KNOWLEDGE ASCRIPTION AND TRADITIONAL EPISTEMOLOGY

by

JON WESLEY BUCKWALTER

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Jesse Prinz

Date

Chair of Examining Committee

Dr. Iakovos Vasiliou

Date

Executive Officer

Dr. Michael Devitt

Dr. John Greenwood

Dr. Stephen Neale

Dr. Jonathan Schaffer

Supervising Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

KNOWLEDGE ASCRIPTION AND TRADITIONAL EPISTEMOLOGY

by

Jon Wesley Buckwalter

Advisor: Professor Jesse Prinz

The principle thesis of this dissertation is that understanding the psychological factors that underlie epistemic judgment through knowledge ascription is essential for progress in traditional epistemology, and that the tools developed across the cognitive sciences are necessary for collecting accurate evidence concerning the nature of these factors. Chapters are displayed as cumulative proof of concept for this thesis. Chapter 1 begins with a discussion of the role of ascription and ordinary language practices in epistemic arguments. The subsequent sections of the dissertation then present experimental evidence advancing new understanding of the judgments we ordinarily make about knowledge ascription, together with a discussion of how this understanding comes to bear on a series of significant and ongoing debates in contemporary epistemology. Chapter 2 displays evidence for the claim that the ordinary concept of knowledge is *factive*. Chapter 3 explores the ways in which pragmatic factors like *stakes*, *error salience*, or attributor *accommodation* influence knowledge ascription, and subsequently, the implications these findings have for adjudicating between certain arguments given in support of contextualism and interest-relative invariantism. Chapter 4 gives evidence for a *moral component* of knowledge attribution, and shows how this effect of normative judgments on epistemic judgments may provide new insight into enduring philosophical puzzles like Gettier problems. Lastly, Chapter 5 surveys recent evidence suggesting that epistemic judgments are prone to performance errors and

demographic variation, that may well threaten to undermine a substantial set of epistemic projects unless the empirical study of epistemic intuitions is incorporated into methodological approaches to the study of knowledge. Chapter 6 is a brief conclusion suggesting areas for further study, as well as how applying these new methods may relate to larger research programs in psychology and cognitive science.

Preface

Experimental philosophy is a subfield that uses methods and techniques typically associated with psychology and the cognitive sciences to investigate questions or phenomena normally of interest to philosophers (for a review of this research see Knobe et al. 2012 and Buckwalter 2012). As a work of both experimental philosophy and epistemology, the content of this dissertation may be viewed by some as seditious, or somehow subversive, to philosophical tradition. But I submit that the ideas presented here flow directly from the lessons we (ought to) have learned from the history of modern philosophy, together with a great longing to return to that august tradition. As one of my highest philosophical influences, and for without whom this work would not have been possible, it was David Hume, who wrote:

> Here then is the only expedient, from which we can hope for success in our philosophical researches, to leave the tedious lingering method, which we have hitherto followed, and instead of taking now and then a castle or village on the frontier, to march up directly to the capital or center of these sciences, to human nature itself; which being once masters of, we may every where else hope for an easy victory. From this station we may extend our conquests over all those sciences, which more intimately concern human life, and may afterwards proceed at leisure to discover more fully those, which are the objects of pore curiosity. There is no question of importance, whose decision is not comprised in the science of man; and there is none, which can be decided with any certainty, before we become acquainted with that science. In pretending, therefore, to explain the principles of human nature, we in effect propose a complete system of the

sciences, built on a foundation almost entirely new, and the only one upon which they can stand with any security.

And as the science of man is the-only solid foundation for the other sciences, so the only solid foundation we can give to this science itself must be laid on experience and observation. It is no astonishing reflection to consider, that the application of experimental philosophy to moral subjects should come after that to natural at the distance of above a whole century; since we find in fact, that there was about the same interval betwixt the origins of these sciences; and that reckoning from THALES to SOCRATES, the space of time is nearly equal to that betwixt, my Lord Bacon and some late philosophers [Mr. Locke, my Lord Shaftesbury, Dr. Mandeville, Mr. Hutchinson, Dr. Butler, etc.] in England, who have begun to put the science of man on a new footing, and have engaged the attention, and excited the curiosity of the public. So true it is, that however other nations may rival us in poetry, and excel us in some other agreeable arts, the improvements in reason and philosophy can only be owing to a land of toleration and of liberty.

Nor ought we to think, that this latter improvement in the science of man will do less honour to our native country than the former in natural philosophy, but ought rather to esteem it a greater glory, upon account of the greater importance of that science, as well as the necessity it lay under of such a reformation. For to me it seems evident, that the essence of the mind being equally unknown to us with that of external bodies, it must be equally impossible to form any notion of its powers and qualities otherwise than from careful and

> exact experiments, and the observation of those particular effects, which result from its different circumstances and situations. And though we must endeavour to render all our principles as universal as possible, by tracing up our experiments to the utmost, and explaining all effects from the simplest and fewest causes, it is still certain we cannot go beyond experience; and any hypothesis, that pretends to discover the ultimate original qualities of human nature, ought at first to be rejected as presumptuous and chimerical.
>
> A Treatise of Human Nature
> Introduction to Book I: 6-8

I believe that there is much we can learn from Hume. The subsequent pages follow in his example, applying the Humean idea of the study of the science of man to questions of contemporary epistemology. For perhaps the nature of knowledge within the grasp of human understanding resides principally in the study of *our experience of it*—the conditions under which we ascribe or deny it, and the psychological mechanisms that support our epistemic judgments and knowledge ascriptions.

DEDICATION

I dedicate this thesis to my parents John and Lizabeth, who together with their love and encouragement have always worked to make my education their first priority; to my sister Lacie Rae, by far the most promising academic in the family; and to my wife Carolyn, a constant source of inspiration, scientific creativity, and strength.

Acknowledgments

This project would not have been possible without the support of many people. I would like to thank my committee members Stephen Neale and John Greenwood, as well as Michael Devitt, for our lively critical discussions of experimental philosophy over the years, and to Jonathan Schaffer of Rutgers University, for his friendship and keen insight into the epistemic questions explored in this dissertation. Special thanks to my advisor and committee chair Jesse Prinz, for his continued support throughout my graduate studies, and whose work has served as the highest model of how sophisticated scientific inquiry can shed new light on perennial philosophical questions. Also many thanks to Edouard Machery of the University of Pittsburgh, and to James Beebe of the University at Buffalo, for early guidance as I began graduate school, and whose work continues to greatly inspire me.

I would also like to express my deep appreciation and gratitude to Professor Joshua Knobe of Yale University, whose passion and love of philosophy is matched only by his incredible generosity and willingness to share his time and talents with others. Lastly, I am indebted to Professor Stephen Stich of Rutgers University, the person who first inspired me to become a philosopher, and the person who, through extensive training, funding, and collaboration (not to mention the occasional and well deserved philosophical ire), has given the most to ensure that I become a good one. None of this—as well as a great many other philosophical projects, I suspect—would be possible but for him.

Table of Contents

List of Figures

A Brief Introduction

The questions of when, or under what conditions we ordinarily ascribe different sorts of mental states or propositional attitudes to others—how we reasons about their beliefs, desires, or actions—are questions central to both philosophy and psychology. Yet compared to paradigmatic states, such as belief and desire ascription studied across folk psychology, relatively little empirical research has been done examining how adults ordinarily reason about the *knowledge* of others.[1] Furthermore, and as a result of this perhaps, the ordinary concept of knowledge, together with the underlying processes and mechanisms that cue and govern ordinary ascription, are currently not well understood. The main idea of this essay is relatively straightforward. It is that this situation in epistemology must be remedied, and that new research aimed at closing the gap in our lack of understanding of ordinary knowledge behaviors and processes is needed.

The principle thesis of this dissertation is that understanding the psychological factors that underlie the judgments we ordinarily make about knowledge ascription is essential for progress in traditional epistemology, and that the tools developed across the cognitive sciences are necessary for collecting accurate evidence concerning the nature of these factors. The chapters of this dissertation are presented as cumulative—*proof of concept*—for this thesis. Within each chapter you will find original experimental evidence advancing new understanding of several factors that underlie the judgments we ordinarily make about knowledge ascription. In fact, the data will demonstrate how many of the factors discussed have been significantly misunderstood or misrepresented by philosophers prior to the present empirical scrutiny. You will then find a discussion of how these factors, properly understood using controlled scientific

[1] For important exceptions in the developmental psychology literature see Birch 2005, Birch and Boom 2007, and Woolley & Wellman 1993.

methods, come to bear directly on a series of significant and ongoing debates in contemporary epistemology. By the end, the main goals of this essay are to have convinced you of the crucial importance understanding the psychological factors underlying our epistemic judgments have for epistemology today, and that the use of experimental methods in philosophy is the best way to help achieve this new understanding.

But this method has something for everyone, not just those philosophers who have participated in what some may judge to be esoteric metaphysical debates only within contemporary epistemology. For by using this method, we will be engaging in a straightforward empirical investigation of knowledge behavior and underlying processes. And in turn, these data may be used to inform theorizing about the semantics of the word 'know' (of interest to linguists generally), as well as theorizing about the ordinary concept of knowledge that people tacitly hold (of interest to psychologists generally). Additionally, progress on these empirical questions may even begin to inform normative speculation about the value of that ordinary concept as it is ordinarily used. And since the way we use the word 'know' or apply our concept of knowledge when we engage in knowledge ascribing behavior seems to be at the very heart of the way we, for instance, often form decisions about assertion or correct action, engage in punishment practices, blame or praise others in social contexts, and so on in our daily lives, discovering the factors underlying our epistemic judgments holds intrinsic interest for those across several diverse debates and domains of inquiry.

Here is how the dissertation will proceed. I begin in Chapter 1, "The Return to Ordinary Language," with a brief review of the evidentiary role that everyday knowledge ascriptions, or uses of the word 'know', have typically begun to play in contemporary epistemology. Once this is in place, we will then proceed directly with an empirical investigation of the nature of these

behaviors and practices. Chapter 2, “The Factivity of Knowledge” starts by addressing the initial question whether or not the folk concept of knowledge is factive or non-factive. We will discover that despite appearances to the contrary, experimental data indicates that philosophical orthodoxy regarding the *factivity* of ‘knows’ in ordinary language withstands empirical scrutiny—and that the truth of a proposition is an important factor that cues ordinary knowledge ascription.

Given this basic starting point into the ordinary concept, we then proceed in Chapter 3, “Stakes, Error, and Accommodation: Contextualism and Pragmatic Encroachment” with a discussion of a cluster of phenomena we might label collectively as ‘pragmatic load’. This work attempts to isolate and characterize the *pragmatic factors* that affect whether or not an attributor will say that a subject has knowledge. The two main questions that are explored are whether or not intuitions about uses of ‘know’ are affected by the attributor’s situation (like the salience of possible attributor error or something called attributor accommodation), or the subject’s situation (like what is at stake for the person an attributor is evaluating). We find that understanding these factors plays a crucial role in evaluating the arguments given for two leading epistemic theories, contextualism and interest-relative invariantism.

In Chapter 4, “The Moral Component of Knowledge: Back to Gettier” we discover that in addition to pragmatic factors like those discussed in Chapter 3, there may also be a substantial *moral component* to knowledge attribution. The present set of studies demonstrate that this basic effect of morality on epistemic judgment is so strong that it can be extended to overturn intuitions in some of the most theoretically central experiments in contemporary epistemology: Gettier cases. It is argued that this phenomenon, in addition to marking an important factor that underlies ordinary ascriptions of knowledge, may also provide new insight into the Gettier problem.

Lastly, Chapter 5, “Methodology in Epistemology,” surveys the latest findings in experimental epistemology suggesting that attributions of knowledge may be highly variable to differences among persons due to *performance errors* like order and framing effects, and highly variable between persons due to *demographic differences* like culture, native language, and age. It is argued that further empirical investigation into the nature, scope, and limits of these two sorts of effects is necessary before epistemologists can continue to rely on empirically unexamined intuitions about ordinary usage, or count their intuitions about such things as good evidence for philosophical analyses of knowledge. The final chapter, Chapter 6, “Conclusions and Future Study” is a brief conclusion suggesting new avenues for future research in connection with ongoing projects in both psychology and cognitive science.

Chapter 1. The Return to Ordinary Language

> The systematic development of truth in scientific form can alone be the true shape in which truth exists. To help to bring philosophy nearer to the form of science—that goal where it can lay aside the name of *love* of knowledge and be actual knowledge—that is what I have set before me.
>
> - Hegel

In the physics classroom, one is reacquainted with a number of terms that we regularly use in everyday discourse. Yet part of any introduction to physics is to teach students that familiar words such as ‘force’ or ‘work’ have very different definitions in the present context than the ways typically used in ordinary language. Such differences are often quite substantive. If the physicists are right about what it means to do work for instance, then it’s doubtful that professional philosophers have done a day of it. Nevertheless, learning that force equals mass times acceleration, or work the total amount of energy transferred, inspires no great controversy among philosophers. Theories of physical laws are not written which take these differences into account. And few physicists feel the need to address the tension between the ordinary conception of ‘work’ and the notion employed across the natural sciences. The reason, it seems, is that the latter is simply accepted as a technical notion within the field of physics.

In contrast, what one does in the epistemology classroom feels very different. While an introduction to this subject also includes the sophisticated analysis of familiar words from everyday language, theories of knowledge or justification that radically diverge from ordinary intuitions, or acceptable uses of ‘knows’ or ‘evidence’ in English inspire tremendous controversy. Entire theories of knowledge are often weighed by their ability to account for their own intuitiveness to some degree or other—and may suffer considerable costs when they fail to

do so. But why, in the relevant respect, is the study of knowledge seemingly so different from the study of work or force? Perhaps the need in epistemology to account for commonsense intuitions suggests that 'knowledge' is not just a technical notion within the field. Instead, the presumption is that the analysis of knowledge that professional philosophers give should be expected to capture some important aspect or aspects of the concept that people came into the classroom tacitly holding.

Of course, the basic idea that ordinary judgments in philosophy are important is nothing new. Even as Aristotle writes in *Nichomachean Ethics* when discussing arguments concerning the good and function of humans:

> We must consider it, however, in the light not only of our conclusion and our premises, but also of **what is commonly said** about it; for with a true view all the data harmonize, but with a false one the facts soon clash.

And later in the same passage, when discussing several intuitive views often held with respect to happiness:

> Now some of these views have been held by **many men and men of old**, others by a few **eminent persons**; and it is not probable that either of these should be entirely mistaken, but rather that they should be right in at least some one respect or even in most respects. (1098b, 27-28)

As Aristotle later notes in this passage, there are some conditions or features of a position, both "old" and "agreed on by philosophers" that a philosophical theory should not fail to meet. However, and at the same time, one would also be foolish not to countenance, "what is ordinarily said" about the content of that theory when arriving at the truth.

It seems clear—and especially at least, compared to ordinary and professional judgments in physics today— that ordinary judgments, and ordinary language evidence in recent epistemology, continues to play a similar kind of important role in philosophical theorizing about knowledge (prominent examples of this range, of course, from Austin 1956; Wittgenstein 2009). But to illustrate this phenomenon in practice, consider one prominent recent example of when the two—ordinary intuition and professional theorizing—come into conflict. A particularly good example can be seen in the philosophy of David Lewis, in reacting to the threat of infallibilist knowledge. It is often thought that, according to commonsense, we all command a fairly large body of everyday knowledge. This knowledge ranges from mundane Moorean facts about our place in the external world, to complex attributions concerning the existence of unobservable entities. On the other hand, nearly all of the thoughts that we have and the decisions that we make occur in conditions involving some amount of risk or uncertainty. If you start taking courses in contemporary epistemology, you will soon discover that this fact has the potential to threaten many of the claims we ordinarily consider ourselves to know. Under the infallibilist picture of knowledge, to know something requires certainty, or that we eliminate all possible chances for error. We can formulate a definition for this classic infallibilist position as follows:

> *Classic infallibilism*. Subject S knows that proposition p iff p holds in every possibility left uneliminated by S's evidence.

According to classic infallibilism, one must rule out every single possibility where one could be mistaken in order to know. Alarmingly however, one is hardly ever in such a position of completely avoiding every possibility of error. As David Lewis encourages, "let your paranoid fantasies rip—CIA plots, hallucinogens in the tap water, conspiracies to deceive, old Nick himself—and soon you find that uneliminated possibilities of error are everywhere" (1999, 418). Since there are very few things about which we can be absolutely certain, it follows that we might actually know next to nothing at all. The result is that when considering whether one knows p, one's evidence is almost never strong enough to rule out every possibility that not-p.

Many people are deeply uncomfortable with this picture of knowledge. If infallibilism is true, then this challenges whether even our most mundane beliefs really count as knowledge. In order to avoid this extremely skeptical conclusion, most philosophers have denied that knowledge requires certainty. It's a good thing too, since after philosophers leave the office, they join all the non-philosophers in using language, ascribing mental states, and making judgments in the absence of it. These facts lend support to the idea that making allowances for uncertainty regularly plays an important role not only in theoretical reasoning about epistemic concepts like knowing or believing, but also in the practical choices we ordinarily make about acting, feeling, desiring, forming relationships, or doing almost anything at all. As Lewis famously notes, accounting for ordinary intuitions seems to force the good epistemologist into a choice between "the rock of fallibilism and the whirlpool of skepticism" (1999, 419).

