompatible with any epistemic possibility (2006a, 79). When a thought involving a concept is deeply epistemically necessary, or compatible with any hypothetical description of what the actual world is like, then it is plausible that that thought captures the concept’s epistemic content.

For example, it is plausible that the epistemic intension of the concept water is roughly captured in the following description “the dominant, clear drinkable liquid that falls from the sky and fills the oceans, rivers and lakes.”²³ This thought is compatible with the possibility that the stuff that fills the oceans, rivers and lakes etc. is XYZ or H₂O; furthermore, it is plausible that it is compatible with any hypothetical description of the underlying character of the actual world. If so, the description “the clear drinkable liquid that falls from the sky and fills the oceans, rivers and lakes etc.” roughly captures the epistemic content of the concept water.

Epistemic necessity is very different from metaphysical necessity. This notion of epistemic necessity allows for a response to Kripke’s challenge to the link between necessity

²³ The word roughly, is important here. Based on the foregoing discussion, we should expect that any such description approximates the concept’s true intension.

and reason. To address Kripke's challenge “we simply have to strongly distinguish ...epistemic evaluation ... (which turns on epistemic necessitation) from the usual sort of counterfactual evaluation (which turns on metaphysical necessitation)” (2006b, 589). Epistemic necessitation is reflected in rational inferences concerning a concept's extension in worlds considered as actual. These rational inferences are justified without any further empirical evidence and plausibly have an a priori justification. Metaphysical necessitation is reflected in judgments about what concepts pick out in counterfactual worlds. These judgments are justified in relation to what that concepts picks out in the actual world. Kripke's challenge to the purported link between necessity and what is knowable a priori holds if we understand the relevant type of necessity metaphysically. But we don't need to do this. If we understand necessity epistemically, as grounded in our rational judgments about what concept's pick out in worlds considered as actual, then the challenge does not hold. This leaves the “golden triangle” of constitutive connections between intensions, modality, and reason intact, and allows for a defense of a variation on the Neo-Fregean thesis that “intension is a notion of meaning, defined in terms of modality, that is constitutively connected to reason”. In particular, we can defend the thesis that a concept's epistemic intension is a kind of conceptual content, defined in terms of modal judgments, that is constitutively tied to reason. Conceptual analysis will then be understood as relying on modal judgments about cases, which have an a priori justification, to learn about a concept's epistemic content. This reliance on intuitions has no quarrel with Putnam's claim that “human intuition has no privileged access to metaphysical necessity” (1973, 709). Intuitive judgments, given the normative idealizations discussed above, are here taken to reveal that

part of conceptual content that is epistemically necessary or “true under any coherent hypothesis about what the actual world is like” (2006, 100).

3.10 Epistemic Intentions and Narrow Content

We have been discussing the view that conceptual analysis can proceed by relying on modal judgments about hypothetical cases to produce theories about a particular type of conceptual content: epistemic content. On the two-dimensional view of conceptual content presented above, conceptual analysis should be understood as an activity that targets a concept’s epistemic intension. This is a kind of truth-conditional content that contributes to the truth conditions of statements in which it is a part. When a concept is part of an utterance the epistemic content of that concept provides an “epistemic truth-condition for the utterance, specifying how the truth of the utterance depends (epistemically) on which epistemic possibility turns out to be actual” (2006b, 594). The variety of conceptual analysis defended here shares some common explanatory goals with good old armchair conceptual analysis. Conceptual analysis, as it is here understood, can give us *a priori* knowledge of truth-conditional content. Of course, the epistemic content of a concept does not exhaust that concept’s contribution to the truth conditions of utterances of which it is a part. The subjunctive content of the concept also contributes a separate kind of truth condition to the utterance, specifying how the truth of the utterance depends (metaphysically) on how the actual world has turned out. As Chalmers notes, for a two-dimensionalist about content “there should be no question about whether the primary [epistemic] intension or the

secondary [subjunctive] intension is *the* content... both can play some of the roles that we want contents to play" (2006b, 593-4).

One of the explanatory roles that epistemic content can play is making sense of the view that content is narrow. We saw motivation for this view above. Even the Twin Earth thought experiment, that it is meant to convince us that content is wide, raises the question of narrow content. As Chalmers says, "the cognitive similarities between Oscar and his twin are striking" (2002b, 608). We can explain these similarities here by noting that Oscar and his twin share a concept with similar epistemic content, while the subjunctive content of their concepts are different. By being grounded in the role that a concept plays in reasoning and thought, narrow content has explanatory resources for handling puzzles about meaning that the wide theory of content does not.

Indeed, Chalmers goes further in arguing that epistemic content really is a kind of narrow content. To show that epistemic content really is narrow, Chalmers argues that all of the descriptive elements in the description of an epistemically possible world are themselves narrow. If the descriptive elements of the possibilities that are in the domain of a concept's epistemic intension include wide elements, then the type of content discussed above will not really be a variety of narrow content. Though I won't be able to get into this argument in the depth that it deserves, the following discussion should give the reader some sense of Chalmers' argument for the narrowness of epistemic content.

Suppose we ask what kind of descriptive elements are contained in the description of an epistemically possible world. Following Chalmers and Jackson, one might think that the following kind of truths are essential: all truths of physics, phenomenal truths, indexical truths and a “that’s all” truth (2012, 110-112). It seems that we must include both physical and phenomenal truths in the base because it has been argued that neither can be inferred from the other. Goodman and Quine argued that physical truths are not defintionally entailed by phenomenal truths, while Jackson’s “Mary” thought experiment shows that phenomenal truths are not entailed by physical truths (2012, 108-109). We cannot restrict the base to physical and phenomenal truths, because indexical statements like “I am Jill Cumby” will not be scrutable form physical and phenomenal truths alone. In order to account for these truths we have to build indexical truths into the base. Finally, we will have to build in a “that’s all” truth, showing that the world in question is completely captured in the set of truths considered.

To see how this is a narrow dimension of conceptual content, we’re going to have to shrink the class of truths that belong in this set to those that are narrow, or those that depend only on the internal state of the thinker. Doing so will allow for a defense of the “narrow scrutability these: all truths are scrutable from truths whose content is determined by the internal state of the thinker” (2012, 23).

We can begin to minimize the base by getting rid of macro-physical truths, instead macro-physical truths can themselves be thought of as scrutable from microphysical truths. For example, truths about the location of a macro-physical object can be reconceived as

truths about the spatio-temporal region consisting of the location of its fundamental parts.

For example, the mass of a macroscopic object is just the total mass of its fundamental parts, “to determine the location and mass of a macroscopic object, it suffices to know the location and mass of its fundamental parts” (2012 291). Furthermore, “something similar plausibly applies to velocity, force, momentum, energy and the like” such that all macrophysical truths about the spatio-temporal position, mass, velocity, force, momentum, energy etc. of regular macro-level objects are scrutable from (classical) microphysical truths about the position, mass, velocity force momentum of its parts (2012, 291).

We can continue to minimize by noting that microphysical truths can be thought of as scrutable from still other truths. Microphysical truths utilize terms such as mass and spin etc., which are theoretical terms. As such, they can be gotten out of the base by “regimenting and Ramsifying” (2012, 319). First, one lists the core principles in one’s theory of “electron” and then replaces the term “electron” with a description that corresponds to these core principles; “perhaps: ‘that kind E such that objects with E have negative charge, orbit around the nucleus, play such and such a role in bonding, and so on’” (2012, 319). The term ‘electron’ can be replaced with a description that does not contain electron, but will just contain a number of other terms. Following Lewis, we can call these terms O terms for observational terms or old terms (2012, 319).

Now, in our set of truths we can take out the microphysical truths and instead put in O terms, where all microphysical truths are themselves scrutable from O truths. We can keep going and minimize even further. What O terms are needed? It is plausible that the set

of O terms include primary qualities (space, time, mass), secondary qualities (color sound etc.) and causal notions. Chalmers argues that primary and secondary qualities are themselves scrutable from phenomenal truths. “To a first approximation, ‘red’ might be analyzed as ‘the property that normally causes phenomenally red experiences in me. A closer approximation might be (for example) holistically define all the colors as that manifold of properties that serve as a causal basis for a corresponding manifold of color experiences, while saying more about what counts as normal” (2012, 322). It is worth noting that this move depends of adopting a functionalism about color terms and will not work if one adopts a primitive view on color terms. One can also adopt a functional view of spatio-temporal terms “On such a view, spatial concepts pick out that manifold of properties that serves as the normal causal basis of a corresponding manifold of properties of our spatial experience. Temporal concepts pick out a manifold of properties that serve as the normal causal basis of the corresponding manifold of properties in our temporal experience”. On a functionalist view of the sort Chalmers defends “spatial and temporal truths are a priori scrutable from phenomenal and causal truths along with uneliminated background truths” (2012, 335)

What about causal truths? Causal truths fall into the category of nomic truths, or truths that express empirical regularities or natural necessities, other concepts in this class include “law” and “chance”. On some views (Humean) these nomic truths are grounded in non-nomic truths, perhaps our own expectations of how things will turn out. However, on the view adopted here, nomic truths are fundamental conceptual primitives and cannot be analyzed further (2012, 340). As such they belong in the scrutability base. Now our scrutability base includes phenomenal expressions, nomic expressions, indexical expressions,

logical and mathematical expressions and whatever expressions are involved in a that's all clause" (2012, 340). We can call this scrutability base the minimal scrutability base.