Notice what Lewis didn't say. Lewis did not dismiss everyday knowledge ascriptions as irrelevant to his epistemic theorizing in philosophy, in the way that perhaps ordinary notions of 'work' are thought by physicists irrelevant to the technical notions of 'work' employed in

science. Nor did Lewis reject the idea that the conditions under which we ordinarily take ourselves to know count as powerful evidence against infallibilism. Instead, what Lewis set out to do was build a theory of knowledge that somehow took account of these claims. His goal was to preserve the preferred philosophical position of infallibilism while resisting total skepticism—but to also do justice to the ordinary judgments we make regarding robust knowledge. To accomplish this, Lewis turns his attention to the semantics of knowledge attribution. What he suggests is that when we typically ascribe knowledge, we do not expect that one must rule out all possibilities of error as traditional infallibilism demands. Instead, the possibilities of error that must be eliminated by one's evidence in order to know are fixed by the context in which the relevant knowledge claim is uttered. Adding this condition to classic infallibilism above results in the following definition:

> *Lewisian knowledge*. S knows that p iff S's evidence eliminates every possibility in which not-p—*Psst!—except for those possibilities that we are properly ignoring.*

Under this proposal, "Psst!" indicates the unarticulated restriction of the quantifier 'every' such that the context of the attribution distinguishes epistemically relevant from irrelevant possibilities of error. Then, when deciding what a subject knows, only the relevant possibilities need to be ruled out by a subject's evidence in order for an attributor to ascribe knowledge. In theory, this basic strategy allows Lewis to maintain infallibilist knowledge, while simultaneously providing for the ordinary intuitions that skeptical challenges are thought to contest. All that remains for Lewis is to specify the particular conditions according to which speakers ignore uneliminated

possibilities of error. And this brings the epistemology back in line with ordinary knowledge ascription. Of course, the question remains whether or not Lewis' theory was actually successful in its attempts to reconcile the two, or in providing a suitable epistemic theory more generally. But it is important to recognize what I'll call *the core insight*—and the new type of approach to epistemology it helped inspire.

Prior to this, the major traditional philosophical approach in epistemology involves analyzing the *constituents* of propositional knowledge. These analyses include specifying the individually necessary and jointly sufficient conditions that must be met for knowledge (see for instance Chisholm 1977; or Shope 1983). And in proposing and testing these conditions, philosophers proceed by appealing to intuitions. In a typical episode, they describe an imaginary situation and ask whether the subject in that situation exhibits some epistemically interesting property or relation. When things go well, the philosopher takes the answer to be intuitively obvious, and subsequently, as evidence for or against the proposed conditions for knowledge. Analyses that are compatible with the content of the intuition are supported, and those that are incompatible with the content of the intuition are challenged. This method has been, and continues to be, of central importance in epistemology. For example, the intuition that a character in a Gettier thought experiment does not know that p (where p is a true proposition that he believes and is justified in believing), is often taken as unanimous evidence against the justified-true-belief account of knowledge.

Alternatively however, the core insight, illustrated in the example from Lewis, is that new progress in epistemology could be made by analyzing the semantics of 'know' and factors that underlie ordinary ascriptions of knowledge. By analyzing these factors, Lewis found that we might discover new features of the concept typically neglected by the traditional approach

described above. For instance, if it's the case that knowledge attribution is sensitive to eliminating only relevant error possibilities from those that may be properly ignored, this may provide evidence for certain contextualist theories of knowledge. Such theories may, in turn, provide new responses to perennial challenges in epistemology that had not previously been considered. And that is exactly what happened. The resulting contextualist theory of knowledge, and the developments that would soon follow, have been widely extolled for their resistance to skepticism (DeRose 1995; Cohen 2000) and Gettier problems (Cohen 1999).

Today, this key insight seems to have spread widely. Popular linguistic theories like standard epistemic contextualism (DeRose 2009; Cohen 2004) and theories of pragmatic encroachment like interest-relative invariantism (Fantl & McGrath 2010; Hawthorne 2004; Hawthorne & Stanley 2008; Stanley 2005) mark the return of ordinary language approaches to epistemology. Indeed, many leading epistemologists have begun to turn their attentions away from the sort of epistemology embodied in the wake of post-Gettier theorizing. In its place, epistemologists have begun to argue for theories of knowledge—not by enumerating necessary and sufficient conditions of knowledge—but rather, by appealing to the factors that influence the judgments people ordinarily make about knowledge attribution, the way we actually use the word 'knows' in non-philosophical talk, or the ordinary concept of knowledge that people tacitly hold.

For example, consider the two main factors discussed in connection with contextualism and theories of pragmatic encroachment said to make a difference when people ascribe knowledge: the practical interests of an ascribee (like what is at stake for the subject of a knowledge claim), and the salience of attributor error (or whether or not the possibility of mistake has been made salient to an attributor of a knowledge claim). These factors, and their

philosophical role in debates between epistemic contextualism and interest-relative invariantism will be presented in detail in Chapter 3. But for the moment, we might use stakes and salience as examples to help begin to illustrate the importance of ordinary knowledge behaviors in epistemic theorizing.

Following Brown (2011), we can distinguish three non-mutually exclusive positions (behavioral, semantic, and metaphysical) various epistemologists have held regarding subject stakes:

- *Folk stakes sensitivity* (behavioral): People are less likely to ascribe knowledge to a high stakes subject for whom the practical consequences of error are severe, than to a low stakes subject for whom the practical consequences of error are slight.
- *Pragmatism* (semantic): The truth of a knowledge ascription depends in part on what is at stake for the subject.
- *Interest relativity* (metaphysical): The holding of the knowledge relation depends in part on what is at stake for the subject.

And regarding the epistemic role of mistakes salient to the ascriber:

- *Folk salience sensitivity* (behavioral): People are less likely to ascribe knowledge in a high salience context in which uneliminated possibilities of a mistake are relevant, than in a low salience context in which uneliminated possibilities of a mistake are irrelevant.
- *Contextualism* (semantic): The truth of a knowledge ascription depends in part on which mistakes are salient for the ascriber.

- *Contrastivism* (metaphysical): The holding of the knowledge relation depends in part on the value of a contrast argument for the mistakes in play.

While these positions regarding stakes or regarding salience are all logically independent, they are often presented together. For instance, we can read Fantl & McGrath (2002), Hawthorne (2004), Stanley (2005), and Pinillos (2011, 2012) as holding all three views of stakes, and Schaffer (2008) holding all three views of salience. But perhaps most often, *Folk stakes sensitivity* and *Folk salience sensitivity* are presented as the lead premises in inference to the best explanation style arguments for radical epistemic conclusions, such as, the semantic conclusion for the role of stakes in *Pragmatism*:

1. *Folk stakes sensitivity*: People are less likely to ascribe knowledge to a high stakes subject for whom the practical consequences of error are severe, than to a low stakes subject for whom the practical consequences of error are slight.
2. If *Folk stakes sensitivity* is true, then the truth of a knowledge ascription depends in part on what is at stake for the subject (inference to the best explanation).
3. *Pragmatism*: The truth of a knowledge ascription depends in part on what is at stake for the subject.

And for the semantic conclusion for the role of salience in *Contextualism:*

1. *Folk salience sensitivity*: People are less likely to ascribe knowledge in a high salience context in which uneliminated possibilities of a mistake are relevant, than in a low salience context in which uneliminated possibilities of a mistake are irrelevant.

2. If *Folk salience sensitivity* is true, then the truth of a knowledge ascription depends in part on what mistakes are made salient for the ascriber (inference to the best explanation).
3. *Contextualism* (semantic): The truth of a knowledge ascription depends in part on which mistakes are salient for the ascriber.

Of course, these arguments can be extending using the same pattern of inference to reach the metaphysical conclusion for stakes in *Interest relativity:*

4. If *Contextualism* is true, then knowledge depends in part on what is at stake for the subject (inference to the best explanation).
5. *Interest relativity*: The holding of the knowledge relation depends in part on what is at stake for the subject.

And the metaphysical conclusion for salience in *Contrastivism*:

4. If *Pragmatism* is true, then knowledge depends in part on which mistakes are salient for the ascriber (inference to the best explanation).
5. *Contrastivism* (metaphysical): The holding of the knowledge relation depends in part on the value of a contrast argument for the mistakes in play.

One can read DeRose (2005) as presenting precisely this type of argument for *Contextualism*. We can read Schaffer (2008) as presenting this extended argument to *Contrastivism,* and Stanley (2005) for *Interest relativity*. But it is important to note that *Folk stakes sensitivity* and *Folk*

salience sensitivity—the premises taken to provide abductive support for *Contextualism*, *Contrastivism,* and *Interest relativity*—are each at core empirical claims. The result is that many leading views in epistemology today rely on substantial and explicit use of empirical or descriptive claims.

Indeed, Stanley motivates his defense of *Interest relativity* largely based on "the intuitive reactions we have," to pairs of cases that manipulate a subject's practical interests. His claim is that our intuitions regarding knowledge ascription are stakes sensitive, and that his "central interest is to evaluate accounts that make as much sense of these intuitions as possible" (2005). Likewise, Hawthorne touts *Interest relativity* as offering "the best hope yet for respecting the intuitive links between knowledge, assertion, and practical reasoning" (2004). DeRose goes as far as to say that "the best evidence" for *Contextualism* comes from "what ordinary speakers will count as 'knowledge'" (2005). And Schaffer employs a similar style of argument for *Contrastivism*: "The most direct explanation of why knowledge ascriptions are contrast-sensitive is that the knowledge relation has a contrast slot" (2008).

These evidential strategies mark a major new trend in contemporary epistemology. But while the *approach* in epistemology has begun to change to include these kinds of factors as premises in epistemic arguments, the *methods* epistemologists have used remain largely the same. The way that philosophers have gathered evidence for these empirical claims about ordinary usage, everyday knowledge judgments, or the ordinary concept itself still largely involves the familiar appeal to intuition about knowledge attribution, together with armchair claims about "what people will say" in response to philosophical thought experiments. In my view, this is a *massive oversight.* And it is one as highly prone to error as it is easily enough to fix.

For much is at stake. If for instance, *Folk stakes sensitivity* is false (as I will argue in Chapter 3) then the argument to *Pragmatism*, or the extended argument to *Interest relativity* given above never get off the ground. And moving beyond simply evaluating *Folk stakes sensitivity*, what other folk sensitivities out there are there, what is the epistemic import of these factors on knowledge attribution, and could they also contribute in valuable and important ways to our study of knowledge? We currently do not know. But one way that we are *not* going to find out the answers to these questions is by appealing to armchair claims about "what people will say" in response to philosophical thought experiments. Conversely, the overall aim in this dissertation is to show that the sole method of using intuitions described above is no longer a solely appropriate method to use when evaluating these latter kinds of arguments in epistemology. Rather, I intend to demonstrate that new philosophical progress can be made in epistemology by preceding with the study of knowledge in just the same way that philosophers and psychologists have studied belief or desire ascription. That is, by using the tools developed in cognitive science and experimental philosophy to empirically examine ordinary knowledge behaviors, and specifically, the conditions and underlying processes by which knowledge is ascribed, we can gain proper insight into the ordinary concept of knowledge. And it turn, these insights, as Lewis and others have shown us, can lead to important new revelations in epistemology. As I will argue, they can demonstrate the factivity of 'knowledge' and hence, provide evidence for the factivity of knowledge (Chapter 2). They can adjudicate arguments between contextualism and interest relative invariantism that rely on the straightforward empirical premises in the manner above (Chapter 3). They can provide new responses to perennial Gettier problems (Chapter 4). And sometimes, they can even shed new light on the

evidentiary status of certain epistemic intuitions, adding to ongoing arguments in philosophical methodology (Chapter 5). This is what I attempt to show in the pages that follow.

But as I've suggested in the introduction, the progress this method may lead to is not only limited to progress regarding certain debates only among epistemologists. Of course, since this is a dissertation in epistemology, these debates will serve as the main focus by which the relevant findings of each chapter are presented. But new understanding of the factors underlying epistemic judgments promises almost as much interest to philosophers and psychologists outside of debates in epistemology as those engaged in them. For instance as displayed above, a behavioral investigation of folk salience or stakes sensitivity can lead to evidence that informs our understanding of the semantics of 'know'—or when people think that a sentence involving a knowledge claim is true. And whether or not the semantics of 'know' includes sensitivity to stakes or salience is not only of interest to linguists generally, but to all those philosophers who wish to understand how practical interests or relevant alternatives factor into people's judgments about the truth values of knowledge ascribing sentences, regardless of how this impacts arguments in epistemology about the holding of the knowledge relation.

Additionally, a third upshot of applying an empirical method to questions of knowledge ascription is that we can use information about the semantics of 'know' to gain clearer insight into the concept of knowledge that people tacitly hold. Since the concept of knowledge seems closely linked to other important concepts such as belief or those that inform action and decision-making, the inner workings of the concept will have extensive domain general appeal. For instance, and as it will be argued in the next chapter, it can often appear that people ascribe knowledge of a proposition to others in a way that is insensitive to the truth of that proposition. But after careful empirical investigation of that linguistic behavior, we will discover that the best

explanation of this is that the concept of knowledge that people hold is factive after all. This is great news for traditional epistemology—which has long emphasized the obviousness of factivity. But the discovery that factivity is an important part of the ordinary concept of knowledge may have a series of important implications about how we use this concept in our daily lives, despite the role it plays in professional arguments for or against factivity in the holding of the knowledge relation.

Lastly, by using this method to grasp a deeper understanding of the concept of knowledge that people tacitly hold, the fourth upshot is that we will also at least begin to be in a position to evaluate the normative value of that concept in the wild. Is that concept a useful one that *should* be relied on when acting or forming other sorts of mental states? Or instead, does the factors we will uncover suggest that it is somehow detrimental to rational decision-making and should be revised? Through empirically investigating the factors that cue ascription, and the ordinary concept that informs these behaviors, we can also begin to form normative predictions about the usefulness, reliability and rationality of that concept or practice. Given these four potential goals of applying this method, we now turn to the research—an empirical investigation into knowledge ascription, and the factors at work underlying our epistemic judgments.

Chapter 2. The Factivity of Knowledge

From the outset, some philosophers may be quite skeptical that the study of ordinary language behaviors regarding 'knows' can actually be of very much use to philosophers. The reason is that even a cursory glance at folk usage of epistemic terms appears to indicate that the concept of knowledge that non-philosophers hold may in fact be very different than the concept of knowledge that professional philosophers are interested in analyzing. Specifically, these two groups appear to differ in their positions on something as fundamental to an analysis as the factivity of knowledge.

Of course, nearly all philosophers agree that only true things can be known. As Socrates asks, "If one fails to get at the truth of a thing, will he ever be a person who knows that thing?" The resulting dialogue in Plato's *Theatetus* reveals his interlocutor's answer. Knowledge cannot be mere opinion, because there may be a false opinion (186c-187b). So since one of the first recorded attempts in western philosophy to analyze what it means to know, philosophers have unanimously accepted the standard truth condition for propositional knowledge. A subject can know a proposition only if that proposition is true. And unlike virtually all other questions debated in connection with the requirements for knowledge today, "the truth condition has not generated any significant degree of discussion. It is overwhelmingly clear that what is false cannot be known" (Steup 2006).[2] One particular way in which philosophers might choose to justify the truth condition is by appealing to the linguistic thesis that the word 'knows' is factive. The utterance 'S knows that p' is true only if p is true. And most epistemologists consider this linguistic thesis to be obvious from the armchair.

[2] Steup does consider one objection to the truth condition proposed by Nicholas Maxwell that Newtonian physics, though false, constitutes part of our overall scientific knowledge.

But does this principle actually reflect ordinary usage, or have things changed considerably since Plato? There seem to be several examples in ordinary language today showing that non-philosophers do not find this linguistic thesis obvious, and that they frequently use 'know' in what appear to be blatantly non-factive ways. This fact has led a growing number of philosophers to begin to speculate about the role of factivity in the actual knowledge judgments people make, as well as the significance these ordinary judgments might have for traditional epistemic theorizing. John Turri for instance, has advanced a performance-view of knowledge that allows for knowledge of "approximate truths", which are strictly speaking, false beliefs (2011a; *forthcoming*). Sympathy for non-factivity has been expressed in research by Daniel Nolan (2008). And, the possibility that epistemic contextualism might allow for contexts under which certain kinds of false beliefs qualify as knowledge has also been discussed in the works of Keith DeRose (2009).

Perhaps the most comprehensive challenge against orthodoxy to date is presented by Allan Hazlett. In a pair of papers, "The Myth of Factive Verbs," and "Factive Presupposition and the Truth Condition on Knowledge" Hazlett presents evidence that the way people use purportedly factive verbs like 'knows', 'learns', 'remembers', and 'realizes' in utterances of the form '*S* knows that *p*' frequently do not require that *p* is true (2010, 2012). He argues that the best explanation for these linguistic data—and particularly for uses of 'know'—is that the underlying folk concept of knowledge allows that false things can be known. The result, it seems, is that the standard truth condition for knowledge may not be justified by ordinary language, but only at the cost of it. (Indeed Hazlett argues that epistemologists should give up ordinary ascriptions as evidence, not just for the truth condition, but also for theorizing about the philosophical concept of knowledge altogether on the basis of this claim).