Chalmers argues that it is at least plausible that all of the truths in the minimal scrutability base are narrow (2012, 393-395). If so, it is possible to defend the Narrow Scrutability thesis: "all truths are scrutable from truths involving only narrow expressions" (2012, 394). Furthermore, the narrow scrutability thesis underwrites the claim that the kind of conceptual content represented by a concepts epistemic intension is itself narrow.

"The Narrow scrutability thesis has important consequences for the theory of content. In particular, it can be used to support the claim that ...thoughts have narrow content, a sort of content that is independent of the environment... To make the case for narrow content, we use the scrutability framework to associate...thoughts with [epistemic] intensions (intensions defined over epistemically possible scenarios...) and use Narrow Scrutability to argue that these intensions are a sort of narrow content" (2012, 396).

3.11 Remarks

This chapter presented an alternative to the two main theories of what the target of intuitive analysis is. Instead of seeing intuitive analysis as targeting natural kinds or properties on the one hand, or personal psychological concepts, on the other, we can think of intuitive analysis as targeting a concept's epistemic intension. A concept's epistemic intension is a narrow component of conceptual content, though the thesis that concepts have this dimension of narrow content is not in tension with the view that concepts also have a wide dimension of content; represented here, by that concept's subjunctive content. We have seen that adopting the view that intuitive analysis targets a concept's epistemic content, provides a response to Kripke's challenge against conceptual analysis and allows for a defense of the

idea that the target of conceptual analysis is knowable *a priori*. Now that we've seen that we can target an aspect of conceptual content by relying on rational judgments of how that concept applies in hypothetical cases from the armchair, in the next chapter, I want to support this account of conceptual analysis by showing that it has resources for replying to one of the criticisms concerning intuitive analysis that came up in previous chapters. This criticism asks those who rely on intuitive judgments in theory construction to provide some constraints on that practice. We know that the reliance on intuitive judgments to tell us about conceptual content is not infallible. The challenge is to supply a means for deciding when intuitive judgments are likely to lead us astray. According to the view presented in this chapter, intuitive judgments about cases reflect the epistemic content of a concept when those judgments are based on idealized rational reflection. When do we meet this requirement and when do we not? Is our ability to base our judgments on rational reflection impacted by kind of case that we are considering? Are some far out cases more likely to impede the reliance on rational reflection? I think so. I'll discuss this further in the next chapter.

Chapter 4

Towards a General Criterion of Compellingness

4.1 Successful and Unsuccessful Appeals to Hypothetical Cases

I begin with two episodes from the history of philosophy. The first occurred in the early 1960's with the publication of Gettier's "Is Justified True Belief Knowledge?" The paper presents two hypothetical scenarios wherein the agent at the center of those scenarios satisfies what were generally held to be the generally accepted necessary and sufficient conditions for knowing a proposition. Gettier holds that it is clear that the agents at the center of these scenarios do not know the proposition in question, despite meeting the necessary and sufficient conditions for knowing.

The second episode occurred in the late 1980's with the publication of Davidson's paper "Knowing One's Own Mind". Davidson presents a hypothetical scenario meant to support the thesis that ordinary mental states are both "inner" in the sense that they are identical with states of the body and yet also "non-individualistic" in that they are at least partially constituted by relations to events and objects outside the subject who's states they are. The hypothetical scenario that Davidson presents is perhaps less well known than the Gettier scenarios so it might be useful to present the scenario here.

"Suppose lightning strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, Swampman, moves exactly as I did; according to its nature it departs the swamp, encounters and seems to recognize my friends, and appears

to return their greetings in English. It moves into my house and seems to write articles on radical interpretation. No one can tell the difference" (1987, 433)

Davidson finds that there is a difference between his own past and Swampman's current mental states. Though Swampman is an exact physical replica of Davidson, Swampman, according to Davidson, cannot properly be said to have thoughts. This is because Swampman's mental states lack the requisite meaning generating causal historical ties to the external environment that Davidson's own mental states had prior to his unfortunate demise.

I draw attention to these two episodes not to argue for or against either of the views that Gettier or Davidson are promoting. The present concern is methodological. These hypothetical scenarios are of interest because they represent a widespread practice characteristic of philosophical theorizing. Philosophers often appeal to hypothetical cases to support their positions. Furthermore, the foregoing examples illustrate how philosophers can either succeed or fail in this practice. Gettier cases are widely held to be successes, having changed the development of epistemological theorizing. As Jackson puts it "support for the justified true belief account of knowledge dropped like a stone consequent on the publication of Gettier, 1963" (2011, 468). Reaction to the Swampman case has been very different. The case is largely seen as a failure, having achieved what Malpas, in the Stanford Encyclopedia entry titled 'Donald Davidson', calls "a very limited usefulness" (2015).

Reflection on such cases raises a question about the role of hypothetical scenarios in philosophical theorizing. Why are some hypothetical cases successful while others are not? It is typically said that a hypothetical scenario is successful just in case that scenario presents a compelling counterexample to a given thesis. In order to be compelling a scenario must

invoke a certain kind of reaction in its audience. Upon considering the scenario one must be motivated to reevaluate the thesis at issue, due to judgments made about the scenario that are at odds with those predicted by that thesis. This does not mean that a compelling case must generate consensus on whether the theory in question is a good one. A compelling counterexample will provoke a widespread reaction but it needn't happen that everyone considering the case will go on to reject the theory in question. It is sufficient that the reaction to such cases provokes further refinements or clarifications from those who hold the theory in question.

It is sometimes suggested that whether a scenario is compelling depends on one's being already inclined, on the basis of background theoretical commitments, to reject the theory to which the scenario is supposed to provide a counterexample. But this suggestion does not withstand scrutiny. Many epistemologists who held the JTB account of knowledge found Gettier's cases compelling, as evidenced by their proceeding to either reject or make substantial revisions to their views based on consideration of Gettier's cases. If the suggestion that hypothetical cases are compelling only to those who hold a background theory consistent with rejection of the theory in question were right, then one would not expect this to happen.

Given the widespread use of hypothetical scenarios in philosophical theorizing, a deeper understanding of what makes a scenario compelling is not only of theoretical interest, but also instrumentally valuable. In what follows, I'll attempt to contribute to this understanding of what makes a given hypothetical scenario compelling by providing a

general criterion of compellingness. Namely, a hypothetical case will function as a compelling counterexample to the extent that the audience considering the case shares a perspective with the agent at the center of the case being considered. I'll show how this follows from considerations concerning the nature of the hypothetical cases that are appealed to in conceptual analysis. I'll then provide an account of what constitutes perspective, such that it can be shared between the audience considering a case and the agent at the center of that case. I'll then apply this criterion to various instances of hypothetical cases, including those discussed above, and show how it explains reactions to these cases.

4.2 Hypothetical Scenarios, Counterfactual Worlds and Thought Experiments

What is the character of the hypothetical cases that are appealed to in philosophical theorizing? Philosophers who rely on hypothetical scenarios in theorizing have often not given a lot of thought to this question. Yet, this has proven costly. The lack of a proper understanding of the nature of these cases encourages skepticism about the practice of relying on them. For example, here's Weinberg outlining how skeptical worries about philosophers practice of appealing to judgments made in hypothetical scenarios to support their theories are based, in part, on this gulf between the heavy reliance on hypothetical cases on the one hand and an incomplete understanding of these cases, on the other:

"the practice [of appealing to judgments about hypothetical cases] appears to set no constraints to how esoteric, unusual, far-fetched or generally outlandish any given case can be. Everyone is familiar with Davidson's Swampman, or Searle's Chinese Room, but one can look at the recent literature and find the likes of double-leashed testifiers, new evil demons, and fissioning/fusioning/teleporting pairs (or are they?) of persons. So this anything goes

aspect of the practice is what makes it particularly ripe for the opponent's challenge" (2007, 321).

I agree with Weinberg that the time has come for further reflection on the method of cases. But I'm more hopeful about our ability to come up with constraints that can both increase our understanding and improve the effectiveness of this method. I believe the criterion of compellingness outlined above is one such constraint. When developing effective scenarios, we are constrained by the requirement that we design scenarios so that the audience considering the case and the agent at the center of the case share a perspective. If the reader of a scenario has reason to doubt that they do in fact share a perspective with the agent at the center of the case being considered, then the scenario will not function as a compelling counterexample. Before getting to how this constraint follows from a proper understanding of the hypothetical cases that are used in theorizing, let's look at a different way of understanding these cases that gives rise to a different constraint that will not work.

One way to understand the nature of the hypothetical scenarios that are relied on in philosophical theorizing is to think of these as counterfactual worlds. To describe a counterfactual world is to describe a way that the actual world might have been but is not. It is typically held that when we consider a possibility counterfactually, judgments about a concept's extension relative to that possibility are determined by facts about that concept's extension in the actual world. This is especially true when the concept in question is that of a natural kind or name, which are standardly held to refer rigidly, such that "in every possible world it designates the same object" (Kripke, 1980, 48) If it is true that concepts pick out their extensions rigidly in counterfactual worlds, then the truth value of judgments involving

those concepts in counterfactual worlds will be determined by the extensions of those concepts in the actual world. Thus, truths concerning what concepts pick out in counterfactual worlds are constrained by facts about the actual world.

For example, Putnam's twin earth scenario is typically thought of counterfactually. The judgment "This is water" made by an agent on twin earth is typically taken to be false. This is so because the concept 'water' picks out substances primarily composed of H₂O molecules in the actual world, and this fact about the concept's extension in the actual world, places constraints on judgments involving that concept in the Twin Earth scenario, or any counterfactual world. The agent at the center of Putnam's scenario cannot speak truly when he says that the watery stuff, which is actually XYZ and with which he is acquainted, is water, because water refers to H₂O in this world, and therefore in every possible world, including Twin Earth.

Thinking of the hypothetical scenarios that are appealed to in philosophical theorizing as counterfactual worlds does put constraints on descriptions of those worlds and thus, may alleviate some of the skeptical concern directed toward this practice. However, the constraints put in place can seem overly strong. The practice of allowing judgments about hypothetical cases to guide theorizing is most characteristic of the project of conceptual analysis. Conceptual analysts typically hold that judgments made in relation to hypothetical cases have value and reflect the concepts that they are attempting to analyze, just in case those judgments result from conceptual competence. It is not clear that this competence depends on possession of a background theory of what that concept picks out in the actual

world. After all, one can be competent with the concept water without knowing that the concept's extension in the actual world is captured by the definition "the substance primarily composed of H₂O molecules". Understanding the hypothetical scenarios as counterfactual worlds does not seem to square with the dominant view among conceptual analysts of how judgments made in those worlds get their value.

The same sort of concern applies to thinking of the hypothetical scenarios relied on in philosophical inquiry as thought experiments. Thought experiments are widely utilized in other disciplines and generally held to be effective. Perhaps one way to respond to the skepticism that surrounds the method of cases in philosophical theorizing is to associate the hypothetical scenarios relied on here with thought experiments. However caution is needed here, as Bealer nicely illustrates:

"Traditionally, in a thought experiments one usually elicits a physical intuition ... about what would happen in a hypothetical situation in which physical, or natural, laws (whatever they happen to be) are held constant but physical conditions are in various other respects non actual and often highly idealized (e.g., so that it would be physically impossible for observers to be present or it would be physically impossible for anyone to conduct the experiment). A classic example is Newton's thought experiment about a rotating bucket in an otherwise empty space. Would water creep up the side of the bucket (assuming that the physical laws remained unchanged)??" (1998, 207)

In Newton's thought experiment, the physical conditions are changed, we assume an empty space, but physical or natural laws are held constant, and it is in relation to these laws that we are to try and answer the question posed. If these laws were not held constant, then it is hard to see how we would go about answering the question at all. Perhaps this is why Bealer notes that rational intuition, "remains silent " on this type of question and goes on to differentiate

this type of question from those usually considered by philosophers, who are “generally concerned about whether a case is possible (logically or metaphysically) and whether a concept applies to such cases” (1998, 207)

Though I agree with Bealer’s distinction, he does little to provide a positive answer to the question of how we are to understand the hypothetical cases relied on in philosophical inquiry if not as thought experiments. Below I’ll present what I take to be the proper way of understanding the nature of the hypothetical scenarios that are appealed to in conceptual analysis. I’ll then show how the general criterion of compellingness outlined above, follows from this understanding of hypothetical cases.

4.3 Hypothetical Scenarios, Centered Worlds, and Perspective

We have considered two ways of understanding the types of hypothetical cases that are relied on in conceptual analysis: as counterfactual worlds and as thought experiments. I have tried to show that neither approach fits well with what it is that philosophers understand themselves to be doing when they appeal to hypothetical cases to analyze concepts. I think we can make better sense of the character of the cases that philosophers typically appeal to when they investigate concepts. This involves understanding the hypothetical cases appealed to in theorizing as centered worlds. Adopting a centered world account of hypothetical scenarios avoids the “anything goes” objection and puts in place constraints that are more in line with the aims of conceptual analysis.

The description of a centered world, unlike a counterfactual world, will necessarily involve truths concerning the perspective of an individual who occupies a position in that world. Chalmers makes this point, when discussing centered worlds, “We can think of the center of the world as representing the perspective of the [individual] within the world” (2006, 60). So does Liao, who arrives at the following definition of a centered world, after careful consideration of how the idea is presented in Stalnaker, Perry and Lewis - “I take a centered world to be the combination of a possible world and a perspective within it – the center.” (2012, 298)

If conceptual analysis utilizes centered worlds to investigate conceptual truths, and if descriptions of centered worlds always contain truths concerning the perspective of an individual at the center of that world, then it will also be true that the conceptual truths that we arrive at through these analyses, will hold relative to a certain perspective, the perspective of the agent at the center of that world. This makes understanding the nature of this perspective important. The question of what exactly constitutes the perspective of an individual at the center of a centered world is one to which we will now turn.

There are at least two ways to understand the notion of perspective in discussions of centered worlds. On a ”thin” more familiar view, perspective will be wholly constituted by indexical concepts, such as “I” and “now”. Liao defends something like the “thin” view of perspectives in his paper. Discussing how to understand the individual at the center of a centered world, Liao argues for the “primitive identification account” according to which we

“let each center be picked out by a cheap identity property. On the Lewisian view of properties, since every set of possible individuals is a property, trivially, there is a property that corresponds to every possible individual (or more precisely, every unit set of an individual). Call such properties identity properties. Trivially, since every possible individual has its own unique identity property, no two possible individuals can have the same identity property.”(2012, 313)

Liao goes on to note “on this account, the individuation of centers and the individuation of possible individuals are primitive, unable to be elucidated through other non-trivial features”. We might here add to the parameters that individuate centers a “time parameter (t)” but this additional parameter is “otiose” and does not add anything substantial to the account of perspective that is necessarily included in any description of a centered world (2012, 313).

The primitive identification account, though useful perhaps for Liao’s purposes, can seem impoverished in the present context. Here, centered worlds are of interest because they provide an explanation of the kind of hypothetical cases that philosophers are reliant on when they use judgments about how concepts apply in hypothetical cases to gain knowledge of conceptual truths. It is important to note that the “thin” view of the individual perspective that characterizes a centered world sheds no light on how the application judgments involving concepts are made relative to that centered world. Understanding how concepts are applied in centered worlds requires broadening the account of perspective that is at play in descriptions of centered worlds.

The “thin” view of perspective that is necessarily constitutive of an individual who constitutes the center of a centered world can be contrasted with a “thick” view according to which the perspective that constitutes a centered world is composed of indexical concepts,

(like “I” and “now”) but is not limited to these. On the thick view of perspective, this is also composed of primitive concepts.

Chalmers argues that it is possible to divide the concepts that occupy conceptual space into two groups, primitive concepts, and non-primitive concepts. Non-primitive concepts get their content in virtue of their inferential role. The non-primitive concept “lying” for example, will get its content via its inferential relations to the concepts “assertion” and “truth” (316, 2012). To understand what it is to tell a lie, one must first understand what it is to assert something and what it is for a statement to be true. The content of a primitive concept, on the other hand, is not constituted via inferential relations to other concepts. A primitive concept is held to be “a concept such that no other concept is conceptually prior to it” (2012, 472). Primitive concepts here serve as “anchors” that ground inferences involving non-primitive concepts (2012, 463).