Given that a number of epistemologists have begun to reconsider the truth condition in light of intuitions about ordinary usage, we may wonder, could it really be that the folk concept of knowledge is truly a non-factive concept? Before attempting an answer, notice that this question, as well as questions regarding the ordinary usage of 'know,' are straightforwardly empirical. Thus it is unsurprisingly that in recent years, experimental philosophers have focused their attentions on investigating these claims. And generally speaking, the result of such research has revealed several experimental challenges to the orthodox conception of knowledge, as regards the necessity of belief (Myers-Shulz & Schwitzgebel *forthcoming*), the Gettier condition (Starmans & Friedman 2012), and the presence of contrast classes (Schaffer & Knobe 2010). So this seems like a perfectly natural place to begin an empirical study of the factors that influence and underlie ordinary attribution. After all, if it's true that non-philosophers hold a non-factive concept of knowledge, some philosophers have questioned whether or not one ought to count folk intuitions as very good evidence in discussions about *their* philosophical concept. And more generally, we might also be interested in this from a purely scientific perspective. If people are typically insensitive to the truth of propositions when ascribing knowledge because their concept is non-factive, this may have a series of shocking implications for our use of that concept, as compared to say, just belief, when it comes to things like justifying action or ascribing praise or blame.

So might research on the truth condition threaten traditional philosophical wisdom about the factivity of knowledge in this way? The goal of this chapter is to *defend* the linguistic thesis (the orthodox philosophical view, as well as ordinary language approaches in epistemology more generally) from this new threat, by attempting to *confirm armchair intuitions* about the factivity of 'know'. Using prior proposals in the literature as a starting point, four experiments are

conducted demonstrating that while it is true that people often use verbs like ‘know’ in what appear to be non-factive ways, the best explanation of these phenomena is not that the folk concept allows for knowledge under false belief. Conversely, new experimental evidence suggests that the concept of knowledge that people tacitly hold does support the standard truth condition for knowledge, together with the fact that the ordinary uses of ‘know’ are often highly sensitive to something called ‘protagonist projection’ (Holton 1997).

In what follows section (2.1) reviews the linguistic evidence against factivity presented by Hazlett, and introduces protagonist projection as a rival explanatory hypothesis to these data. Section (2.2) argues that the purportedly non-factive uses of ‘know’ that we will consider are better explained by the folk tendency to adopt the perspective of the putative ‘knower’ when attributing, and not because they think that subjects with false beliefs really do have ‘knowledge’. Sections (2.3-2.4) are a discussion of these results, and the implications they may have for ordinary language approaches in epistemology.

2.1. The Argument Against Factivity

Nearly all epistemologists agree that only true things can be known. Call this the standard truth condition of knowledge:

(STC) *S* knows that *p* only if *p* is true.

Many philosophers have justified STC in their analyses by appeal to the linguistic thesis that the word ‘know’, much like the English words ‘learn’, ‘remember’ and ‘realize’ is factive. Let this be defined as follows:

> (Factivity) Certain two-place predicates, including 'knows', 'learns', 'remembers', and 'realizes', which denote relations between persons and propositions, are *factive* in this sense: an utterance of 'S knows p' is true only if *p*, an utterance of '*S* learned *p*' is true only if *p*, and so on. (From Hazlett, 2010, 499)

Our question is whether or not Factivity is true. To test whether or not Facticity is true our task will be to investigate whether ordinary uses of these words—and in particular the word 'know' in utterances of the form '*S* knows that *p*'—truly reflect this principle. From the start, the prognosis for this principle seems grim. A quick Google search (headline: "The Cuban missile crisis: what you know is wrong") reveals that one is surrounded by what appear to be blatantly non-factive uses in ordinary language.[3] Consider four more examples given in the literature we might assume ordinary speakers will find acceptable. If true, the following data—taken from actual press releases, movies, and newspapers—would count as evidence against Factivity:[4]

> (Know Ulcers) Everyone knew that stress caused ulcers, before two Australian doctors in the early 80s proved that ulcers are actually caused by bacterial infection.

> (Know Sailing) He figures anything big enough to sink the ship they're going to see in time to turn. But the ship's too big, with too small a rudder…it can't corner worth shit. Everything he knows is wrong.

[3] See this opinion piece at the New York Times website: <http://www.nytimes.com/2008/06/13/opinion/13iht-edbeam.1.13693498.html>

[4] These items are originally presented in Hazlett (2010), and remain the best linguistic evidence to date for non-factivity.

> (Learn History) In school we learned that World War I was a war to "make the world safe for democracy," when it was really a war to make the world safe for the Western imperial powers.
>
> (Realize Death) I had trouble breathing, sharp pains in my side, several broken ribs and a partially collapsed lung, and I was in the middle of nowhere without any real rescue assets – it was then that I realized I was going to die out there. (From Hazlett, 2010, 501, labels added)

But of course, these items only count as *prima facie* evidence against Factivity, as most epistemologists are aware that there are some contexts in which we appear willing to attribute knowledge of falsehoods. The question is what to make of such claims. In other words, we need to know whether or not these uses of 'know', 'learn', 'remember', and 'realize' really do indicate that people think 'S knows p' but not-*p* is true as philosophers understand them, or if these sentences can be accounted for in some other way.

Some epistemologists have begun to advocate for the former reading (see Turri *forthcoming*, Hazlett 2010; 2012). Again using Hazlett (2010) as an example, one option is to reject Factivity outright, and to argue that the folk concept serving as the meaning for 'knows' in ordinary language allows for knowledge under blatantly false belief. In place of orthodoxy, Hazlett eliminates the truth condition and argues for a concept of knowledge based on the following two necessary conditions:

(NF1) An utterance of '*S* knows that *p*' is true only if *S* believes *p*.

(NF2) An utterance of '*S* knows that *p*' is true only if *S* possess epistemic warrant for (her belief that) *p*.

It is an open empirical question as to whether or not this kind of non-factive concept is better equipped than the traditional philosophical concept of knowledge to provide for the variety of different ways that 'know' is used in ordinary talk. For our present purposes, the main selling point of a concept based on something like NF1 and NF2 is that it is better able to explain the *prima facie* counterexamples to Factivity considered above. For example, the reason why 'everyone knew that stress causes ulcers' is an acceptable sentence is because—given a belief backed by sufficient epistemic warrant—people really did have knowledge that stress causes peptic ulcers before it was proven to be *H. pylori* (Marshall et al. 1985).[5]

If something like a justified belief theory is correct, then philosophers could no longer look to ordinary language to justify STC. Instead, philosophers would need to defend STC in their analyses of knowledge in some other way. And furthermore, if the folk concept of knowledge informing the definition of 'know' in ordinary language really is non-factive, then this might also question whether or not epistemologists interested in giving a traditional analysis of knowledge should be all that interested in ordinary uses of 'know' to begin with. After all, the concept of knowledge that philosophers have historically held is factive through and through. So

[5] For a fascinating account of this discovery (in which Marshall describes infecting himself to prove his theories correct against overwhelming medical consensus) see the Noble Prize speech available at <http://www.nobelprize.org/mediaplayer/index.php?id=614>. Ironically, Marshall begins with a quote from historian Daniel Boorstein, "The greatest obstacle to knowledge is not ignorance, it is the illusion of knowledge."

data concerning the way people ordinarily use the word 'know' may not be the best evidence for those epistemologists whose specific goal it is to account for what has traditionally been considered the *philosophical concept* of knowledge.[6]

Given this potential threat to philosophical orthodoxy, and the significant methodological implications at stake for epistemologists, a series of objections have been given in this literature in response to the suggestion that the ordinary concept of knowledge is non-factive. For instance, Turri (2011b) and Tsohatzidis (2012) both argue along Gricean lines that even if a non-factive concept of knowledge presented by Hazlett above has the upper hand in the sentences considered above, something closer to the traditional factive account is still preferable to handle the *majority* of knowledge sentences we use.[7] And in response to these criticisms, an impressive amount of scholarship has been devoted to providing for what the technical inner-workings of a non-factive concept, should it exist, might look like (see Hazlett, 2012). However the question we consider presently is about the *actual linguistic data* itself, in those individual cases taken to provide the strongest evidence for the general claim that Factivity is false. When considering such cases we ask of the larger question at issue, is it really true that the best explanation of those linguistic data is that non-philosophers have a non-factive concept of knowledge?

Proposed instead is an account of the evidence collected against Factivity to date that is far less costly. This explanation involves the phenomenon of *protagonist projection* proposed by Richard Holton:

---

[6] It is important to note that the existence of different concepts of knowledge between philosophers and non-philosophers is not to be taken lightly, and may have a series of important didactic and methodological implications (see arguments by Buckwalter & Stich 2011, Stich *in prep*).

[7] These arguments allege that although Hazlett's non-factive theory relies on conversational implicature to account for occasions under which '*S* knows that *p*' does imply *p*, these implicatures fail Grice's "cancelation test of conversational implicatures" (1989). See Hazlett (2012) for a response to these arguments.

> I suggest that these sentences work by projecting us into the point of view of the protagonist; let us call the phenomenon protagonist projection. In each case the point of view into which we are projected involves a false belief. We describe the false belief using words that the protagonists might use themselves, words that embody their mistake. So we deliberately use words in ways that do not fit the case. (Holton, 1997, 626)

The standard story is that some sentences invoke *projective* readings. For instance, Holton considers the following examples, 'She sold him a pig in a bag. When he got home he discovered it was really a cat,' and 'I saw a shooting star last night. I wished on it, but it was just a satellite.' Sentences like these are acceptable to us because presumably, we take up the protagonist's perspective and then imagine what seems true from their point of view. And since the protagonist's perspective involves a false belief in each case, we use words the protagonists might use to describe those beliefs from their own points of view. So when we read sentences like the examples above, we don't take them literally as evidence that some pigs are actually cats, or that some satellites are actually meteoroids. Rather, we find the sentences acceptable because from the protagonist's point of view, 'he *thought* she sold him a pig,' or 'she *thought* she saw a shooting star.' Another way to think about this phenomenon is that it is roughly equivalent to when one talks about someone else, and—with sufficient cues, e.g. imitating their bodily language and tone of voice—can then use 'I' to refer to this other person.[8]

[8] Similarly, see DeRose (2009) for an argument that the special intonation of 'knows' provides crucial cues that sentences implying that protagonists *know* things that are false are not to be taken literally.

So might this also be happening when people evaluate the knowledge sentences that appear to count as evidence against Factivity? Consider for instance, Know Ulcers. It could be that 'everyone knew that stress caused ulcers' does not literally count as evidence that 'know' is non-factive, in just the same way that 'I saw a shooting star last night' does not literally provide evidence that some satellites are meteoroids. Instead, perhaps people think Know Ulcers is acceptable only because they assume the protagonist's perspective, in this case the perspective of people living before the 1980s.[9] Then, they describe how the situation may have appeared from that perspective. And, people living before the 1980s had relatively decent evidence supporting their beliefs about the relationship between stress and ulcers, in a way typically thought to be consistent with knowledge. Therefore 'everyone knew that stress caused ulcers' is judged acceptable because it appeared acceptable from *that* perspective—and not because it's judged true, from the more informed evaluator's perspective, that the protagonists actually knew something false.

We now have at least two hypotheses available to explain linguistic evidence like *Know Ulcers*—a non-factive folk concept of knowledge on the one hand, and Factivity plus protagonist projection on the other. And whether or not people are adopting projectionist readings in the relevant way is a straightforwardly empirical issue. So we can test if protagonist projection can explain people's willingness to accept seemingly non-factive uses of the purportedly factive verbs considered over and above postulating a non-factive concept. We simply need to collect more linguistic data. We now turn to four such experiments for this purpose.

[9] Turri (2011b) points out that this knowledge sentence might be dismissed as simple exaggeration (e.g. all 4.5 billion people alive in 1985 knew *what*, exactly?) and that it's crucial to take steps to promote strict literal readings. Though it seems plausible that correcting for this exaggeration would still leave the intuitive force of Hazlett's initial example intact (see follow-up experiment in footnote 19).

2.2. Experiments For Projection

The following four experiments have been devised to test if protagonist projection can account for people's judgments involving the seemingly non-factive use of 'knows', 'learns' and 'realizes' in ordinary language. And specifically for 'knows', our key question is whether or not people really do attribute knowledge to subjects with false beliefs, or if they are evaluating the acceptableness of each statement through the perspective of their protagonists.

2.2.1. Projection Experiment 1 (Hazlett Cases)

To begin, Experiment 1 simply sought to test whether protagonist projection could account for people's judgments concerning the specific examples presented in *Know Ulcers*, *Know Sailing*, *Learn History*, and *Realize Death.*[10] Participants (N = 73, 37 female, median age = 35) were randomly presented with items very similar to each of the four cases considered as counterevidence to Factivity in Chapter 2.1. These included two sentences containing the verb 'know' and one sentence each involving 'learn' and 'realize'.[11] For example, here is one of the four items given to each participant:

> "**Everyone knew** that stress caused ulcers, before two Australian doctors in the early 1980s proved that ulcers are actually caused by bacterial infection."

[10] All studies used the internet-based commercial research tools Mturk and Qualtrics. Online samples were restricted to participants located in the United States.
[11] A full list of materials is available for Experiment 1 and Experiment 2 available in Appendix and Supplemental Materials.

Directly after seeing each of the items, participants were given questions of the following form, respective to the different verbs imbedded. They were then asked to select one of the explicit paraphrases provided regarding the usage of that verb (order of items and answer choices also randomized in all experiments).[12] Below is the question that was pared with the ulcer case above:

> Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:
>
> A) Everyone thought they knew (*projectionist answer*)
>
> B) Everyone really did know (*non-factive answer*)

If participants are reading these sentences and interpreting the verbs as factive, then we would expect them to be more likely to adopt the projectionist answer—that from protagonist *x*'s perspective, '*x* thought that *x* Φ'. Alternatively, if the sentences considered by Hazlett really are rightly interpreted as counterexamples to Factivity, then we would expect participants to give the non-factive answer—that from the participant's more informed perspective, "*x* really did Φ."

What we find is a split in the way participants think about 'knows' on the one hand, and 'learns' and 'realizes' on the other. In Experiment 1 participants were significantly more likely to give the projectionist answer for the former, but non-factive answer for the latter two verbs.[13] The result was that only 9% (ulcer case) and 21% (sailing case) gave the non-factive answers to

[12] This explicit paraphrase task has also been used successfully in other domains, including the study of mental state attributions to groups (see Phelan 2010, Phelan et al. *manuscript*).

[13] Using McNemar's test, a significant tendency was found whereby participants were much more likely to give projectionist answers in both ulcer and sailing cases than in both history and death cases, respectively ($p < 0.01$ for all comparisons).

the knowledge sentences, while 56% (history case) and 49% (death case) gave non-factive answers to the learning and realizing sentences. This is reflected in Figure 2.1 below:

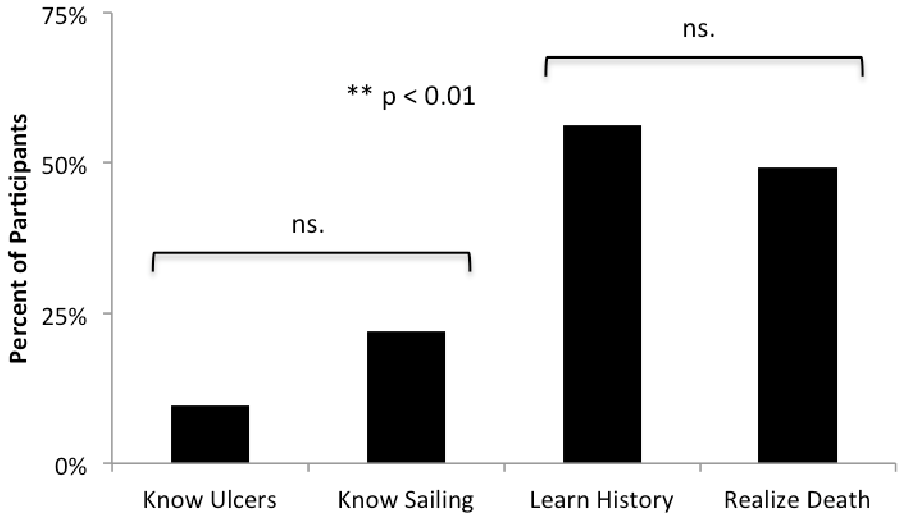

Figure 2.1. Projection Scores in Hazlett Cases. Depicted is the percent of participants giving "really Φ" answers in Experiment 1.

Interestingly, there seems to be some disagreement about the best interpretation of the sentences involving 'learns' and 'realizes'.[14] Perhaps this is a result of the melodramatic feel of the particular sentences selected. Or maybe it really is the case that a significant number of people consider Factivity false for 'learns' and 'realizes' specifically. By comparison however, the data do seem to clearly indicate that participants are interpreting the sentences involving 'knows' from the protagonists' perspective. Though we cannot be sure of the precise reason why, it does not seem correct to say that participants actually think that everyone before the 1980s really had knowledge under false belief when evaluating that particular sentence.

[14] John Turri (personal communication) suggests that perhaps one salient reading of 'learn that' is 'is taught that', and of course 'is taught that' certainly isn't factive.

2.2.2. Projection Experiment 2 (Crab cases)

Findings in Experiment 1 give us some reason to begin doubting whether the sentences considered in Section 1 really count against Facticity. But it is still a bit unclear what these data are telling us. After all, the cases differed not only in the verb used (e.g. knows vs. realized) but also in the various complement clauses expressing the false propositions. So Experiment 2 sought to replicate projectionist findings for 'knows' in Experiment 1, but this time by displaying the different factive verbs in the same contexts and with the same complement clauses. Also, the non-factive verb 'believe' was included in this study, serving as an experimental control for comparison to the factive verb findings.

In a between-subjects experiment, participants (N = 120, 70 female, median age = 34) were presented with a vignette about two biologists trying to indentify the genus of a crab.[15] While each participant was only given one of four possible stories, each story varied by which verb was embedded: 'know', 'realize', 'learn', or 'believe' (a full list of materials appear in the appendix). For example, below is the story involving knowledge:

[15] This test case is borrowed from Hazlett, an example using 'know' to guarantee the truth of a proposition, where the proposition is false.

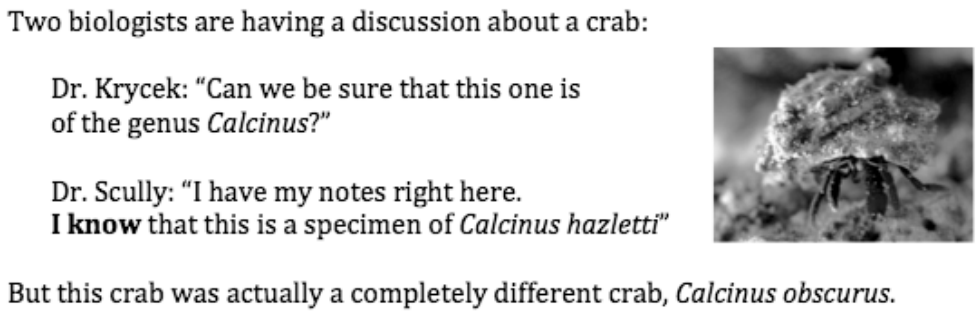

Figure 2.2. Example of experimental materials in Experiment 2. Depicted are materials for crab cases in the knowledge condition.

After each story, participants were asked the same basic question involving the relevant verb of the vignette they were given. Below is the question seen by participants given the knowledge case above:

Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:

A) Dr. Scully thought she knew

B) Dr. Scully really did know

As in Experiment 1, if participants think that Dr. Scully's utterance, 'I know that this is a specimen of *Calcinus hazletti'* actually describes a true state of affairs, then we would expect them to answer from *their perspective* that 'Dr. Scully really did know'. Alternatively, if participants interpret this statement by projecting, then the prediction would be that they will be more likely to answer that 'Dr. Scully thought she knew'. In other words, this answer would

count as good evidence that the reason participants find the incorrect thing Dr. Scully said acceptable is because it is true from the *protagonist's perspective* that she thought she knew, and not because she was actually able to know something false.

What was found was again a significant split in answers depending on the verb in question. Unsurprisingly, a large majority of participants gave the non-factive answer for 'believe' (63%). Additionally, only a small percentage gave this answer in the knowledge case as compared with 'believes' (10%, $p < 0.01$, Fisher's exact test). And puzzlingly, people were once again more likely to give non-factive answers for 'learn' and 'realize' (33% and 36%) than for 'knows' ($p < 0.05$, Fisher's exact test) though still significantly less than for 'believe' ($p < 0.05$, Fisher's exact test). These results are displayed in Figure 2.3:

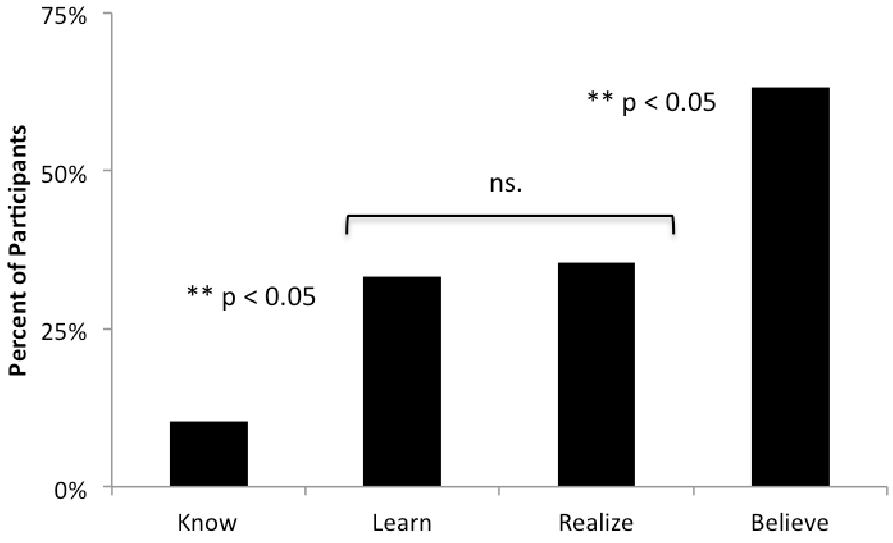

Figure 2.3. Projection Scores in Crab Cases. Depicted is the percent of participants giving "really Φ" answers in Experiment 2.

The results of Experiment 2 appear to replicate the findings in Experiment 1. This evidence again questions the explanation that participants think that protagonists with false beliefs really do have knowledge. Compared to the other verbs included across Experiments 1-2, a consensus is emerging for a projectionist reading of 'knows'. It seems unlikely then, that people's treatment of these particular knowledge sentences really are best interpreted as counterexamples to Factivity, insofar as they reflect an underlying non-factive folk concept of knowledge.

2.2.3. Projection Experiment 3 (FBI cases)

Experiment 2 continues to build our case in support of Factivity. But although it showed that people are more likely to give projectionist rather than non-factive readings for 'knows' than 'believes' in cases with identical false complement clauses, this does not give us evidence that the false complement clauses are directly responsible for their answers. It's possible that participants only tended to give projectionist answers to the sentences because they did not think the subjects had actually Φ-ed, independently of whether or not they think the verbs are or aren't factive. For example, participants may have answered that Dr. Scully only 'thought she learned' about the crab because she made an error when copying down her notes, or misremembered the lesson, and not because it's impossible to learn false things.

So Experiment 3 sought to find evidence that the truth or falsity of the proposition in question is what is directly responsible for projectionist or non-factive answers across the various different verbs tested. In a between-subjects design, participants (N = 217, 115 female, median age = 34) were presented with the following story about two police offers relaying some

testimony.[16] While each participant only saw one story, the stories varied only by the verb used in bold ('know', 'learned', 'realized', or 'believe'). Below is the vignette involving 'knows':

> Officer Ted asks the Police Sergeant, "Is there any information from the FBI about how the bomb was constructed?"
>
> The Sergeant told him, "From the investigation, **they know** the bomb was homemade."

After seeing one vignette, roughly half of participants were presented with a complement phrase expressing a false proposition.

> *False Complement:* But actually, the bomb in question only appeared homemade. Instead, professionals constructed it in a high-tech chemical plant.

The other half received a complement expressing a true proposition:

> *True Complement:* And as it turns out, the bomb in question was homemade. Non-professionals had constructed it in a basement apartment.

Lastly, participants were presented with the same question used in Experiments 1-3 to test for projectionist interpretations of the various different verbs:

[16] The FBI vignette is paraphrased from Hazlett (2010). Accounting for this case is taken to be a major advantage of the non-factive theory.

Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:

A) The FBI thought they knew

B) The FBI really did know

If the data in Experiments 1-2 really do support Factivity, then we would expect a large majority of participants presented with the factive verbs to answer that the 'FBI really did Φ' when the accompanying complement clause expresses a true proposition, and that 'FBI thought they knew' when given a false complement. Again, projection would predict that the incorrect testimony relayed by the sergeant when she said 'they know' is acceptable to participants only in so far as it is true from the *protagonists' perspective*, and not because false things are known. Alternatively, we would expect a null effect for 'believes', whereby the truth or falsity of the complement still results in a majority of non-factive answers.

And that was exactly what was found. Most participants gave the 'really Φ' answer for 'know' (86%), 'learn' (78%), 'realize' (90%), and 'believe'(80%) when presented with the true complements. But participants were significantly less likely to give the 'really Φ' answer for all factive verbs in false complement cases: 12%, 21%, and 31%, respectively ($p < 0.01$, Fisher's exact test for all comparisons). And as expected, the false complement case had no significant effect on judgments about 'believe' (61%). Lastly, the difference between projectionist answers in all the factive verb sentences compared to the 'believe' sentences—within just those cases

with false complements—continues to be significant ($p < 0.05$, Fisher's exact test for all comparisons).[17] These results are shown in Figure 2.4 below:

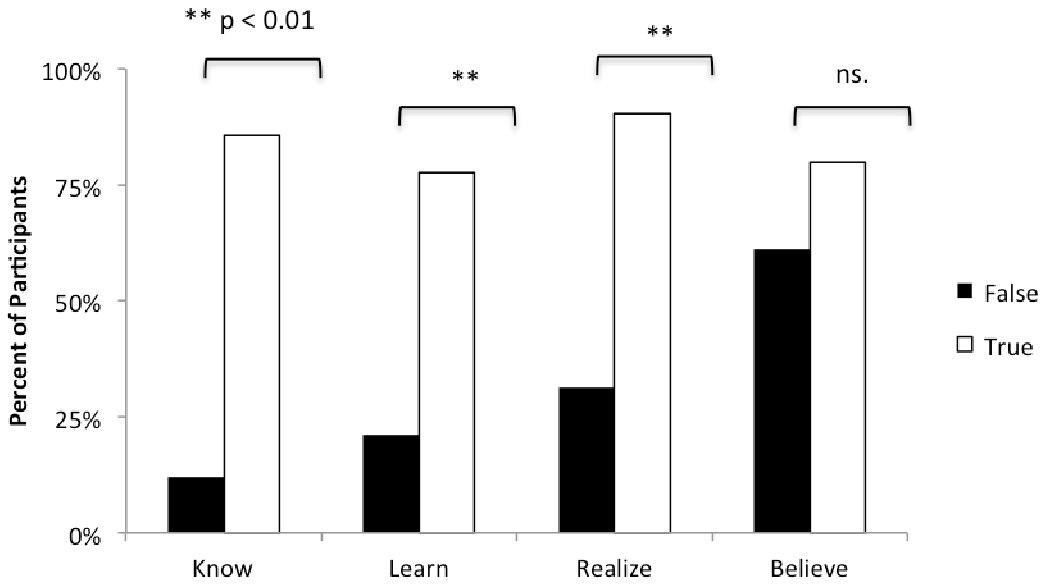

Figure 2.4. Projection Scores in FBI Cases. Depicted is the percent of participants giving "really Φ" answers grouped by true or false complement in Experiment 3.

2.2.4. Projection Experiment 4 (Bank cases)

The experiments so far have all asked participants about which interpretation best describes the uses of certain verbs in short sentences. But usually, epistemologists are interested in evaluating considered intuitions in detailed epistemic vignettes. They rarely ask participants to simply interpret individual sentences. So perhaps the findings in Experiments 1-3 are somehow the

[17] Unlike the biology cases, a significant effect was not detected in the FBI cases between 'knows' on the one hand, and the other purportedly factive verbs 'learn' or 'realize' on the other (though it does continue to be suggestive in the same direction).

result of presenting people with sentences involving non-factive usage, and do not reflect their more considered epistemic judgments about actual cases. So for stronger evidence that people are projecting when using the factive verbs above in what appear to be non-factive ways, we would want the same evidence for projection in vignettes like those typically used by philosophers. Additionally, we would also want to first, explicitly verify that participants are judging the false knowledge claim of the putative epistemic knower as acceptable, before explaining those acceptability judgments in terms of projection.

Experiment 4 sought to provide exactly this kind of evidence. Given the extensive research done in traditional and experimental epistemology on what are known as "bank cases" to date (DeRose 1992; Buckwalter 2010; Schaffer & Knobe 2010), these cases were also chosen to further investigate people's judgments about factivity. And since these cases have typically been used to study the salience of attributor error (or the possibility that a mistake might be made) they seem like particularly apt cases to investigate the effects that false beliefs (or that an actual mistake has been made) may have on ordinary knowledge judgments. The idea is to ask participants to make knowledge judgments about protagonists in these cases (rather than simply supply them with non-factive usage), and then test the actual judgments people make about these cases for projection.

In a between-subjects study, all participants in Experiment 4 (N = 126, 77 female, median age = 34) saw the following text, very similar to the original bank cases, and to previous research done with bank stimuli (Buckwalter 2010):

> Hannah and her sister Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. As they drive past

the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Since they do not have an impending bill coming due, and have plenty of money in their accounts, it is not important that they deposit their paychecks by Saturday. Hannah says, "I was just at this bank two weeks ago on a Saturday morning, I know that the bank will be open on Saturday. Let's leave and deposit our paychecks tomorrow morning."

After reading this basic story, participants randomly saw one of two different conclusions to the story varying whether the bank was actually open on Saturday was either true or false:

True belief: When they returned the next day, they found that the bank was open for business on Saturday, as it usually is most mornings.

False belief: When they returned the next day, they found that the bank was closed on Saturday due to a lightning strike that occurred earlier that morning.

Regardless of the various different bank vignettes received, each participant was then asked three questions. The first question was designed to serve as a basic comprehension and manipulation check to insure that participants read the story, as well as understood that the subjects above had explicitly true or false beliefs:

According to the story, was the bank actually open for business on Saturday morning? [Yes/No]

Then, on a five-item scale (1 being "definitely false," 3 being "neither true nor false," and 5 being "definitely true") participants were asked to judge the knowledge statement made by the subject in the story above:

> When Hannah said, "I know that the bank will be open on Saturday," is what she said true or false?

Lastly, participants were then asked the question designed to reveal whether or not protagonist projection was present:

> Which do you think best describes Hannah in the story?
>
> A. Hannah thought she knew
>
> B. Hannah really did know

The result of the study was that in accord with traditional philosophical wisdom, the truth of the epistemic subject's belief had a large effect on participant's knowledge judgments.[18] Participants were less likely to agree that the protagonist's knowledge claim was true when the protagonist's belief was false. These results are represented in figure below:

[18] 20 participants were removed for failing *comprehension check*. Means and standard deviations for the knowledge question were as follows: True Belief (M = 4.52, SD = 0.57), False Belief (M = 3.71, SD = 1.34). An independent samples t-test confirms this difference is significant, t (104) = 4.18, $p < 0.01$.

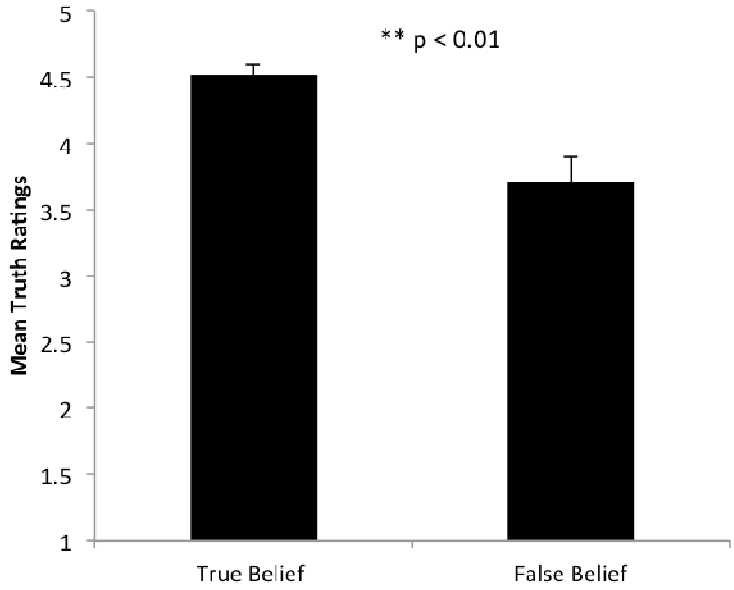

Figure 2.5. Knowledge ratings in Bank Cases. Depicted is the mean truth ratings of protagonist knowledge statement in Experiment 4. Scales ran 1-5.

While truth did have a large impact on participant judgments, answers in the false belief case were nonetheless still quite high. And in fact, mean answers are significantly above the midpoint of three on this five item scale, suggesting that participants do in fact consider Hannah's knowledge sentence 'I know that the bank will be open on Saturday,' to be true when her belief is false.[19]

Thus, we have found exactly what we were looking for, what appears to be a non-factive folk usage of a purportedly factive verb. The further question now remains as to whether or not this is evidence of an instance of non-factive knowledge or for projection. To find out we need to have a look at people's answers to the projection question. Again, for judgments that Hannah's knowledge sentence is true to count as good evidence against Factivity, we would expect them to

[19] A one-sample t-test shows False Belief means are significantly greater than the midpoint, $t(47) = 3.67$, $p < 0.01$.

be just as likely to answer 'Hannah really did know' despite the truth or falsity of her belief. Alternatively, if people are projecting, then we would predict significantly more "Hannah really did know" answers when Hannah's belief is true, and more "Hannah thought she knew" answers when Hannah's belief is false.