Chalmers (2012) provides several heuristics for identifying primitive concepts. I want to focus here on two features that primitive concepts share: Non Twin Earthability and transparency.²⁴ To identify primitive concepts we can use the Twin Earth test. If a concept is primitive then it will not be Twin Earthable. Primitive concepts are unlike the concept “water”, which Putnam’s thought experiments have taught us, are Twin Earthable - functional and phenomenal twins have different water concepts reflecting differences in their environments. Contrast this with a concept such as “consciousness”, or “zero”. It is not clear how environmental differences can cause functional and phenomenal twins to have

²⁴ For Chalmers’ discussion of these two ideas see, *Constructing The World*, pp. 316-319 and 327-378, respectively.

different concepts here. The content of a primitive concept remains stable through environmental differences. If a concept is not Twin Earthable, then this is a good indication that it is primitive.

Furthermore, unlike a concept such as water, primitive concepts are conceptually transparent. “conceptually they appear to be transparent: no referential opacity is revealed by conceptual reflection. On the face of it, ‘zero’ transparently picks out zero, and ‘consciousness’ transparently picks out consciousness... they seem to give a direct grip on their referent”. (2012, 374) We do not need a set of necessary and sufficient conditions to know what consciousness, is, neither do we need an account of the underlying physical correlates of consciousness in the actual world to know what consciousness is. Of course, it would be nice to know these things, and we could and should attempt to approximate the content of a primitive concept by supplying a set of necessary and sufficient conditions, just as we could and should dive into empirical investigations to try and figure out what the physical correlates of consciousness are in the actual world, but it would be strange to claim that we do not know what consciousness is until we have settled either of these issues. Grasping a primitive concept grants us the ability to categorize certain experiences as belonging or not belonging within its extension. It is because we possess the ability to make these distinctions that we are licensed to claim some knowledge of the phenomenon, such that we can then go out and construct theories that attempt to make sense of these judgments.

Which concepts are Non Twin Earthable and conceptually transparent? Chalmers (2012, 390) provides the following list of primitive concepts, noting that the list is extremely tentative and flexible.

Logical concepts, such as *not* and *there exists*

Mathematical concepts, such as a *set*

Phenomenal concepts, such as *consciousness*

Spatiotemporal concepts, such as *space* and *time*

Secondary qualities, such as *color*

Nomic concepts, such as *law*

Fundamentality concepts, such as *in virtue of*

Normative/evaluative concepts, such as *ought*

Indexical concepts, such as *I, now, this*

The question of which concepts are members of the set of primitive concepts is open to debate. However, as long as there is some limited set of primitive concepts, then we will have found a good candidate for the constitutive elements out of which the perspective of the individual at the center of a centered world is built. Because conceptual truths are tested against application judgments made in relation to descriptions of centered worlds, primitive concepts will provide a context in which conceptual truths are investigated and in which conceptual truths hold. When we investigate concepts by looking at how they are applied in centered worlds, these judgments will be made relative to the perspective of the “I”, with its set of primitive concepts in tow. Conceptual truths thus arrived at will hold from the perspective of the individual or individuals that share similar sets of primitive concepts.

Furthermore, primitive concepts ground inferences involving non-primitive concepts and so can shed light on how it is that the individual at the center of a centered world goes about making the judgments that are relied on in conceptual analysis.

4.4 Criterion of Compellingness

Let's recap. It has been argued that the best way to understand the hypothetical cases that are relied on in conceptual analysis is as centered worlds. The description of a centered world, always involves truths about the perspective of an individual, the center, of that world. Perspective, on the “thick” view presented above, is constituted by primitive concepts. Thus, conceptual truths hold relative to a perspective constituted by the set of primitive concepts. If all of this is true, then, it follows that in order to function as a compelling counterexample to a given conceptual truth, a hypothetical case must be described so that the agent at the center of that scenario has a perspective that is constituted by primitive concepts. Otherwise, it is not clear that we could expect the conceptual truth at issue to hold for that case in the first place.

Take the Gettier case, the prototype of a compelling counterexample. The conceptual truth being analyzed here is expressed in the statement “Knowledge is justified true belief”. At the time Gettier was writing, concepts were thought of as having definitional structure. ‘Justified’, ‘true’, and ‘belief’ are thought of as independently necessary and jointly sufficient conditions for application judgments involving the concept “knowledge”. Based on what has been said in the last chapter, we can see things differently. We reject the

definitional view of content in favor of an intensional view, wherein concepts are thought of as functions that map possibilities to extensions. Here, the purported conceptual truth should be taken as saying that, in any centered world, the complex concept “justified true belief” and the concept ‘knowledge’ have the same epistemic intension, returning the same extension in every centered world. If so, the statement “Knowledge is justified true belief” will be epistemically necessary. In order to show that this is not so, we can describe a centered world wherein these concepts pick out different extensions. Gettier cases are cases wherein the intensions of these concepts come apart. The subject at the center of this hypothetical case, the subject whom we are attempting to use as a test case for the conceptual truth at issue, is judged to have a justified true belief, but to lack knowledge. What I want to claim is that one reason why this case functions as a compelling counterexample to the conceptual truth at issue, is that we share a perspective with the subject at the center of that case. It is only if we do so that we would expect that the conceptual truth would hold for them in the first place, or that this case would be a legitimate one to rely on inquiry. The reason is that conceptual truths, as they are here understood, are taken to explain the inferential roles of various concepts. According to the picture that was supplied earlier, the inferential role of a concept will be anchored in that subject’s set of primitive concepts. So we wouldn’t expect, that this truth would hold for subjects who do not share sets of primitive concepts.

Let’s go back to Swampman. Davidson does not explicitly claim to be investigating a conceptual truth, but from our perspective, we can say that the conceptual truth at issue here is something like “meaningful mental states are both inner (identical to states of the body)

and non-individualistic (constituted by events and objects outside the subject whose states they are)”. One way to test this claim is against a subject who has inner mental states, but very meager ties to the external environment. Swampman seems to fit the bill. If we judge that Swampman, despite having inner states identical to Davidson, nevertheless does not have mental representations that qualify as being meaningful, then this will support the claim that strong ties to the external environment are also required for possessing meaningful mental states. In order for the case at issue to provide support for this claim, it needs to be found compelling. It needs to invoke strong intuitions that Swampman (the subject at the center of the scenario and whom we are using as a test case for the claim at issue) lacks meaningful mental states. But the Swampman case has not been found to be compelling. I want to claim that the reason why, as Ludwig puts it, “our intuitions about whether Swampman has thoughts are remarkably silent”, is that the assumption of a shared perspective here breaks down (1996, 101). If we don’t share a perspective with Swampman, then it is not clear that the conceptual truth being tested should hold for him anyway. This explains why the case has failed as a compelling counterexample.

What I want to suggest then, is that we should recognize how important the notion of having a shared perspective is to the functioning of a compelling counterexample. Once we do so, we can outline a constraint on the type of cases that are likely to be most profitably used in conceptual analysis. In order to increase the chances that others will find a given case compelling, we should construct the proposed counterexample such that the agent at the center of that counterexample (the agent whom we are using as a test-case for

the conceptual truth at issue) is one that we can reasonably expect to share a perspective with.

One question that arises is how we decide whether we do share a perspective with the agent at the center of a case we are considering. One heuristic that is useful here is to attempt to imaginatively put oneself in the agent's shoes. When philosophers offer a hypothetical scenario meant to elicit an intuitive judgment either for or against a given thesis, they do not preface the discussion by directing the reader to understand the hypothetical scenario as a centered world, they do not say "Imagine that you are at the center of the world described below", or "Imagine that the actual world were such that ..." but I think that this move is implicit in our consideration of these cases. We try to imagine that we are at the center of the case described. If we can't easily do this, or if doing so changes important features of the case, then this gives us a good indication that we do not share a perspective with the agent at the center of that case. As Papineau says of the Swampman case, "Can we really be sure when we imagine Swampman that we are really imagining a randomly created being with no history of selection or learning? It's quite a hard thing to imagine" (1996, 131). This inability to imagine what it is like to be the agent at the center of the scenario described is a good indication that we do not share a set of primitive concepts with this agent, and that we cannot reasonably expect that their perspective is one that we share.

4.5 Objections

Thinking of hypothetical scenarios as centered worlds does put constraints on the construction of these scenarios. Setting constraints on the practice of relying on hypothetical cases in theorizing allows for a response to “anything goes” objection that was noted above. At the same time, increasing understanding of the type of cases that can be most profitably relied on in analysis helps to increase the effectiveness of this practice. However, while the presence of such a constraint can seem advantageous from this point of view, it might also be viewed as an unnecessary limitation. There are three potential worries here that I would like to address.