As it turns out, participants were much more likely to answer 'Hannah thought she knew' in the false belief case (60%), but that she 'really knows' in the true belief case (69%, $p < 0.01$, Fisher's exact test). These results are reflected in Figure 2.6 below:

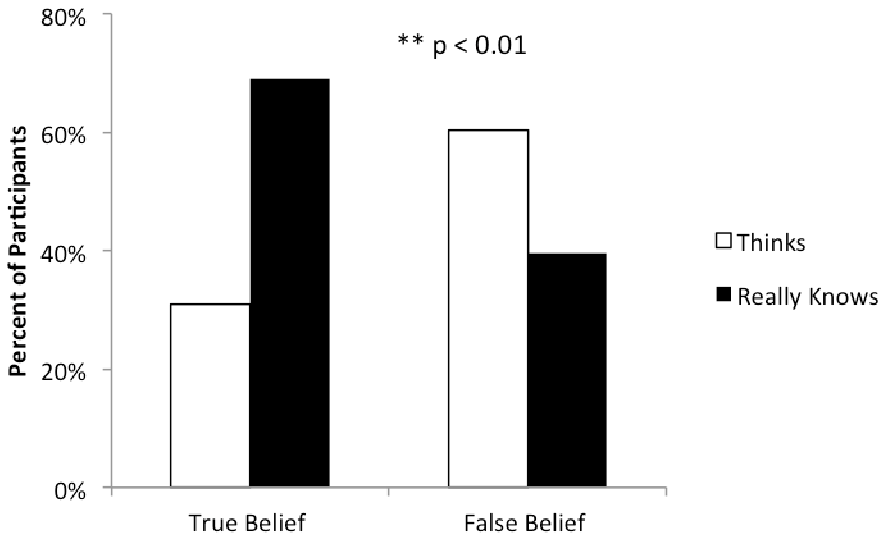

Figure 2.6. Projection Scores in Bank Cases. Depicted is the percent of participants answering 'Hannah thought she knew' versus 'Hannah really did know' in Experiment 4.

This demonstrates that there are cases in which participants do appear wiling to assent to the truth of '*S* knows that *p*' when p is false. But Experiment 4 also urges caution when assuming that this result counts as a counterexample to Factivity. As we have seen, participants are much

more likely to interpret such sentences from the perspective of the protagonist, than they seem willing to attribute knowledge to protagonists with false beliefs.

2.3. Do we know what's false?

We began with a discussion of the standard truth condition in epistemology, and noted that one of the ways that philosophers might justify STC is with the armchair assumption that Facticity is true. Our question was whether or not this assumption is warranted, or if a careful investigation of ordinary uses of 'know' actually presents evidence that Facticity is false. We then considered several sentences that seem to be acceptable or non-deviant to speakers, but also appear to use factive verbs—including 'knows'—in transparently non-factive ways. Two hypotheses were presented to explain these linguistic data: the theory that non-philosophers tacitly hold a non-factive concept of knowledge, or the more parsimonious theory that knowledge for non-philosophers is factive, but that uses of 'knows' in everyday language often involve protagonist projection.[20]

[20] Turri (2011b) and Hazlett (2012) both speculate about the existence of different folk concepts for *know* and *know for certain* or *know it is true*. This prompted the subsequent follow-up to Experiment 3 to make sure participants aren't reading the 'really' in answer choices (e.g. 'really Φ') as an intensifier, cueing the latter concept. The same true and false FBI cases for 'know' and 'believe' were re-administered between-subjects ($N = 63$) with the following changes. First, the cases involved a particular FBI *agent* named Smith. Second, participants were asked, 'Which do you think best describes Agent Smith?' just in case bolding the verbs was contributing to different readings of 'know'. Third, answer choices dropped the intensifier (now 'Agent Smith thought he Φ the bomb was homemade' or 'Agent Smith Φ that the bomb was homemade'). Lastly, these participants were given a comprehension check explicitly acknowledging the truth or falsity of the complement. Out of 56 remaining participants, 92% of those in the *false 'know'* condition gave the *projectionist answer*. Alternatively 93% said that 'Agent Smith Φ-ed' for *false 'believe'*, 83% for *true 'know'*, and 94% for *true 'believe'* (Fisher's exact test, $p < 0.01$ for all three comparisons). Thus is seems unlikely that 'really' was doing much work in generating projectionist answers.

Current data suggest that the linguistic evidence considered, at least so far, has not been enough to show that Facticity is false, or that non-philosophers' concept of knowledge is non-factive. Conversely, the results of Experiments 1-4 seem to indicate that people accept uses of factive verbs in cases where complement clauses express false propositions, but that this fact can be explained without calling into question the very idea of Factivity for 'knows'. Instead, the results point towards projection. In Experiment 1 we saw that generally speaking, people were very wiling to adopt the perspective of their protagonists. In Experiments 2-3, we saw that projectionist paraphrasing of factive verbs arose specifically due to the truth or falsity of the complement clause across the same context of usage. And lastly, Experiment 4 demonstrated that projection is present in the actual judgments that experimental participants make when presented with the kind of epistemic thought experiments often considered by philosophers.[21]

A second interesting result of these studies is that there may be striking differences between folk usage of 'knows' and other purportedly factive verbs, e.g. 'learns', and 'realizes'. While still giving a majority of projectionist answers, participants in Experiment 1-2 gave significantly more non-factive answers for these verbs than for 'know'. Though this effect in Experiment 3 did not reach statistical significance, the same pattern of results is still suggestive. Of course, more research is needed to discover whether it really is the case that multiple readings are available for these verbs. But at the very least, it is beginning to seem that there is something special about 'knows' that doesn't also show up as readily for these other verbs. People seem to be adamantly adopting projectionist paraphrases rather than genuinely ascribing knowledge to protagonists with false beliefs in the cases considered.

[21] This factor will also be of potential interest to experimental philosophers seeking to empirically investigate intuitions in other traditional thought experiments where projection might arise.

And perhaps this is good news. Many epistemologists consider everyday uses of 'knows', utterances of the form '*S* knows that *p*', and the judgments people ordinarily make about knowledge ascription to play important roles in epistemic theorizing (for variations on this theme see Cohen 1988, Lewis 1999, Schaffer 2005, Stanley 2005). This renewed interest in ordinary language has recently inspired many new collaborations between traditional and experimental philosophers regarding the psychological factors that underlie knowledge attribution (for a summary see Buckwalter, *forthcoming*). But the concept of knowledge these philosophers are interested in analyzing is decidedly factive. So alternatively, if the folk concept underlying uses of 'knows' allows for false knowledge, then at the very least it seems we would need to decrease some level of credence in the idea that ordinary usage counts as good evidence for an analysis of *that concept*.[22]

Yet the actual evidence collected so far undercuts the evidence used to question Factivity, suggesting that (i) the ordinary language support of the standard truth condition for knowledge is left intact, and that (ii) the evidentiary status of ordinary usage in the analysis of the *philosophical concept* of knowledge is not questioned by this particular factor.[23] Of course, the data do not show that all would-be non-factive uses of 'know' are cases of protagonist projection. Nor do they show that there cannot be evidence against factivity independently of projection phenomena. What they do suggest however is that absent strong evidence to the

[22] Turri (2011) points out that ordinary usage may still be good evidence for *this project* in the majority of cases. It should also be noted that not all epistemologists are interested in *this project* (analyzing the concept of knowledge that philosophers happen to hold). Consider instead, normative or metaphysical accounts of knowledge. Nothing has been said about the role of ordinary usage in these latter kinds of epistemic projects.

[23] To be clear, present experiments remain neutral regarding the question of semantic entailment and presupposition. One could still claim on the basis of these data that 'know' only presupposes, rather than entails, that the propositional complement is true in ordinary language (Hazlett 2012). However, the burden of proof rests with those who would question orthodoxy to provide further experimental evidence for this claim.

contrary orthodoxy must still stand. Perhaps there are yet to be explored circumstances that do genuinely challenge Factivity in ordinary language. And it's entirely possible that there is conceptual variation between philosophers and non-philosophers along some other dimension.[24] But regarding the current question, the burden of proof seems to reside with those who would challenge Factivity. Dissidents must collect more empirical evidence to support the existence of a non-factive folk concept of knowledge, let alone justify divorcing traditional theories of knowledge from the relevant study of ordinary usage altogether, before we can consider abandoning orthodoxy.

2.4. Conclusion for the Factivity of Knowledge

We *know* that we are surrounded by non-factive uses of factive verbs like 'knows'—or at least—we thought we knew it. Experimental evidence has demonstrated that finding good linguistic evidence against Factivity, and specifically for an underlying non-factive concept of knowledge, is going to be more difficult than opening the newspaper. Under carefully controlled experimental conditions in which participants are given explicit paraphrasing tasks, we have seen that a more likely explanation of the linguistic evidence considered so far is that participants are adopting projectionist readings, rather than actually attributing false knowledge. And these projectionist readings undercut explanations of the linguistic data that support the argument that non-philosophers tacitly hold a non-factive concept of knowledge.

Thus we can conclude that without further evidence to the contrary, those philosophers who choose to justify provisions of factivity in their philosophical analyses of knowledge by

[24] Again, see for instance, recent work by Myers-Schulz & Schwitzgebel (*forthcoming*) regarding the standard belief condition for knowledge (though also see challenges to this research by Rose & Schaffer *forthcoming*).

appeal to the ordinary usage or the ordinary concept of knowledge now have empirically supported reasons for doing so. And, perhaps more importantly, we might also conclude that this result serves to highlight the need for empirical research in epistemology. For collecting accurate and reliable ordinary usage data should be at the heart of epistemic projects invoking the concept of knowledge best informing the meaning of 'knows' in ordinary talk, and the philosophical conclusions that follow.[25]

[25] Many thanks to Allan Hazlett, Richard Holton, Josh Knobe, Josh May, Eddy Nahmias, Jonathan Schaffer, Eric Schwitzgebel, and John Turri for detailed and helpful comments on previous drafts of this chapter.

Chapter 3. Stakes, Error, and Accommodation: Contextualism and Pragmatic Encroachment

The evidence presented in Chapter 2 suggests that both people tacitly hold a factive concept of knowledge. These results also reinforce the need for conducting carefully controlled experiments when making empirical claims about the ordinary conditions for knowledge.[26] From this basic starting point, we can now move on to evaluate a specific dispute currently underway in epistemology regarding the support that ordinary language practices may provide for theories of both contextualism and pragmatic encroachment. And also, from a more general prospective, we can glean insight into interesting pragmatic factors that might guide ascription and the application of the ordinary concept.

Perhaps the best way to become acquainted with this debate is to begin with the following thought experiment. Imagine that two spouses are waiting to deposit a check at the bank on a Friday evening and the lines inside are very long. One says to the other, "I know the bank will be open tomorrow" and suggests they return on Saturday morning instead. This mundane knowledge *assertion* seems true. However, now suppose instead that their check was actually for a large sum of money meant to cover a series of impending overdue bills, making its immediate deposit extremely important for their financial futures. It even occurs to them that the bank might have altered the open hours since their last Saturday visit. One says to the other, "I don't know the bank will be open tomorrow." In the latter circumstance, this knowledge *denial* seems true. Many philosophers have claimed that these contrary intuitions are best explained by

[26] Perhaps they should also increase our general credence in the idea that the ways non-philosophers use the word 'knows' can be counted as good evidence in analyzing the concept of knowledge in which philosophers have typically been interested.

the idea that practical stakes or the salience of error possibilities can play an important role in everyday evaluations of knowledge ascribing sentences.

The precise way of accounting for intuitions in these famous bank vignette pairs (see DeRose 1992) has played a prominent role in contemporary epistemology. As we observed in Chapter 1, one recent trend is to cite the content of our intuitions to support philosophical analyses of knowledge based on the way we actually make and assess knowledge ascriptions. In turn, the ability to explain how the same knowledge denial and knowledge assertion can both be true in these two different contexts has inspired support for some of our best new competing theories of knowledge. And while debate continues among contemporary philosophers, most agree that this ability to account for bank intuitions—or the purported fact that people are more likely to ascribe knowledge when stakes and error are low than when they are high—counts as one serious theoretical advantage.

Parallel to this theoretical work in epistemology, experimental philosophers of the last few years have also begun to investigate the factors that influence ordinary language practices involving knowledge sentences (for a review see Buckwalter 2012). They've done this by running experiments in the social psychological tradition by constructing stimulus materials closely linked to bank case vignettes. Specifically, these researchers have sought to understand the conditions and the mechanisms at work when we make third-person attributions of knowledge. However, the results these experimental philosophers have uncovered has inspired a great, and yet unsolved mystery. Current data suggest that people's answers in these kinds of cases systematically diverge from the predictions of trained epistemologists. As we saw in Chapter 2, this reinforces the need for careful empirical controls when studying ordinary attributions.

With few exceptions, this evidence to date indicates that when participants in an experiment are presented with bank-case style stimuli, the factor of stakes plays only a marginal role in their knowledge ascriptions to others.[27] This is puzzling since on the one hand, most professionals have agreed that stakes do or should play an important role in how we assess knowledge-ascribing sentences. But on the other hand, experimentalists have been unable to reproduce commensurate findings in actual practices. So could it be that these experiments are somehow deficient in detecting the mechanisms that buttress our ordinary language practices involving knowledge, and if so, how? Or alternatively, are theoretical intuition pumps in the hands of epistemic experts less accurate measures of everyday knowledge ascriptions, and if so, why? To answer these questions is to solve *the mystery of ascriber intuitions*.

Recently, two promising solutions to the mystery have been proposed by Keith DeRose (2011) and N. Ángel Pinillos (2011, 2012) along the former lines by questioning the designs of past experimental studies. They have suggested that folk intuitions only *appear* to radically diverge from professional epistemologists regarding subject stakes and attributor error due to various problems with the particular experiments that have been used to examine them. In response to these claims, this chapter presents two new experiments designed to directly test the solutions put forward by DeRose and Pinillos. After making adjustments for their worries, results vindicate previous findings by suggesting that (i) the solution to the mystery is likely not based on the particular empirical features these theorists have identified, and (ii) that the salience of ascriber error continues to make the difference in folk ratings of third-person knowledge-ascribing sentences.

[27] Since the writing of this chapter, new work by Sripada and Stanley (*forthcoming*) has claimed to detect a stakes effect in unrelated cases. However a critical discussion of these recent findings will be saved for a later occasion (see Buckwalter & Schaffer, *manuscript*).

In order to uncover the true culprits in our caper of stakes and error in ascriber intuitions, the chapter proceeds as follows. Section 3.1 reviews the evidence traditional philosophers and experimental epistemologists have generated involving third-person mental state attributions of knowledge. Sections 3.2-3.3 introduce the solutions proposed by DeRose and Pinillos, and respond with experiments designed to test them. Section 3.4 discusses possible resolutions to the mystery of ascriber intuitions based on these new data, as well as advances two further hypotheses about its origins. Lastly, sections 3.5-3.6, concludes by revisiting the implications such results have for two leading theories of knowledge supported by ordinary intuitions.

3.1. Professional Intuitions and Experimental Data

It is often taken for granted in the contextualist literature that people's intuitions in bank cases fluctuate between *low cases*, where the stakes or error possibilities of the case are minimal, and *high cases*, or where stakes and error are critical. Such intuitions have inspired some epistemologists to develop analyses of knowledge by focusing precisely on ordinary ascribing behaviors. The result is that in addition to traditional factors like evidence or justification, recent discussions of knowledge have also begun to include more practical considerations that seem to be at work in everyday judgments.[28]

One strategy to capture these purported intuitions between bank cases is to focus on the pragmatic conditions that are directly relevant to the subject of the bank vignettes. Theorists such as Hawthorne (2004), Fantl and McGrath (2002, 2010), and Stanley (2005), advance accounts of knowledge by observing that not only the truth-relevant facts, but also the practical facts of the main character's situation, seems to have a large effect on our bank case intuitions. Specifically,

[28] Indeed there is a real debate as to whether evidence must be understood in a truth-conducive way (see for instance Fantl and McGrath 2010).

the claim is that ascriptions of knowledge are sensitive to the practical interests of a subject, the importance that he or she is right, and the personal costs involved with being wrong. While each of these theories are subtly different, for simplicity we will refer to the common metaphysical view, *interest-relative invariantism* (IRI), as the view that roughly, whether or not a true belief counts as knowledge depends in part on what is at stake for the *subject*.

A different strategy to capture bank intuitions is to focus on the use of the word 'knows' and the factors that influence the truth of knowledge-ascribing sentences. Standard *epistemic contextualists* (DeRose 1992, 1999, 2005, 2009; Cohen 1988, 1999, 2004) claim that in everyday usage, the very same knowledge sentence often seems to be true in specific sorts of contexts but false in others. While different contextualists are free to debate the factors that influence these truth conditions, the two factors most often discussed are the stakes of the *attributor* when ascribing knowledge to a third-person, and also, whether or not the possibilities for error have been made salient for the attributor from context to context. The resulting semantic thesis regarding instances of 'knows that p' is that the truth conditions of knowledge-ascribing sentences can be different from one conversational context to another, based on those details of the *attributor's* situation.

One thing that interest-relative invariantists and advocates of contextualism have in common is that they both claim to capture ordinary language practices involving knowledge-ascribing sentences. To find out what these ordinary practices actually are, evidence is usually collected as follows. We are invited to consider vignettes, or pairs of fictional scenarios, in which all the features of the cases are fixed, and where we are asked to make a decision involving what the protagonist knows. Then between the cases, we vary the factors of stakes or error, to see if our intuitions about the character's knowledge changes. When we do this, if the result is that our

intuitions change from case to case, this suggests that these factors play a role in how we assess knowledge sentences. The results of such thought experiments in the philosophical literature have indicated that most philosophers agree that stakes play an important role in ordinary knowledge ascription.[29]

And beyond the careful manipulations presided over by today's leading epistemologists in these specific cases, there is something generally intuitive about the idea that the costs of one's beliefs weigh heavily on our assessments of doxastic states.[30] If this is the case, then it should be a straightforward task to empirically pinpoint stakes effects in our behaviors. And indeed, there are now several results in the social psychological literature surrounding this general issue of stakes and error independently of research on knowledge specifically. Mayseless and Kruglanski (1987) for instance, show that people's subjective probability estimates are affected by the desire to avoid judgmental mistakes, in view of their perceived costliness. Fischer et al. (2008) have shown that people make decisions involving loss with significantly less subjective decision certainty (that loss-decisions are more difficult to make than decisions involving gains), and that this effect of loss-framing systematically increases people's desire to search for confirmatory information like evidence. Similarly regarding error, hypotheses of selective exposure (see Smith et al. 2008) have long since held that people tend to avoid dissonant, but seek out consonant information, suggesting that the salience of error might affect our capacity and motivation for processing evidence in the face of critical information. However, in these discussions it is very important to distinguish the epistemic ascriber, or the person who

[29] It is important to note that this discussion references the received view in the epistemic literate on these intuitions, and not experimental evidence directly measuring philosophers' actual judgments. So it's possible that factors like publication bias against those without stakes intuitions could be playing a role in artificially inflating the near consensus about bank cases.
[30] Though not all philosophers report having stakes sensitive intuitions (see Schaffer 2006). This may point to the existence of important individual differences in bank cases and beyond.

claims someone has knowledge, from the epistemic subject, or the person of whom these claims are made. What the social psychology literature suggests is that regarding the latter, subjects in high stakes situations, or cases of high potential for error, will have significantly less confidence in their judgments.[31] Yet, what these results do not directly speak to is the former, or how such factors might affect third-person knowledge attributions, or our evaluations of other people's first person knowledge assertions.

This further question of third-person attribution is exactly what experimental philosophers have attempted to investigate.[32] Despite the intuitions of epistemologists, as well as the results from social psychology, these data in experimental philosophy have largely turned up negative results for stakes. As Schaffer and Knobe (2010) say when summarizing these findings, "This research suggests...that—contrary to what virtually all of the participants in the contextualism debate have supposed—neither stakes nor salience impacts the intuitions of ordinary speakers." The majority of these findings involve the familiar bank stimuli discussed above. While each experiment is slightly different, they all involve manipulating stakes by varying the importance of a protagonist's financial circumstances (for instance, whether or not they have an impending bill coming due), and error factor by varying whether a character proposes some kind of relevant alternative to the knowledge claim asserted (for instance, that banks can sometimes change their hours without notice).

---

[31] Another thing that social psychology seems to suggest is that any impact of stakes on knowledge goes through an effect on credence. But if this is the case, then this would show that stakes are not necessarily an independent fourth factor in knowledge along with justified true belief, but merely causally connected to belief (see Weatherson 2005).

[32] Phelan (*manuscript*) has also shown that subject stakes have a marginal impact on people's judgments about evidence. Beyond just looking at stakes and error, researchers have also shown that moral judgment can play a large role in people's willingness to ascribe knowledge (Beebe and Buckwalter 2010; Beebe and Jensen *forthcoming*, Buckwalter *manuscript*). Presumably, a theory of knowledge that wished to do justice to folk intuitions would also need to account for these epistemic judgments.

Four independent studies have investigated the way stakes and error influence folk knowledge practices in these particular cases. Buckwalter (2010) found that in cases where a subject in this situation makes the knowledge claim, "I know the bank will be open on Saturday," participants found it to be true no matter the stakes or error of the case. Similarly, Feltz and Zarpentine (2010) found that whether or not the epistemic subject made a knowledge assertion in a low stakes case, or denial in a high stakes case had no effect on people's judgments about their truth or falsity in the bank cases they tested. On the other hand, when participants were specifically asked to ascribe or deny rather than evaluate stated knowledge sentences, May et al. (2010) were able to detect a small effect of stakes (though mean judgments suggested that participants still agreed that subjects had knowledge despite this difference). Lastly, Schaffer and Knobe (2010), were able to detect a significant effect for error by making the error possibilities more salient through vivid and personal anecdotes (and we will revisit this idea in section 3.2).[33]

Of course, while all of these studies involve the same general bank context, they each have their differences (for instance they ask slightly different questions or use distinctive stimulus materials). Overall however, people tended to either strongly ascribe knowledge or judge knowledge statements to be true despite the stakes manipulation of these various bank experiments. Only one study (Schaffer and Knobe) was able to detect an effect for error. So at first glance, experimental data in philosophy appear to be at odds with evidence from three different sources: general intuitive seemings one might have about the role of stakes, the results contextualists and interest-relative invariantists have uncovered in select cases, and also, some

[33] One worry raised by Josh May (personal communication) is that it is often difficult to raise the salience of attributor error without also inadvertently increasing the subject's evidence.

neighboring results in social psychology concerning the impact of stakes on subjects' degrees of confidence.[34] But how could this be?

3.2. Bank Experiments and Epistemic Contextualism

Bank data generated from empirical studies conducted by experimental philosophers and from the intuitions of traditional epistemologists are at odds, and the game's afoot to explain why. We now turn to the first candidate solution given by DeRose (2011), according to which three important deficiencies in the experimental designs of past studies are said to be responsible.

3.2.1. DeRose Challenges the Data

In "Contextualism, Contrastivism, and X-Phi Surveys," DeRose raises several crucial objections against previous bank experiments, as well as the implications these data might have for standard epistemic contextualism. DeRose (2011) points out that the studies all fail to incorporate all of the crucial elements needed to successfully test the empirical predictions of contextualism. The result, claims DeRose, is that "it turns out the intuitive support for contextualism doesn't really face much of a wave of empirical trouble," and that "there are some severe problems with taking the results of any of these studies (whatever their aims) as undermining the intuitive basis for contextualism" (83).

According to DeRose, the first problem with earlier research is that some studies (May et al. 2010; Schaffer and Knobe 2010) use stimulus materials in which the epistemic subjects did not make an explicit knowledge statement within the confines of the actual vignette. Instead,

[34] Marginal results for stakes on ascribers' intuitions in bank cases may also call into question the assumption that the impact stakes have on subject's degrees of confidence mean that stakes will have an impact on ascribers' intuitions.

these researchers asked participants to rate whether or not they thought subjects had knowledge in the various cases. The worry is that in cases where participants are asked to ascribe knowledge, rather than evaluate a knowledge statement, they will be more likely to consider the particular features of their context (the experimental setting) rather than the epistemic subjects' context (as stipulated in the vignette). So, if further research were to explicitly test the predictions of contextualism, a better experiment would have participants evaluate the truth of knowledge statements made by a speaker in the actual vignette (as opposed to simply being asked whether a given character in that vignette has knowledge).

A second worry raised by DeRose is that previous experiments (Buckwalter 2010) fail to test all the crucial factors said to influence ordinary truth judgments of knowledge sentences. While it is true that some contextualists predict differences based on attributor stakes or the possibility for error, others suggest that it is the *interaction* of these two factors that affects attributors' conversational contexts. And as DeRose says in an earlier paper, "The best case pairs will differ with respect to as many of the features that plausibly affect the epistemic standards, and especially those features which most clearly appear to affect epistemic standards as possible" (2009, 54). In the name of simpler experimental designs however, bank cases are often presented by comparing the effect of stakes or error individually. By using vignettes that isolate these two factors, the concern is that the test pairs will not have included all of the relevant factors contextualists claim influence epistemic standards. So a better test would include cases with combinations of both high stakes and salience (as opposed to just one or the other).

Lastly, DeRose points out that the best test cases for contextualism should capture how ordinary speakers actually use the knowledge claims in question. Typically, people in situations with low epistemic standards will tend to ascribe knowledge, and people in high-standards

situations will deny knowledge. But some previous experiments asked participants only about the truth conditions of assertions (Buckwalter 2010; May et al. 2010). So the worry however is that by excluding people's evaluations of knowledge denials, we will have accidentally manipulated some property of knowledge assertion, and not stakes or error more generally. Here, one might appeal to what David Lewis called the "rule of accommodation" (Lewis 1979). According to this rule, if someone says something false in a given conversational context, we will seek to change the features of that context (via any number of specific context altering factors) such that the new context makes the content of that utterance true. Supposing that 'knows' is a context-sensitive term subject to Lewis' rule, it's possible that participants in the bank experiments are more likely to agree that a *high* knowledge denial is true than they are to agree that a *high* knowledge attribution is false simply because of accommodation, and not because they sincerely judge high conversational contexts irrelevant when evaluating knowledge sentences. So in order to test this claim, participants in future experiments would need to receive cases involving both knowledge assertion and denials across the various conversational contexts presented.[35]

DeRose claims that while previous bank experiments may have included just one of these various components, only cases meeting all of these challenges will serve as accurate measures for the intuitive basis of contextualism. In the meantime, the absence of any one of these features could well serve as the explanation to the mystery of ascriber intuitions. So it looks as though experimental philosophers need to go back to the lab in order to more accurately test whether the predictions made by contextualists about stakes and salience in bank cases have ordinary intuitions on their side.

[35] Though we might wonder, if the purported effect was driven primarily by accommodation what would knowledge have to do with it?

3.2.2. Meeting DeRose's Challenges

The worries by DeRose discussed in the last section serve as important challenges to the interpretation of previous bank results. So one straightforward strategy is to move forward by including these challenges as factors in new experimental studies. The following experiment was performed to further test contextualism in precisely this way, by importing DeRose's three worries into a new research design point by point. This is accomplished by independently varying three critical factors within the basic bank-style vignette: (i) the *speech act* of the character in the vignette could either be one of asserting that she has knowledge or denying that she has knowledge, (ii) the *stakes* could either be high or low, and (iii) the *error possibilities* could either be salient or non-salient. And given the findings of Schaffer and Knobe (2010) this manipulation of the salient error possibilities differed from other previous studies by using a *concrete and vivid* example of error.

This resulted in a 2 (stakes) x 2 (error) x 2 (speech act) between subjects design in which each participant was randomly assigned to one of eight possible conditions. The possible combinations of the resulting eight conditions are shown below.[36]

[36] This study used the internet-based commercial research tools Mturk and Qualtrics. Online samples were restricted to participants located in the United States.

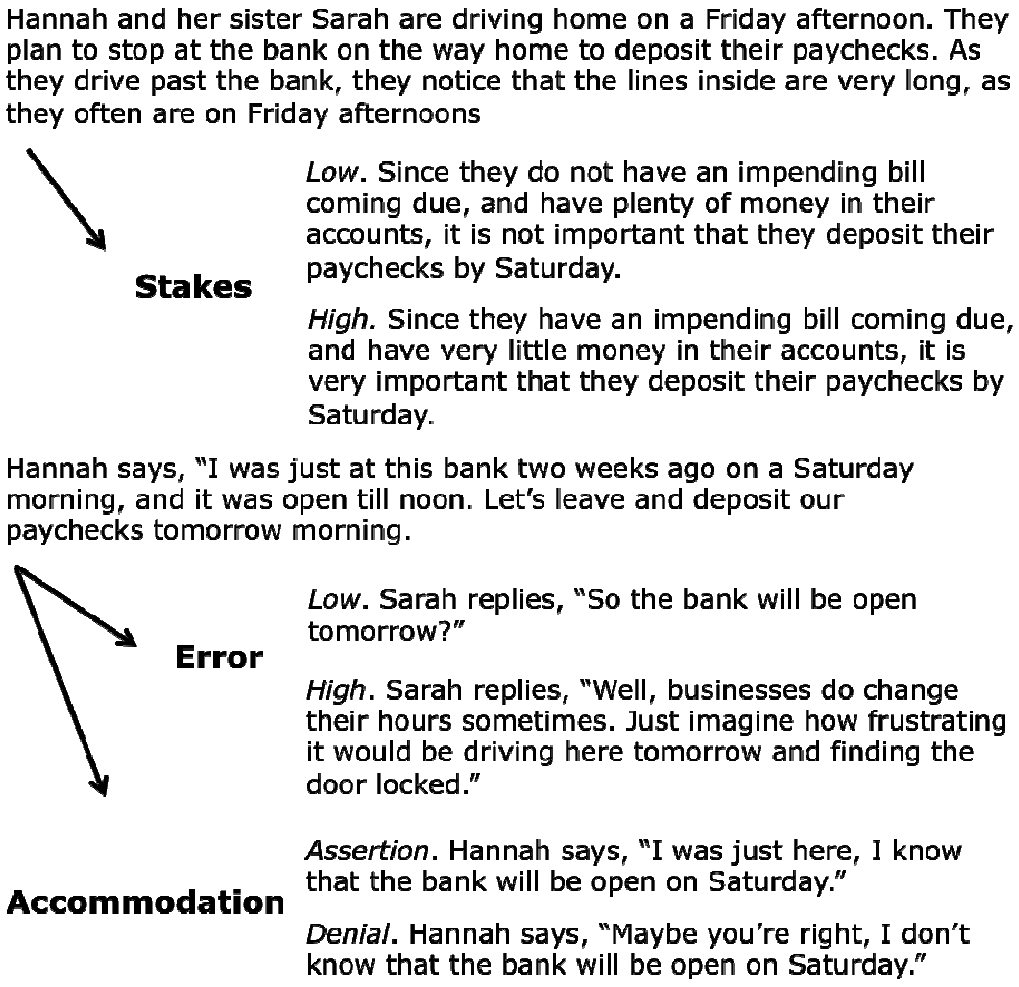

Figure 3.1. Construction of bank experiments through manipulation of stakes, error salience, and accommodation.

After seeing one of the possible bank case combinations, and receiving a pair of comprehension checks, participants (N = 215, 32% male) were then asked the following question.[37]

[37] 30 participants were removed from this study for failure to pass two very basic comprehension check questions.

Assume that as it turns out, the bank really was open for business on Saturday. When Hannah said, "I (*know / don't know)* that the bank will be open on Saturday," is what she said true or false?

Answers were assessed on a five-item scale anchored with truth-value terms (e.g., 1 = false, 3 = in between, 5 = true). Mean truth-value judgments in the bank cases are represented in Figure 3.2 and Figure 3.3 below:

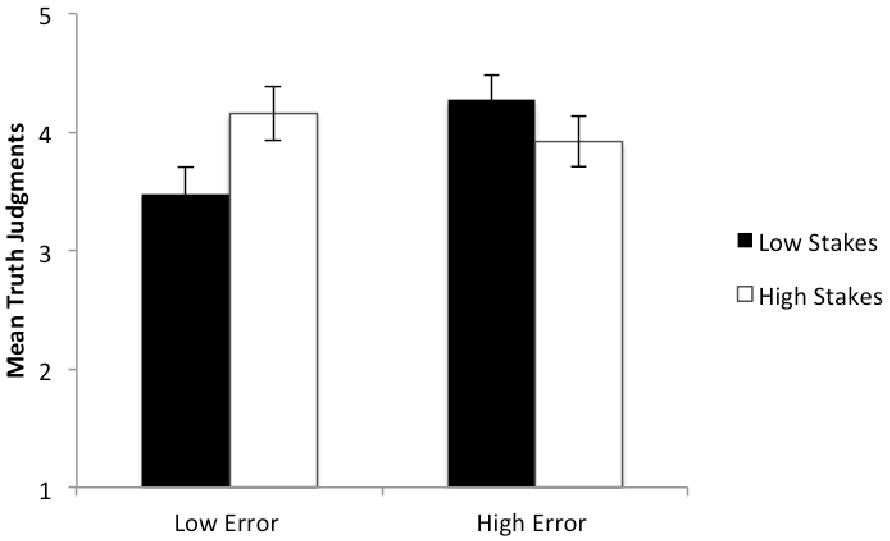

Figure 3.2. Knowledge Denial in Bank Cases. Depicted are mean truth judgments for knowledge *denials* grouped by error. All scales ran 1-5.