According to what has been said above, the conceptual truths arrived at through consideration of intuitive judgments about how the relevant concepts apply in centered worlds will hold relative to a certain perspective, a perspective constituted, in part, by primitive concepts. Because of this, these truths will have a limited applicability. They will only hold relative to scenarios where there is a subject who possesses a set of primitive concepts. But it is possible to describe un-centered scenarios where there is no subject. The truths arrived at through investigation of centered worlds, will not apply to un-centered scenarios. But I don’t think much follows from this. Consider for example a generally held thesis from epistemology. The entailment thesis holds that propositional knowledge entails belief. Does this truth hold in scenarios where there aren’t any agents? Why should we expect that it would? The entailment thesis should be interpreted as a truth about subjects who possess the concepts ‘belief’ and ‘knowledge’. It should be read as saying that for these

agents, knowing a proposition entails believing that proposition. I see no problem with the claim that this thesis does not apply to scenarios where there aren't any agents.

The second potential objection to the view developed above concerns the claim that in order to make a judgment about a scenario the agent considering that scenario must be able to take on the perspective of the agent at the center of that scenario. This involves sharing a similar set of primitive concepts with the agent at the center of that scenario. One might doubt that we do share primitive concepts with other agents, so that we are unable to adopt the perspective of the agent at the center of many scenarios. Thus, the reliance on hypothetical scenarios would be severely restricted.

I think it is reasonable to expect that most people share similar sets of primitive concepts. If we look at the tentative list of primitive concepts suggested above, then this seems like a reasonable assumption. Furthermore, if judgments involving non-primitive concepts are based on primitive concepts, then it is reasonable to expect that judgments involving non-primitive concepts would also be widely shared. But this is not necessarily the case in practice. The variability among judgments relative to any given scenario might make one think that the parties to variability share different sets of primitive concepts. It is important to keep in mind here, that disagreements in practice can result from any number of influences. Differences in judgments can result from what Nagel has called “generic motivational problems” encountered when those considering a particular scenario “are not interested in a particular story” and therefore “may be more inclined to respond randomly” (2012, 516). Disagreements may also arise when the scenario in question “is somewhat under

described, and ... participants are not fleshing out the details the same way (as each other, or as other professionals would)" (2012, 519). Differences in judgments concerning hypothetical scenarios can also result from a failure to make the inferences that are warranted by the conjunction of primitive concepts and the description of the case one is considering. Other times, seeming disagreements about a particular case can be explained away as verbal disputes. For these reasons, the claim that we generally share a similar set of primitive concepts and that primitive concepts "anchor" inferences involving non-primitive concepts is consistent with what at first glance appears to be disagreement in judgments about hypothetical scenarios. Still, after accounting for all of these factors, we might find that disagreement remains, but this still does not necessarily mean that the parties who disagree share different sets of primitive concepts. It may be that the conjunction of primitive concepts and the description of the case in question is not sufficient to warrant an inference about how concepts are to be applied in that case. Here, we may need to make use of further principles to arrive at a conclusion. If this is so, variation can be explained by noting that the parties in question have appealed to different principles to decide the case, rather than supposing that they possess different sets of primitive concepts.

This brings us to a third potential objection. If, as I think, apparent disagreements among responses to hypothetical cases gives us little reason to doubt that we share similar sets of primitive concepts and therefore, in this regard similar perspectives, then it can seem as though the criterion of compellingness presented above is trivial because we will nearly always share primitive concepts with the agent at the center of the hypothetical case that we

are considering. The criterion of compellingness will seem trivial because it will nearly always be met. But this is not so, as we see will by looking at a couple of examples.

We have already seen how the assumption that we share a perspective with the agent at the center of a particular case breaks down in the Swampman case. Let's now turn to a more recent example of the use of hypothetical scenarios where the assumption that we share a similar set of primitive concepts with the agent at the center of those scenarios breaks down. Because this is a more recent case, it is early to assess whether the cases in question will ultimately be successful or function as compelling counterexamples to the thesis at issue. Nevertheless, we can see what predictions would be made here using the criterion of compellingness presented above.

Murray, Sytsma and Livengood (2013) argue that the entailment thesis, which holds the propositional knowledge entails belief, should no longer be considered the default view in epistemology. Their argument is based on four purported counterexamples to the entailment thesis. The authors' show that a significant percentage of adult Americans intuitively judge in four varied scenarios that the agent in those scenarios has knowledge of a proposition but does not believe that proposition. If intuitive judgments do diverge from those predicted by the entailment thesis in this way, the entailment thesis can no longer be defended based on its intuitive obviousness and should no longer be regarded as the default view in epistemology. I agree that finding compelling counterexamples to the entailment thesis provides a reason for doubting that thesis. But I don't think that three of the four

counterexamples to the entailment thesis offered by Murray, Sytsma and Livengood are compelling. Again, the criterion of compellingness developed above can help explain why.

Following on work by Radcliff (1966), and Myers-Schulz and Schwitzgebel (2013), Murray, Sytsma and Livengood (2013) empirically test responses to four new purported counterexamples to the entailment thesis.

The God Scenario

The first experiment utilized a between subject-design, each of 101 participants was randomly assigned to one of two yes/ no questions:

- 1) Does God know that $2+2=4$?
- 2) Does God believe that $2+2=4$?

93.1% of subjects answered that God knows, while 65.9% answered that God believes.

The Dog Scenario

In a second study utilizing a within-subjects design, 190 subjects were given the following vignette:

Researchers have found that some breeds of dogs are surprisingly intelligent. Among the most intelligent are border collies. One border collie, Cassie, is even able to solve simple mathematical problems: If you ask her “what is $2+2$,” for example, she will bark four times; similarly, if you ask here “what is $4+5$ ” she will bark nine times!

Each subject was then asked two yes/no questions (counterbalanced for order):

1) Does Cassie know that $2+2=4$?

2) Does Cassie believe that $2+2=4$?

53.2% of total respondents said that Cassie knows. 52.5% of these respondents said that Cassie does not believe. 61.1% of total respondents said that Cassie does not believe. 45.7% of these respondents answered that she nevertheless knows.

The Cash Register Scenario

In a third study, utilizing a within subjects design, 194 subjects were given the following vignette:

It is an ordinary Saturday morning and Mary is at the supermarket picking up some groceries. When she went to check out there were lines in every lane that had a human cashier; several of the new “self-checkout” lanes had no wait, however. In the self checkout lane there is an item scanner attached to a cash register: You scan your items and when you are done the cash register displays how much you owe and takes your payment (having slots for cash or credit cards).

Mary had never used one of these self-checkout lanes before and was a bit nervous about doing so. Nonetheless she didn’t want to wait in line and decided to give it a try. Everything went fairly smoothly. When she was done she hit total. The cash register then displayed the total amount that Mary owed: \$42.73. Mary double-checked the amount displayed and found that it was accurate. It took her some time to do this, however, and after a minute the cash register started to beep. When Mary inserted her credit card, the beeping stopped. Mary finished paying with no further trouble. Overall, Mary thought the self-checkout lane worked quite well and thought that she would use it again.

Subjects were then asked two yes/no questions (counterbalanced for order).

Did the cash register know that Mary owed \$42.73?

Did the cash register believe that Mary owed \$42.73?

88.6% of total respondents answered that the cash register knew. Of these respondents, 34.9% answered that it did not believe. 38.1% of total respondents answered that the cash register didn't believe. Of these respondents, 81.1% answered that it nevertheless knew.

The Geo-Centrist Scenario

In a fourth study utilizing a within subjects design, 98 participants were given the following vignette:

Karen is a first year student at a prestigious university. She is a good student and has been doing very well in her classes. One of the classes she is taking is introduction to physics. One of the topics covered in this class is the place of the earth in the solar system. For example, one of the things that Karen has been taught is that the earth revolves around the sun.

Prior to starting at the University, however, Karen was home-schooled by her parents. Karen's parents taught her that the earth is at the centre of the universe. Karen accepts what her parents have taught her. In particular, in spite of what has been taught in her physics class, she holds that the earth does not revolve around the sun.

One of the questions on Karen's final exam is the following:

True or False: The earth revolves around the sun.

Karen answers "true" on this question. She gets the answer correct and ends up scoring 100% on the exam.

Each participant was then asked two yes/no questions (counterbalanced for order):

Does Karen know that the earth revolves around the sun?

Does Karen believe that the earth revolves around the sun?

54.1% of total respondents said that Karen knows. Of these respondents, 84.9% answered that she didn't believe. 81.9% of total respondents answered that Karen doesn't believe. Of these respondents, 56.3% answered that she nevertheless knows.