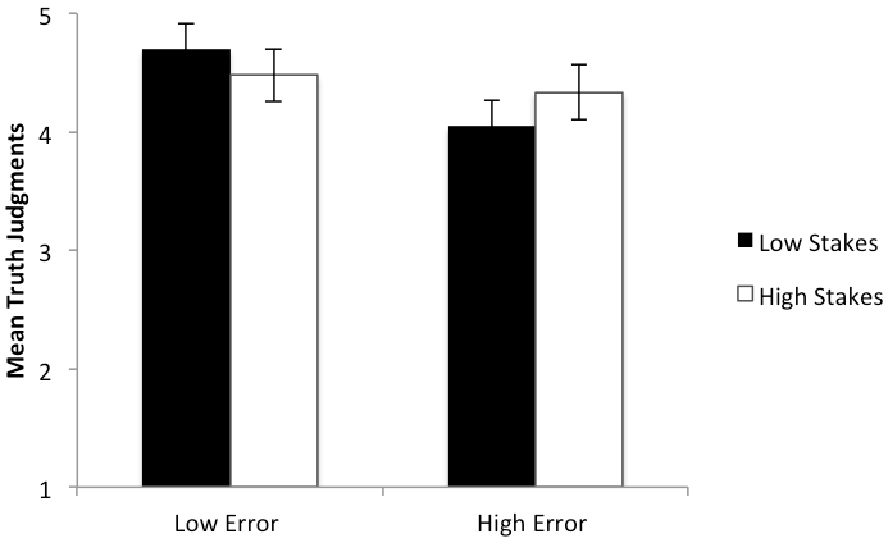

Figure 3.3. Knowledge Assertion in Bank Cases. Mean truth judgments for knowledge *assertions* grouped by error. All scales ran 1-5.

The study yielded three key results.[38] First, there was a main effect for speech act whereby people thought that knowledge sentences involving an assertion, no matter the error or stakes of the case, were more likely true than knowledge sentences involving denial.[39] However despite this main effect, it's also true that participants in the experiment just generally judged everything true across the board. Notice that in the graphs above, responses have not been recoded from their original values. But since a '5' in knowledge assertion conditions is logically equivalent to a '1' in knowledge denial conditions, we can clearly see how high truth value judgments in Figure

[38] Means and standard deviations for denial conditions: Low Error/Low Stakes (M = 3.48, SD = 1.62), Low Error/High Stakes (M = 4.15, SD = 1.05), High Error/Low Stakes (M = 4.27, SD = 1.08), High Error/High Stakes (M = 3.92, SD = 1.12). Means and standard deviations for assertion conditions: Low Error/Low Stakes (M = 4.70, SD = 0.56), Low Error/High Stakes (M = 4.48, SD = 0.59), High Error/Low Stakes (M = 4.05, SD = 1.30), High Error/High Stakes (M = 4.33, SD = 0.73). In the results to be reported below, a three-way between-subjects analysis of variance was conducted to evaluate the effect of error, stakes, and speech-act type on participants' truth value judgments in the bank cases.

[39] Main effect for the factor of speech act, (F (1,177) = 8.6, $p < 0.01$).

3.2 and Figure 3.3 indicate that participants gave very different responses between speech act types. This suggests that while people find the knowledge assertion true in *low* and a knowledge denial true in *high* as contextualists predict in bank cases, intuitions were largely driven by accommodation.

Second, we find exactly the impact of error possibilities that standard epistemic contextualists predict. When the possibilities for error were made salient (in a concrete and vivid way), participants were more inclined to say that the assertion of knowledge was false and that the denial of knowledge was true.[40] This effect is shown below in Figure 3.4 by collapsing across all the various levels of high and low stakes cases administered:

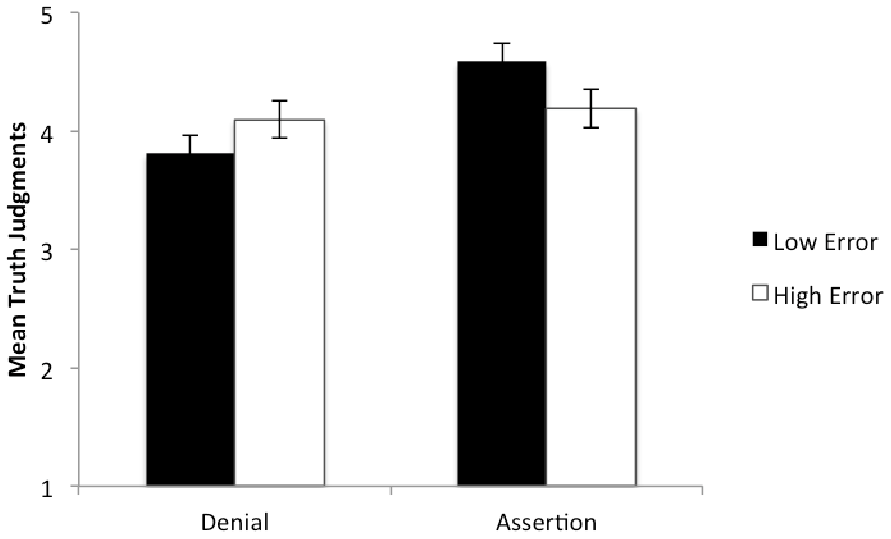

Figure 3.4. Knowledge by Salience in Bank Cases. Depicted are mean truth judgments for low and high error conditions grouped by *assertion* and *denial.* All scales ran 1-5.

[40] A significant interaction effect was found between factors of speech act and error, (F (1, 177) = 4.62, $p < 0.05$).

Lastly, despite obtaining the predicted effect of attributor error, the predicted effect of subject stakes was not found.[41] There was no general tendency for people in the high stakes bank conditions to be more inclined to think that assertions of knowledge were false or that denials of knowledge were true.[42] Even after correcting for all of the previous worries in bank case stimuli, an effect for stakes was not detected in people's evaluations of these particular third-person knowledge-ascribing sentences.

So here's the state of play. Previous experiments have demonstrated that there is little reason to think stakes or error play meaningful roles in folk assessments of knowledge ascriptions in bank cases. Then, DeRose proposed three objections to the extant ascription data that might be responsible for the incongruent results between philosophers and ordinary people. When incorporating those exact worries into the current test (along with the vivid error possibilities advocated by Schaffer and Knobe), we found that the actual problem with previous tests had to do with the error manipulation. That is, data seem to show that when error is vividly presented to participants in cases of both attributions and denials, there is an effect on third-person ascription.[43] So the key message from this further experiment on the bank cases seems to be that the possibility of error made salient to the attributor does have an impact on the evaluation of truth conditions of sentences that attribute or deny knowledge, but that subject stakes do not.

---

[41] No significant interaction effect was found between speech act and stakes, (F (1, 177) = 0.40, p = 0.53).

[42] Indeed, the only significant effect of stakes was an incredibly complex interaction. In cases with salient error possibilities, people were less inclined to say that the denial was true when the stakes were high (M = 3.92, SD = 1.12), whereas in cases without salient error possibilities, people were more inclined to say that the denial was true when the stakes were high (M = 4.15, SD = 1.05). In other words, there was an effect such that high stakes had opposite impacts on denials depending on whether error possibilities were made salient (F (1, 177) = 6.00, $p < 0.05$).

[43] Relative to Lewis (1996), this may suggest that simply mentioning error possibilities is not enough to make them salient in the relevant, epistemic context-altering way.

Of course, it is often difficult to confidently interpret null stakes results. Generally speaking, there could be numerous different reasons for why a given experiment does not find a hypothesized effect. While one reason could be that no such effect actually exists, any number of experimental confounds could also be responsible. One could also object that people simply weren't paying attention, or that they were not holding fixed the important epistemic features of the cases (like the amount of evidence possessed by an epistemic subject for instance) and that is why an effect of stakes on knowledge was not found on this particular occasion. While these are important worries to consider, it is not clear that they serve as viable explanatory hypotheses of the data in hand in the current bank study.

Importantly, the space of candidate explanations of the null results for stakes in the present experiment is constrained by the interaction effect of error possibilities and speech act type. Unlike the previous studies that have turned up negative or null results for both stakes and error, this study found the predicted effect for the latter but not the former in a single experiment across the very same stimulus materials. So a plausible explanation for the absence of the impact of stakes would still need to retain the ability to explain this result for error. Therefore it seems unlikely that one could rely on accusations about epistemic attention or shifting evidence to explain the lack of impact stakes had on knowledge judgments, while participants simultaneously behaved exactly as epistemologists were predicting when it comes to error. In this regard, we can have even more confidence than before to think that participants are considering these cases as epistemologists intend. It's just that they don't arrive at the *same epistemic intuitions* as the experts in these particular cases regarding stakes as they do for error.

3.3. Evidence-Seeking Experiments and IRI

While there now might be some doubt about the specific features of the cases above regarding contextualism and salience of error, these data seem to very clearly question the ability of interest-relative invariantist positions to capture ordinary intuitions regarding subject stakes in bank cases. We now turn to the second candidate solution given by N. Ángel Pinillos (2012) suggesting that problems specifically to do with the stakes manipulations of previous bank case experiments can account for the mystery of divergent ascriber intuitions.

3.3.1. Pinillos Challenges the Data

Pinillos (2012) offers some exciting new experimental evidence suggesting that unlike the bank data consensus in experimental philosophy, subject stakes do in fact influence third-person mental state attributions of knowledge. Specifically, Pinillos uses an experimental paradigm measuring the amount of evidence participants require an epistemic subject to collect before they ascribe knowledge to that subject. The stimulus involves a college student who is proofreading his assignment for typos. In a low-stakes condition, it is not particularly important that the assignment have no errors, while in a high-stakes condition, the student faces disastrous consequences should even one error be discovered by his professor:

> *Typo Low Stakes.* Peter, a good college student has just finished writing a two-page paper for an English class. The paper is due tomorrow. Even though Peter is a pretty good speller, he has a dictionary with him that he can use to check and make sure there are no typos. But very little is at stake. The teacher is just asking for a rough draft and it won't matter if there are a few typos. Nonetheless Peter would like to have no typos at all.

> *Typo High Stakes.* John, a good college student has just finished writing a two-page paper for an English class. The paper is due tomorrow. Even though John is a pretty good speller, he has a dictionary with him that he can use to check and make sure there are no typos. There is a lot at stake. The teacher is a stickler and guarantees that no one will get an A for the paper if it has a typo. He demands perfection. John, however, finds himself in an unusual circumstance. He needs an A for this paper to get an A in the class. And he needs an A in the class to keep his scholarship. Without the scholarship, he can't stay in school. Leaving college would be devastating for John and his family who have sacrificed a lot to help John through school. So it turns out that it is extremely important for John that there are no typos in this paper. And he is well aware of this.

After seeing one of these conditions, participants were asked, "How many times do you think [*Peter/John*] has to proofread his paper before he knows that there are no typos?" They were then told to fill in the blank with the number they thought was appropriate.

The study showed that when participants were presented with either the low-stakes or the high-stakes conditions, people thought that the student needed to collect more evidence in order to know there were no typos when the stakes of the case were high (median = 5) than when they are low (median = 2). The finding suggests that since people require more evidence before ascribing knowledge to epistemic subjects in this way, folk attributions of knowledge must be sensitive to stakes.

Furthermore, Pinillos suggests that this experiment may give us a unique perspective on what went wrong in bank cases. The worry, claims Pinillos, is that there really is no way to track

whether participants are holding fixed crucial details of the cases, such as how much evidence the epistemic subject has between conditions. The thought is that evidence-seeking experiments are better equipped in this regard, since it is precisely the amount of evidence the subject should have which is measured. By detecting fluctuations in the amount of evidence required, these experiments show differences in knowledge judgments by stakes, casting doubt on the divergent intuitions previous experimental philosophers have detected in bank cases. So regarding theories like interest-relative invariantism, it looks as though experimental philosophers also need to return to the lab, in order to find out exactly how stakes are influencing judgments about evidence-seeking behaviors when ascribing knowledge.

3.3.2. Meeting Pinillos' Challenges

One latent worry in the evidence-seeking design is that the differences detected in the amount of evidence collected between low and high stakes subjects in this particular experiment could arise not because third-person mental state attributions of knowledge are or aren't intrinsically sensitive to stakes, but rather because high-stakes subjects are expected to collect more evidence than low-stakes subjects to actually have an outright belief on the issue at all. So a further study was run to answer the question of whether stakes specifically affect mental state ascriptions of knowledge, or alternatively, whether these differences are instead an effect for some other mental state, like belief.[44]

In this study, 100 participants were given a manipulation as close as possible to what is used in Pinillos' study involving subject stakes, but also varied the kind of mental state ascription

[44] For similar stakes results for other verbs besides 'know' and 'believe' see Buckwalter and Schaffer (*manuscript*).

that was attributed to that subject.[45] This resulted in a 2x2 between-subjects experimental design, independently varying the practical stakes (either high or low) and the mental state (either belief or knowledge) of the subject.[46] After seeing one of the same stimuli given above in Pinillos' original experiment, participants were than asked, "How many times do you think Peter has to proofread his paper before he [*believes/knows*] that there are no typos?" and to "Please insert the number you think is appropriate in the space below." Results are represented in Figure 3.5 below by mean scores of the amount of evidence needed in each case:

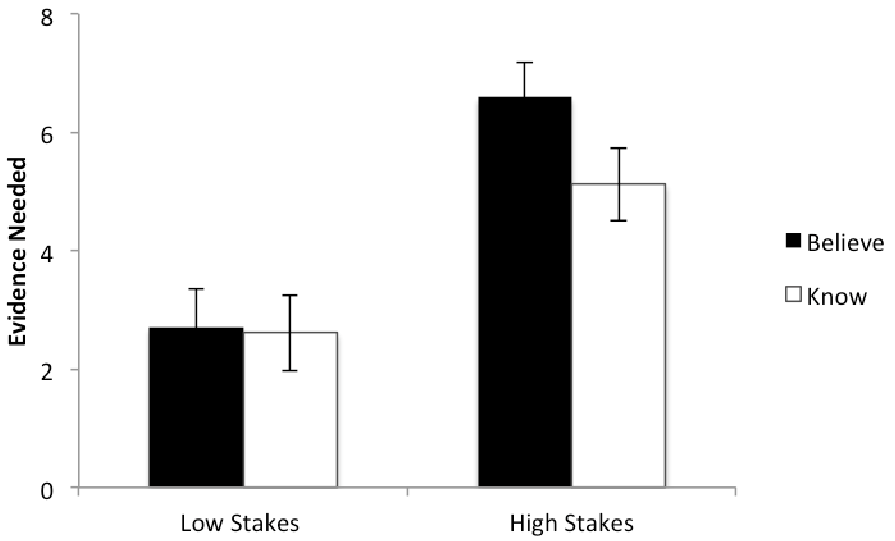

Figure 3.5. Pinillos Typo Case Results. Depicted are mean responses grouped by stakes and mental state.

We find that as before, stakes had a huge impact on ascriber intuitions. However, while the experiment showed a significant difference on people's judgments between low and high stakes

[45] 10 participants were removed from this study for failure to pass a very basic comprehension check.
[46] These materials are borrowed directly from Pinillos (2012).

contexts, the specific mental state they were asked about within these contexts did not.[47] In other words, participants gave roughly the same answers when they were asked how much evidence was needed to be collected before the epistemic subject had knowledge as they did how much evidence needed to be collected before the subject had a belief that a certain result would obtain.

In the dispute between intellectualists and interest-relative invariantists, advocates of IRI usually hold that subject stakes are themselves supposed to bear directly on the criteria for whether a subject's true belief constitutes knowledge. Yet, identical scores between subjects in questions regarding whether an epistemic subject knows something is the case and believes something to be true suggests that the effect shown for stakes does not tell us about the specific criteria people are actually using when deciding if a subject's belief constitutes knowledge. Instead, the fact that participants found that high-stakes subjects are expected to collect more evidence than low-stakes subjects in order to count as believing suggests that that subjects in high-stakes conditions are expected to collect more evidence just to make up their minds at all. It's not that both subjects' preexisting beliefs are transformed into knowledge with a greater amount of evidence in high stakes cases than low stakes cases, but rather that epistemic subjects need a disproportionate amount of evidence when forming the requisite belief.

This suggests that this particular evidence-seeking experiment is a fine-grained enough measure to have detected that subject stakes have some kind of an effect on participants' general responses in the current experiment, but not fine-grained enough to show that the relevant effect

[47] Typo Low Stakes Belief (M = 2.71, SD = 1.27), Typo Low Stakes Knowledge (M = 2.61, SD = 0.89), Typo High Stakes Belief (M = 6.59, SD = 5.05), Typo High Stakes Knowledge (M = 5.12, SD = 3.42). A two-way between-subjects analysis of variance was conducted to evaluate the effect of Mental State and Stakes on participant free responses regarding evidence. A significant main effect was obtained for stakes, F (1, 86) = 23.1, p <0.01. However, no main effect was found for Predicate, F (1,86) = 1.40 p = 0.24, and no interaction between these two factors was detected, F (1, 86) = 1.05, p = 0.31.

reveals something specific about knowledge in particular.[48] For better evidence supporting the empirical predictions made by IRI, data would need to demonstrate that the stakes of the case are what matter for a subject's true belief to count as knowledge.[49] To be clear, nothing in this response to Pinillos rules out the possibility that stakes do actually play such a role in people's judgments, or that such evidence could be collected in the future. Instead, this experiment was designed to show that we are not warranted in inferring support for IRI from these particular data regarding stakes. Further research is necessary before supporters of interest-relative invariantism can reasonably make the inference that knowledge ascription is particularly sensitive to stakes in the relevant way. And, antecedent empirical reasons for the conclusion that stakes do not play this role should inspire caution before overturning previous results to the contrary.