The authors hold that the results obtained in all cases run counter to those that are predicted by the entailment thesis. The entailment thesis, according to Murray, Sytsma and Livengood, makes two predictions about how people will respond to these types of cases. 1) If subjects judge that an agent knows a proposition then they will also judge that that agent believes that proposition. 2) If subjects judge that an agent does not believe a proposition, then they will also judge that that agent does not know that proposition. But these predictions are not born out by the data.²⁵ Thus, it seems, the entailment thesis fails to account for judgments about cases, and because of this, has no right to be regarded as the default view in epistemology based on its intuitive obviousness. I do not find that the first three counterexamples just presented are compelling, that is I do not find that they provide a reason for doubting the entailment thesis. This reaction to the cases is predicted by the criterion of compellingness presented above.

The first step to seeing how the claim that these cases are not compelling is consistent with the criterion of compellingness presented above is to recognize that the concept "belief" is a primitive concept. We have a kind of direct grip on what a belief is. This kind of direct grip on the referent of the concept belief is evidenced in our ability to apply the concept in a wide variety of scenarios. We can typically, though not always, easily

²⁵ The authors report that all results were tested and found to be statistically significant using one-sample test of proportion with continuity correction.

categorize mental states as falling (or failing to fall) within the extension of the concept belief. We can do this without possessing a definition or set of necessary and sufficient conditions. It seems that knowing what a belief is does not depend on knowing what some other concepts are, rather we know what a belief is simply because we are believers.²⁶ Furthermore, it is hard to see how the concept belief is Twin Earthable; changes in the external environment do not seem to necessitate changes in the concept belief between functional and phenomenal duplicates.

If we take the concept belief to be a primitive concept and we take conceptual truths to be grounded in judgments made in centered worlds, and we take a “thick” view of centered worlds, where the center of each world is constituted by primitive concepts, then the entailment thesis should be understood as expressing a conceptual truth that holds relative to the perspective constituted by primitive concepts, including belief. But if we understand the entailment thesis as a conceptual truth that holds relative to the perspective of an individual who has the primitive concept of belief, then it is clear why the first three of the counterexamples that are proposed to the entailment thesis are not compelling. They do not present compelling counterexamples to the entailment thesis because the perspective of the agent (God, a Dog, and a cash register) at the center of these scenarios is different from our own. In particular, there is no reason to think that the agents at the center of these scenarios possess the primitive concept “belief”. There is some empirical support for the claim that the concept belief is not shared by other species. Research by Marticorena, Ruiz,

²⁶ The claim that we can grasp the concept belief without a prior grasp of other concepts is controversial. Williamson (2000) takes knowledge to be more fundamental. Adjudicating this debate lies outside the scope of this paper.

Mukerji, Goddu and Satos (2011) and Kaminski, Call and Tomasello (2008) indicates that non-human primates have difficulty attributing beliefs to others, this is to be expected if they do not possess the concept belief. If the agents at the center of these scenarios do not possess this primitive concept, then it is not clear, that the entailment thesis makes any prediction about how the concepts ‘belief’ and ‘knowledge’ would behave relative to these scenarios. The entailment thesis will only make predictions about how the concepts ‘belief’ and ‘knowledge’ relate to one another, in worlds centered on those individuals whom we can reasonably expect to share a perspective with, and it is far from clear that we should expect to share a perspective with God, a Dog, or a cash register. This still leaves the fourth scenario offered by the authors standing. It seems we do share a perspective with the agent at the center of the Geo-Centrist example. But showing that the first three scenarios are not compelling dampens the arguments against the entailment thesis quite a bit. It is a bit early to say whether the recent interest in the entailment thesis will ultimately show it to be unjustified. However, in so far as the resolution to this debate is decided by judgments about hypothetical scenarios, the criterion of compellingness outlined above predicts that it will be the fourth type of scenario that will ultimately carry the day and that we’d do just as well to leave God, Dogs and cash registers out of the discussion.

We have looked at several instances where the criterion of compellingness predicts that the hypothetical cases appealed to in debate are not compelling because we are unable to take on the perspective of the agent at the center of those scenarios. The criteria is not trivial, despite it’s being true that we can, in general, reasonably expect to share a perspective

with the majority of human agents. It's time to apply the criterion more broadly to see how it handles other cases.

4.6 Chinese Nation, Chinese Room

The practice of appealing to judgments about hypothetical cases has played a significant role in shaping debates in the philosophy of mind. In this section we will look at two hypothetical scenarios from this area to show that the criterion of compellingness correctly predicts reactions to these cases.

The first hypothetical scenario we will consider appears in the debate concerning Functionalism. Here, (briefly) mental states are defined by the causal role they play in a system. Functionalism has the virtue of overcoming the multiple realizability argument that plagues identity theory. The multiple realizability argument is based on the plausible thought that mental states, such as pain, can be realized in various physical systems. Functionalism allows for this. However, critics of functionalism argue that it is too liberal in its descriptions of which systems have mental states. The hypothetical scenario considered below, presented by Block, in his paper “Troubles with Functionalism” is meant to provide a compelling counterexample to Functionalism by providing a case wherein we judge that the system at the center of that case meets the description of what it is to have a mental state provided by functionalism but nevertheless, clearly lacks mental states. Here is the scenario:

“Suppose we convert the Government of China to functionalism, and we convince its officials that it would enormously enhance its international prestige to realize a human mind for an hour. We provide each of the billion people in China with a specially designed two-way radio that connects them in the appropriate way to other persons and to [an] artificial

body... [which contains] a radio receiver and transmitter connected to the input and output neurons... We arrange to have letters displayed on a series of satellites such that they can be seen from anywhere in China."

The display serves as input to the system that then produces the appropriate output. Block notes that such a system would be "functionally equivalent to you for a short time, say an hour" (1978, 279). Thus, if functionalism is true, then the system should have mental states. Now, ask yourself, as reader of this scenario whether the system in question has mental states. Block predicts that you will judge that it does not. The system just described presents "a *prima facie* counterexample to machine functionalism [because there is] *prima facie* doubt whether it has any mental states at all" (1978, 281).

Before discussing this case further, let's consider a second well-known scenario offered by Searle, in his paper "Minds, Brains and Programs." Searle presents his scenario as a counterexample to a view which he refers to as "Strong AI", according to which an "appropriately programmed computer really is a mind - in the sense that computers given the right programs can be literally said to understand and have other cognitive states" (1980, 417). One way to test this view, according to Searle, is to "ask oneself what it would be like if my mind actually worked on the principles that the theory says all minds work on" (1980, 417) To this end, Searle offers the following hypothetical scenario:

"Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore...that I know no Chinese...To me Chinese writing is just so many meaningless squiggles... After this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English and I understand these rules. They enable me to correlate one set of formal symbols with another set of formal symbols... that I can identify ...entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the

third batch. Unknown to me, the people who are giving me all of these symbols call the first batch a “script”, they call the second batch a “story,” and they call the third batch “questions,” and the set of rules in English that they gave me, they call the “program”... Suppose that after a while, I get so good at writing the programs that from the external point of view...my answers to the questions are actually indistinguishable from those of other native Chinese Speakers.... I simply behave like a computer; I perform computational operations on formally specified elements. ...I am simply an instantiation of the computer program.”

Contra to what strong AI predicts Searle holds that “it is obvious in the example that I do not understand Chinese” (1980, 418).

Again the present concern is not with arguing for or against either of these positions. The present concern is with specifying a means of explaining reactions to such cases in order to increase our understanding of the method of cases and to increase its effectiveness. Once again, the criterion of compellingness presented above can help.

The first thing to note here is that while both cases are well known and have had an impact on theorizing, the latter case has had a greater impact than the former. This is evidenced by citation information. According to Google Scholar, “Minds, Brains and Programs” where the Chinese Room appears has more than three times as many citations as “Troubles with Functionalism” where the Chinese Nation appears. Furthermore, both the Stanford Encyclopedia of Philosophy and the Routledge Encyclopedia of Philosophy have separate entries for the Chinese Room, explaining its significance in the development of theorizing, while neither has an entry for the Chinese Nation. Such data supports the claim that Chinese Room case has been found to be more compelling than the Chinese Nation case. This is exactly what the criterion of compellingness would predict.

The criterion of compellingness suggests that a case will be compelling to the extent that we can take on the perspective of the agent at the center of that case, where taking on the perspective of the agent involves sharing a similar set of primitive concepts with that agent. There is reason to think that our ability to take on the perspective of the agent at the center of the case we are considering breaks down in the Chinese Nation example.

Perspective is constituted by primitive concepts. This includes a concept of “phenomenal consciousness”. Perhaps we do not grasp this concept explicitly, but it seems that most people grasp this concept implicitly, having a direct grip on what it is to be phenomenally conscious despite the fact that they may not be able to articulate what phenomenal consciousness is. But it seems unlikely that we would be willing to attribute a similar concept to the system at the center of the Chinese Nation case. The claim that we are unwilling to attribute phenomenal consciousness to something like the Chinese Nation is supported empirically by the work of Knobe and Prinz (2008). Knobe and Prinz’s studies suggest that physical constitution matters to attributions of consciousness. Their research supports the claim that people will be unwilling to attribute phenomenal consciousness to the Chinese Nation. If we do not do this, we cannot easily imagine ourselves at the center of this case. This inability to imaginatively put ourselves at the center of this case, without changing its internal structure alerts us to the fact that the assumption that we share a perspective with the agent at the center of the case breaks down here. Thus, it fails to function as a compelling counterexample, at least when compared with another case like the Chinese Room. The criterion of compellingness presented above predicts that the Chinese Room case will be more compelling than the Chinese Nation case. We can easily imagine

ourselves at the center of this case. Indeed, Searle crafts the case so that it is centered on him. Searle in the Chinese room is the same as Searle out of the Chinese room, except for his increased proficiency translating squiggles into more squiggles.

4.7 Remarks

Of course, there are many hypothetical cases and I cannot consider them all here. Still, I hope to have considered enough cases to show that this criterion of compellingness is at least plausible and deserves further consideration. I would now like to make explicit some predictions that follow from this criterion of compellingness that can be used to test it in future work. The criterion of compellingness presented here holds that given a sufficient description of a case, that case will be found more compelling if the agent at the center of that case, the one whom we are testing the conceptual truth at issue on, is one whom we can reasonably expect to share a perspective with. One way to experimentally test the criterion of compellingness is to provide subjects with the Chinese Nation case and the Chinese Room case that are presented above. Subjects would be split into two groups, each group would be told that they are being asked to consider a potential counterexample to a purported thesis. One group would be given a simplified version of the Strong AI thesis as it appears in Searle's paper, the other group would be given a simplified version of the thesis associated with machine functionalism. Each group would be told that they are going to read a purported counterexample to that thesis. Once they have read the relevant case, they would be asked to indicate on a seven-point scale, how compelling the case is. The criterion of compellingness predicts that the group presented with the Chinese room type

counterexample would report higher levels of compellingness. Another way to test this criterion of compellingness would be to find a case where it is possible to substitute an agent whom we have strong reason to believe we do share a perspective with, with an agent that we have strong reason to believe we do not share a perspective with and give a version of each case to participants to see if the case with the agent that we do share a perspective with, is found to be more compelling as a counterexample. The criterion of compellingness presented here predicts that participants would find more compelling the case with the agent we share a perspective with at the center. Though I need to think more carefully about the most effective way to test this criterion, there is no principled reason why this couldn't be done. Such studies could help to determine the extent to which the criterion of compellingness argued for above does track facts concerning people's judgments about cases.

Conclusion

The introduction to this dissertation stated that its main goal is to arrive at a better understanding of the method of cases. This has been achieved. We have seen how the argumentative burden placed on those who use the method of cases differs depending on what one takes the target of analysis to be. We considered three potential targets of analysis: extra-mental facts, the psychological structures that are used by default in our higher order cognitive processes, and that component of content associated with a concept's epistemic intensions. We looked at possible argumentative strategies that might be used by those who engage in each practice in order to respond to the recent evidence purporting to show that intuitive judgments are untrustworthy evidential sources. In the first chapter we saw that extra-mentalists cannot meet all of the epistemic goals of theorizing without also undertaking investigations of the instrument that is producing the intuitive judgments that are relied on as evidence for uncovering knowledge about external facts. In the second chapter we looked at one way in which the target of intuitive analysis can be refined to meet criticisms directed toward the reliance on intuitive judgments in investigations that target underlying psychological structures. In chapter three, we looked at a recent version of "good old armchair conceptual analysis" and, in chapter four, we saw that it is possible to use resources from this theory to increase our understanding of the kinds of cases that can be most profitably relied on when using intuitive judgments to target a concept's epistemic content.

Undoubtedly, there is a lot of work that needs to be done to fully understand the nature and role of intuitive judgments in inquiry. Much has been made recently about the contributions that can be made to this project by empirical methods of research. I think that

these are great developments. To borrow from Knobe and Nichols, “the thing to do now is just to cast off our methodological chains and go after the important questions with everything we’ve got” (2007, 14). In the rush to incorporate empirical methods of inquiry, there has been a temptation to downplay the significant contributions that can be made from the armchair. Hopefully, Chapter 4 serves as a reminder that it is possible to make contributions to the important questions by reflecting on matters from the armchair as well.

Bibliography

- Alexander, Joshua and Jonathon Weinberg. 2007. "Analytic Epistemology and Experimental Philosophy" *Philosophy Compass* 2(1): 56-80.
- Alexander, Joshua, Ronald Mallon, and Jonathon Weinberg. 2010. "Accentuate the Negative" *Review of Philosophy and Psychology*, 1(2): 297-314.
- Bealer, George. 1998. "Intuition and the Autonomy of Philosophy" in *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Edited by Michael R. DePaul and William Ramsey. Oxford: Rowman and Littlefield Publishers, Inc. 201-240.
- Beebe, James and Ryan J. Undercoffer. 2015. "Moral Valence and Semantic Intuitions" *Erkenntnis*: 1(22): 445-466.
- Block, Ned. 1978. "Troubles with Functionalism" *Minnesota Studies in the Philosophy of Science* 9: 261-325.
- Boyd, Kenneth and Jennifer Nagel. 2014. "The Reliability of Epistemic Intuitions" in *Current Controversies in Experimental Philosophy*. Edited by Edouard Machery and Elizabeth O'Neil. Routledge 109-127.
- Cappelen, Herman. 2012. *Philosophy without Intuitions*. Oxford: Oxford University Press.
- Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, David. 2002(a). "Sense and Intension" *Philosophical Perspectives* 16.
- Chalmers, David, J. 2002(b). "The Components of Content" in *Philosophy of Mind: Classical and Contemporary Readings*. Edited By: David J. Chalmers, New York, NY: Oxford University Press, 608-633.

- Chalmers, David. 2003. "The Nature of Narrow Content" *Philosophical Issues* 13(1): 46-66.
- Chalmers, David. 2006(a). "The Foundations of Two-Dimensional Semantics" in *Two-Dimensional Semantics*. Edited by Manuel Garcia-Carpintero and Josep Macia. Oxford: Oxford University Press, 55-140.
- Chalmers, David, J. 2006(b). "Two-Dimensional Semantics" in *The Oxford Handbook of Philosophy of Language*. Edited by Ernest Lepore and Barry C. Smith. New York, NY: Oxford University Press, 574-606.
- Chalmers, David J. 2011. "Verbal Disputes" *Philosophical Review* 120(4): 515-66.
- Chalmers, David. 2012. *Constructing the World*. Oxford: Oxford University Press.
- Chalmers, David and Frank Jackson. 2001. "Conceptual Analysis and Reductive Explanation" *The Philosophical Review* 110(3): 315-361.
- Cummins, Robert. 1998. "Reflection on Reflective Equilibrium" in *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Edited by Michael R. DePaul and William Ramsey. Oxford: Rowman and Littlefield Publishers, Inc. 113-127.
- Davidson, Donald. 1987. "Knowing One's Own Mind" *Proceedings and Addresses of the American Philosophical Association* 60 (3): 441-458.
- DePaul, Michael R and William Ramsey. 1998. Preface to *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Edited by Michael R. DePaul and William Ramsey. Oxford: Rowman and Littlefield Publishers, Inc. VII-XVI.
- Devitt, Michael. 1994. "The Methodology of Naturalistic Semantics" *Journal of Philosophy* 91 (10): 545-72.

Feltz, Adam, and Edward Cokely. 2009. "Do judgments About Freedom and Responsibility Depend on who you are? Personality Differences in Intuitions About Compatibilism and Incompatibilism." *Consciousness and Cognition* 18: 336-358.

Gettier, Edmund. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121-123.

Goldman, Alvin, I. 2007. "Philosophical Intuitions: their Target, their Sources, and their Epistemic Status." *Grazer Philosophische Studien* 74: 1-26.

Goldman, Alvin, I. 2010. "Philosophical Naturalism and Intuitive Methodology" *Proceedings and Addresses of the American Philosophical Association*.

Goldman, Alvin, and Joel Pust. 1998. "Philosophical Theory and Intuitive Evidence" in *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Edited by Michael R. DePaul and William Ramsey. Oxford: Rowman and Littlefield Publishers, Inc. 180-197.

Huebner, Bryce, Michael Bruno, and Hagop Sarkissian. "What does the nation of China think about phenomenal states?" *Review of Philosophy and Psychology* 1 2010: 225-243.

Jackman, Henry. "Semantic Intuitions, Conceptual Analysis, and Cross-Cultural Variation" *Philosophical Studies* 146 (2): 159-177.

Jackson Frank. 1982 "Epiphenomenal Qualia" *Philosophical Quarterly* 32 (April): 127-136.

Jackson, Frank. 1986. "What Mary Doesn't Know" *Journal of Philosophy* 83 (May): 291-295.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A defense of Conceptual Analysis*. Oxford: Oxford University Press.

Jackson, Frank. 2003. "Narrow Content and Representations, or Twin Earth Revisited" *Proceedings and Addresses of the American Philosophical Association* 77 (2): 55-70.

Jackson, Frank. 2011. "On Gettier Holdouts." *Mind and Language* 26 (4): 468-481.

James, William. 1968. "What Pragmatism Means" in *James: Essays in Pragmatism* Edited By Alburey Castell. New York, NY: Hafner Publishing Company. 141-158.

Kaminski, Juliane, Josep Call and Michael Tomasello. 2008. "Chimpanzees know what others know, but not what they believe." *Cognition* 109: 224-234.

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

Knobe, Joshua. Forthcoming. "Experimental Philosophy is Cognitive Science" in *A Companion to Experimental Philosophy*. Edited By Justin Sytsma and Wesley Buckwalter. Blackwell.

Knobe, Joshua and Shaun Nichols. 2007. "An Experimental Philosophy Manifesto" in *Experimental Philosophy*. Edited By Joshua Knobe and Shaun Nichols. New York, NY: Oxford University Press. 3-14.

Knobe, Josh and Jesse Prinz, 2008. "Intuitions about Consciousness: Experimental Studies" *Phenomenology and the Cognitive Sciences* 7 (1): 67-83.

Kornblith, Hilary. 1998. "The Role of Intuition in Philosophical Inquiry: An Account with no Unnatural Ingredients" in *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Edited By Michael R. DePaul and William Ramsey. Oxford: Rowman and Littlefield Publishers, Inc. 129-141.

Kornblith, Hilary. 1999. "Timothy Williamson's Philosophy of Philosophy", *Analysis* Vol. 69, (1): 109-116.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Laurence, Stephen and Eric Margolis. 1999. "Concepts and Cognitive Science" in *Concepts: Core Readings*. Edited By Eric Margolis and Stephen Laurence. Cambridge, MA: MIT Press, 3-81.

Laurence, Stephen and Eric Margolis. 2003. "Concepts and Conceptual Analysis" *Philosophy and Phenomenological Research*, Vol. LXVII, No.2, 253-282.

Lepore, Ernest and Barry Loewer. 2011. "Solipsistic Semantics" in *Meaning, Mind and Matter: Philosophical Essays* Edited by Ernest Lepore and Barry Loewer. New York, NY: Oxford University Press, 131-151.

Liao, Shen-yi. 2012. "What are Centered Worlds?" *The Philosophical Quarterly* 62 (247): 294-316.

Liao, S. Matthew, and Alex Weigman, Joshua Alexander, and Gerard Vong. 2012. "Putting the Trolley in order: Experimental Philosophy and the Loop Case" *Philosophical Psychology* 25 (5): 661-671.

Ludwig, Kirk. 1996. "Duplicating Thoughts" *Mind and Language*, 11 (1): 92-102.

Machery, Edouard. 2005. "Concepts are Not a Natural Kind" *Philosophy of Science* 72: 444-67.

Machery, Edouard. 2009. *Doing Without Concepts*. New York, NY: Oxford University Press.

Machery, Edouard. 2010. "Précis of Doing without Concepts" *Mind and Language* 25 (5): 602-11.

Machery, Edouard, Ronald Mallon, Shaun Nichols, and Stephen Stitch. 2004. "Semantics, Cross Cultural Style" *Cognition* 92 (3): B1-B12.

Machery, Edouard and Selja Seppala. 2009. "Against Hybrid Theories of Concepts" *Anthropology and Philosophy* 10: 99-127.

- Mallon, Ron, Edouard Machery, Shaun Nichols, and Stephen Stitch. 2009. "Against Arguments From Reference" *Philosophy and Philosophical Research* 79(2): 332-356.
- Malpas, Jeff. 2015. "Donald Davidson" *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edited By Edward N. Zalta.
URL = <<http://plato.stanford.edu/archives/spr2015/entries/davidson/>>
- Marticorena, D.C., Ruiz, A.M., Mukerji, C., Goddu, A., and Santos, L.R. 2011. "Monkeys represent others' knowledge but not their beliefs." *Developmental Science* 14(6): 1406-16.
- Murray, Dylan, Justin Sytsma and Jonathon Livengood. 2013 "God Knows (but does God believe?)" *Philosophical Studies* 166 (1): 83-107.
- Myers-Schulz, Blake and Eric Schwitzgebel. 2013. "Knowing That P without Believing That P." *Noûs* 47 (2): 371-384.
- Nadelhoffer, T. and E. Nahamias. 2007. "The Part and Future of Experimental Philosophy" *Philosophical Explorations*, Vol. 10 (2), 123-149.
- Nagel, Jennifer. 2012. "Intuitions and Experiment: A Defense of the Case Method in Epistemology." *Philosophy and Phenomenological Research* 85 (3): 495-527.
- Nagel, Jennifer, Valerie San Juan, and Raymond A. Mar. 2013. "Lay Denial of Knowledge for Justified True Beliefs" *Cognition* 129: 652-661.
- Papineau, David. 1996. "Doubtful Intuitions" *Mind and Language* 11(1): 130-132.
- Petrinovitch, Lewis, and Patricia O'Neil. 1996. "Influence of Wording and Framing Effects on Moral Intuitions" *Ethology and Sociobiology* 17: 145-171.
- Putnam, Hilary. 1973. "Meaning and Reference" *Journal of Philosophy* 70 (19): 699-711.
- Putnam, Hilary. 1975. "The Meaning of 'Meaning'" *Minnesota Studies in the Philosophy of Science* 7:131-193.

- Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- Radford, Colin. 1966. "Knowledge, By Examples." *Analysis* 27 (1): 1-11.
- Rosch, Eleanor. 1999. "Principles of Categorization" in *Concepts: Core Readings*. Edited By Eric Margolis and Stephen Laurence. Cambridge, MA: MIT Press, 189-206.
- Schroeter, Laura. 2004. "The Limits of Conceptual Analysis" *Pacific Philosophical Quarterly* 85 (4): 425-453.
- Schulz, Eric, Edward Cokely, and Adam Feltz. 2011. "Persistent Bias in expert judgments about free will and moral responsibility: A test of the Expertise Defense." *Consciousness and Cognition* 20:1722-1731.
- Schwitzgebel, Eric, and Fiery Cushman. 2012. "Expertise in Moral Reasoning? Order Effects on Moral Judgments in Professional Philosophers and Non-Philosophers" *Mind and Language* 27 (2): 135-153.
- Seale, John R. 1980. "Minds, Brains and Programs" *Behavioral and Brain Sciences* 3 (3): 417-457.
- Singer, Peter. 1972. "Famine, Affluence and Morality" *Philosophy and Public Affairs* 1(3): 229-243.
- Smith, Edward and Douglas Medin. 1999. "The Exemplar View" in *Concepts: Core Readings*. Edited By Eric Margolis and Stephen Laurence. Cambridge, MA: MIT Press, 207-221.
- Sosa, Ernest. 2007. "Experimental Philosophy and Philosophical Intuition" *Philosophical Studies* 132 (1): 99-107.
- Swain, Stacey, Joshua Alexander, and Jonathon Weinberg. 2008. "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp." *Philosophy and Phenomenological Research* 76 (1): 138-155.

Turri, John. 2013 “A conspicuous art: putting Gettier to the test” *Philosophers' Imprint* 13(10): 1-16.

Weinberg, Jonathon. 2007. “How to Challenge Intuitions Empirically without Risking Skepticism.” *Midwest Studies in Philosophy* 31 (1): 318-343.

Weinberg, Jonathon. 2009. “On doing better, experimental style.” *Philosophical Studies* 145: 455-464.

Weinberg, Jonathon, Shaun Nichols, and Stephen Stitch. 2001. “Normativity and Epistemic Intuitions.” *Philosophical Topics* 29 (1-2): 429-460.

Weinberg, M. Jonathon, Stephen Crowley, Chad Gonnerman, Ian Vanderwalker, Stacey Swain. 2012. “Intuition & Calibration” *Essays in Philosophy* Vol. 13, (1), Article 15: 256-283.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

Zamzow, Jennifer, and Shaun Nichols. 2009. “Variations in Ethical Intuitions” *Philosophical Issues* 19: 368-388.