3.4. Towards Solving the Mystery

We began with a great mystery. How can we make sense of the systematic differences between professional intuition and folk judgments in bank cases? Data indicate it is unlikely the mystery can be entirely solved by appealing to the specific problems that DeRose and Pinillos have identified with previous experiments. However, neither does the evidence suggest that philosophers using more traditional methods were completely mistaken about actual knowledge practices. Joining with past research, further data on bank case intuitions suggest that while error is a factor that influences ordinary third-person mental state ascriptions of knowledge, stakes are

[48] The supporter of IRI might respond to this objection by claiming that if knowledge is a norm of belief, then identical results between mental states would be compatible with the impact of stakes on people's criteria for knowledge ascription. This is certainly a *possibility*, just one that remains to be proven experimentally.

[49] Indeed, another worry is that a possible ambiguity exists whereby a more natural reading of the question asked in these experiments is something like, "how many times *should* the subject proofread his paper in this situation?"

not. The latest experiments show that knowledge ascriptions in the bank cases fluctuate when vivid error possibilities are made salient. Data from evidence-seeking experiments on the matter of stakes is shown to be inconclusive, and without further testing, does not yet undermine the growing consensus in experimental philosophy that subject stakes play but marginal roles in attributor judgments.

Given these data, it seems that the solution to the mystery of ascriber intuitions is twofold. Something went wrong when previous experimental epistemologists claimed that the salience of ascriber error does not affect people's knowledge judgments, and something went wrong when professional epistemologists claimed that their intuitions about the importance of subject stakes *actually reflect ordinary people's evaluations* of knowledge-ascribing sentences. While the empirical data is a good start to solving this mystery, an interesting further question remains why experimentalists and traditional philosophers made the mistakes that they did. Before going on to speak of the philosophical ramifications of the data in these cases, we will pause to hypothesize about why or how this mystery developed.

One hypothesis to explain why previous bank case studies were unable to detect the intuitional variance epistemic contextualism predicts between contexts of low and high ascriber error seems relatively straightforward. In past experiments on bank cases, the factor of error was manipulated by only minimally mentioning the possibility of error. In Buckwalter (2010) for instance, the high error bank case interlocutor challenges her epistemic subject by speculating on only one general way in which she could be wrong (e.g., "Banks are typically closed on Saturday. Maybe this bank won't be open tomorrow either"). Similarly, May et al. (2010) take a similar tack when constructing case of high error by having the epistemic subject point out that generally speaking, banks do change their hours. However, current empirical evidence suggests

that the effect was detected in the present experiment simply by making the particular error manipulation in the bank vignettes more vivid, "Just imagine how frustrating it would be driving here tomorrow and finding the door locked." Following Schaffer and Knobe, who showed a salience of error effect for knowledge attribution, present results demonstrate a similar effect when participants are asked to make truth judgments of knowledge ascriptions. [50]

The first culprit then, that helps explain the advent of the mystery of ascriber intuitions, is a specific problem with earlier experimental research. The problem was that the error possibilities were not made salient enough in bank cases to detect the difference in intuition between low and high conversational contexts.[51] Regarding the disagreement about the factor of stakes however, it still remains incredibly puzzling why the intuitions found in these experimental studies continue to diverge so systematically from both the intuitions of trained epistemologists, as well as the predictions that results concerning first-person confidence might have made for third-person knowledge ascription. Further experimental evidence continues to support the hypothesis that knowledge is sensitive to error, while continuing to question the joint hypotheses that practical stakes, and the link between subjects' degrees of confidence and stakes, have anything but marginal impacts on ascribers' intuitions in bank cases.

One hypothesis to explain this difference is that ordinary people and trained epistemologists approach these thought experiments in different ways (see for example Phelan, *manuscript*). On the one hand, participants of an experiment usually experience one particular case, and are then asked to report an immediate intuition about whether or not the epistemic subject has knowledge. By contrast, trained philosophers often proceed by considering pairs of

[50] The effect shown here for error is smaller than what was shown in work by Schaffer and Knobe (2010).
[51] Feltz and Zarpentine investigating a range of life or death cases outside of bank contexts where subject stakes are quite vivid, but are also unable to detect stakes effects.

bank vignettes together and then engage in a kind of reflection about whether the relevant differences in context have any epistemic importance. The philosophers then go on to make predictions about the judgments ordinary people will make in these cases on the basis of that evaluation. And these two different types of approaches to epistemic judgments may be shaped by two very different kinds of psychological processes. The former seems to be an implicit system-one intuition generating capacity that enables us to respond to epistemic intuitions in particular cases with which we are presented. The latter is a system-two process, involving a more abstract set of theoretical beliefs about epistemic principles, as well as predictions about how others might conform to those principles.[52]

So could these different decision making approaches between these two different groups account for the mystery in bank cases? Indeed, it's possible that the predictions made by philosophers are subject to a kind of *distinction bias* often discussed in behavioral economics (Chatterjee et al. 2009; Hsee and Zhang 2004, 2010). The basic idea is that when presented with several vignettes differing by stakes—in similar to what Hsee and Zhang call "joint evaluation mode"—professional philosophers identify this feature of the vignettes and then make choices and form predictions based on their training and knowledge of the abstract epistemic principles involved. But conversely, the processes underlying ordinary people's judgments when they are presented with singular cases—or in "single evaluation mode"—are based on preferences related to actually experiencing those particular cases. And since it seems likely that the preferences that philosophers use in the former mode of evaluation will be different those of non-philosophers in

[52] Phelan (*manuscript*) tests a similar hypothesis by asking participants to judge which factors *should* affect an epistemic subject's confidence in her beliefs. While participants profess to the general principle that sakes should affect confidence judgments (about evidence at least), they fail to allow the costs of being wrong to influence the *actual* judgments they in these cases when they experience them.

the latter, these different modes of evaluation may encourage philosophers to *overpredict* the impact of stakes in people's actual knowledge judgments in bank cases.

So regarding the mystery of ascriber intuitions, perhaps the second culprit is the bias that arises from the combination of formal philosophical training, together with making predictions when cases are presented under joint evaluation rather than experienced. If true, this hypothesis may be able to help explain how judgments made by expert epistemologists gave rise to the importance of stakes, but also how the role of stakes in ordinary bank case judgments was mispredicted or exaggerated.[53]

3.5. Implications and Philosophical Importance

Though we have made some empirical progress in resolving the mystery of ascriber intuitions, the mystery's denouement raises perhaps an even more complicated philosophical question: how does this explanation about stakes and error in bank cases bear on epistemic contextualism and interest-relative invariantism?

Beginning with contextualism, many philosophers have argued that as a semantic theory, it makes particular linguistic commitments about word usage. Hawthorne (2004) and Stanley (2005) argue for instance that the relevant kind of context-sensitivity of 'knows' is objectionable because our usages of that particular word frequently deviate from usages of other common indexicals said to be context-sensitive. Going even further, Brown (2011) argues that since contextualism provides a linguistic model for 'knows', and given that such models provided by

[53] If this solution is correct, then these differences seem to highlight the need to institute more careful controls when utilizing the evidence-by-intuition method in philosophy. Such methods may be just as susceptible to criticisms one might make of any research program in psychology regarding experimental design confounds, or biases (see for example, order effects in trolley problem intuitions among professional philosophers by Schwitzgebel and Cushman, *manuscript*).

leading theories of contextualism are committed to certain kinds of context sensitivity, contrary behavioral data about folk knowledge practices regarding such sensitivities would actually threaten to undermine the view entirely.

If these arguments are correct, then getting the right results in bank case experiments seems crucial. As it stands, one outcome of research up to this point is that semantic theories of standard epistemic contextualism are supported by experimental data showing an effect for at least one specific kind of context sensitivity. Particularly, data show that such views that wish to include a correct theory of people's ordinary language practices regarding sensitivity to conversational contexts should focus not on the practical stakes, but rather on *the error possibilities made salient to the attributor*. In such cases, contextualism would not be undermined by the current experimental results. The clear upshot is not only has contextualism been shown to be compatible with the relevant knowledge behaviors, but also that such empirical evidence can be used to forge more detailed versions that specify with greater accuracy the relevant linguistic model claimed for the word 'know'.[54]

Unlike the data relevant to contextualism showing folk sensitivity to error possibilities however, current results continue to suggest that third-person attributions were insensitive to stakes. And, such findings seem to be clearly at odds with the premise that interest-relative invariantism best explains bank intuitions. In response to this tension, Brown (2011) argues that such experimental data only threaten to undermine one popular way of arguing for IRI, but not the position itself. Following Brown, we might note that there is no necessary dependence of the truth of metaphysical theses regarding things like temporal parts, the nature of substance—and likewise the determinants of knowledge—on concordant folk judgments or practices regrinding

[54] Specifically, the present evidence may tell against "pragmatist" contextualists, according to which practical matters partly determine the truth of knowledge ascriptions.

their central theoretical entities. As a metaphysical theory of knowledge, the truth of IRI does not turn on, and is not committed to ordinary intuitions about knowledge *per se*. Therefore, despite the results of the current studies, interest-relative invariantists are free to continue to include premises about the epistemic roles of subject stakes for the metaphysical determinants of knowledge at the cost of ordinary language.[55]

What the empirical evidence does seems to suggest, *so far* at least, is that IRI may not provide the best explanation for our epistemic behaviors regarding stakes in bank cases and beyond. In other words, the data from Pinillos has not been enough to convince us that ordinary assessments of these particular types of knowledge ascribing sentences count in favor of the metaphysical view that knowledge is stakes sensitive. And, while a metaphysical view need not enjoy folk agreement to be true, a safe bet is that the conclusions of such a thesis are generally speaking, more likely true when supported by true premises rather than false ones. So this may encourage future supporters of IRI to develop and embrace alternative arguments for their view based on something other than folk practices regarding stakes (again see for example Brown 2011; Fantl and McGrath 2010). Of course, there's always the possibility that future experiments will discover the long lost case that does display persistent stakes effects. And such cases will have to be evaluated on their own merit as they arise. But at the very least, the current data generated in response to DeRose and Pinillos—joined with the difficulty several independent

[55] Though it is important to note that this is nonetheless a considerable theoretical blow, since IRI was not on the board before it was claimed that the alleged intuitions needed to be accounted for.

researchers have faced in detecting anything but negligible stakes effects—begin to question whether building an epistemology around folk stakes sensitivity is a very good idea.[56]

3.6. Conclusions for Contextualism and IRI

Experiments continue to suggest that accommodation and the salience of ascriber error, but not subject stakes, makes the difference in the ordinary evaluation of third-person knowledge sentences. But research exploring the ways in which people actually evaluate knowledge sentences can still be a benefit to the more traditional research in the field, serving to help supplement, and not supplant, such methods when appropriate. One of the main goals of this chapter has been to show that the empirical investigation of the mystery of ascriber intuitions can help contextualists and interest-relative invariantists become the best versions of themselves.

By suggesting which specific features of an attributor's context ordinarily affect the standards for knowledge, this research begins to allow a more accurate estimate of the linguistic model of the word 'knows'. The result is that one way to *be a better contextualist* is to develop versions of the theoretical view in which the context sensitivity of 'knows' varies by accommodation and error. Similarly, experiments also continue to question the evidence supporting the claim that ordinary third-person knowledge judgments are sensitive to subject stakes. Such results may undermine one particular way of arguing for the thesis that knowledge is stakes-sensitive. Yet, they may also suggest that one way to *be a better interest-relative invariantist* might be to accept versions of the view that do not rely on premises concerning ordinary knowledge practices that people may not—or may not prevalently have. In both cases,

[56] See for instance, Jonathan Weinberg's notion of "philosophical effect size" whereby the simple detection of a psychological effect may not always be sufficient for supporting certain roles in philosophical argument without meeting a series of further conditions (2011).

empirical research in epistemology is additive to the theoretical work and methods in the field, inciting new directions for the partnership between future theoretical and experimental work on contextualism and interest-relative invariantism.[57]

[57] Many thanks to James Beebe, Keith DeRose, Mikkel Gerken, Josh Knobe, Josh May, Jennifer Nagel, N. Ángel Pinillos, Jonathan Schaffer, Jason Stanley, Jonathan Weinberg, and other blog members who participated in lengthy discussions on the *Certain Doubts* blog, for helpful comments and suggestions. I am grateful to Josh Knobe, Jesse Prinz, and Stephen Stich for insightful comments on previous drafts.

Chapter 4. The Moral Component of Knowledge: Back to Gettier

Chapter 3 has demonstrated that in addition to traditional things like justification or evidence, knowledge attribution may also be sensitive to some *very surprising factors*. As we have seen, leading pragmatist views in epistemology claim that the truth of a knowledge ascription is sensitive to stakes, or the practical interests of an ascribee (Fantl & McGrath 2010; Hawthorne 2004; Stanley 2005). On the other hand, many contextualists claim that the truth of a knowledge ascription depends on the degree to which error possibilities have or have not been made salient to an attributor (DeRose 2009; Cohen 2004).

Since these claims are largely descriptive in nature, this new trend emphasizing ordinary ascription has also inspired new debates between epistemologists and experimental philosophers (DeRose 2011; Nagel 2010, *forthcoming*; Schaffer & Knobe 2010). As we saw in Chapter 3, these debates mostly focus on testing the claims or predictions of pragmatist and contextualist theories in epistemology. The work presented in the previous chapter suggests that when subjected to empirical scrutiny in properly controlled experiments, there is some evidence to support the claims epistemologists have made regarding error salience and accommodation (Schaffer & Knobe 2010; Buckwalter *forthcoming*).[58]

But importantly, experimental philosophers have also discovered that these are *not the only surprising factors* at work in ordinary attributions.[59] Following the work of Joshua Knobe on intentionality judgments (Knobe 2003; 2004), it has also been suggested that there is an

[58] For a debate about the best way to interpret these data regarding stakes see Buckwalter (2010), Feltz & Zarpentine (2010), May et al. (2010) and Buckwalter & Schaffer (manuscript).
[59] Also see effects for culture (Weinberg, Nichols & Stich, 2001), native language (Vaesen and Peterson, *manuscript*), and order bias (Swain et al., 2008). For a review of the latest research in experimental epistemology see Buckwalter (2012).

important *moral component* to knowledge attribution (Beebe & Buckwalter 2010). This effect, known as the *Epistemic Side-effect Effect* (hereafter ESEE), demonstrates that people's prior evaluative judgments—and in particular their moral judgments—can influence judgments they make about knowledge attribution. Though there has been significant debate among cognitive scientists regarding the best explanation of this effect, what researchers have found time and again is that morality seems to have a robust impact on what we take others to know (Beebe & Jensen, *forthcoming*).

In comparison to the attention given to empirical data in experimental epistemology regarding stakes or error salience though, the impact of morality on epistemic judgments has gone largely overlooked. But should epistemologists also care about the influence that morality seems to have on knowledge attribution practices? In this chapter, new experimental evidence is presented suggesting that the moral component of knowledge attribution may have a series of important implications for epistemologists sympathetic to the ordinary language approach. These data show that the general effect shown in ESEE can be extended, to overturn standard intuitions in some of the most theoretically central thought experiments used in contemporary epistemology: Gettier cases.

In what follows, section 4.1 reviews the extant literature in experimental philosophy regarding the moral component of knowledge attribution, as well as the disagreement surrounding rival psychological explanations for this effect. Section 4.2 is a short discussion of the role of Gettier intuitions in epistemology. Section 4.3 presents three studies demonstrating that when typical Gettier cases are *made ESEE*, moral valence significantly influences participants to attribute knowledge to Gettier subjects. In Section 4.4, it is argued that while this research is currently ongoing, the correctness of each rival psychological explanation reviewed

in section 4.2 will bear heavily on (i) the typical use of Gettier intuitions in epistemology, and (ii) the recent practice of using data concerning ordinary knowledge attribution as evidence for theories of knowledge.

4.1. The Moral Component of Knowledge

Perhaps one of the most famous discoveries in experimental philosophy to date is that moral judgments can have a pervasive impact on the application of non-moral concepts.[60] For instance, in Knobe's well-known study on intentionality judgments (2003), he constructed two vignettes that described the circumstances of a company chairman learning of a potential new program to increase profits. However in each case, adopting this new program has a side effect—it will either help or harm the environment. Since the chairman is indifferent toward these side effects, he decides to start the new program in both cases. The result is that the environment ends up being helped or harmed, respectively. What Knobe found is that people's moral judgments in these cases had a large impact on their intentionality judgments. Specifically, he found that 82% of participants in the harm condition agreed that the chairman intentionally harmed the environment, whereas 77% of subjects in the help condition said that the chairman did not intentionally help the environment. This effect is known as the side-effect effect.

What Beebe & Buckwalter (2010) have demonstrated in ESEE is that moral considerations like those prominent in the side-effect effect can also affect people's intuitions about the chairman's knowledge. In their study, they gave undergraduate students (N = 749) the same cases Knobe used above, but instead asked participants to ascribe knowledge to the chairman regarding the occurrence of either the good or bad side-effect. And surprisingly, they

[60] For a review of these effects see Knobe et al. 2012.

found the same basic result. Participants were more likely to say that the chairman in the vignettes knew that his actions would bring about the side effect when the outcome was bad, and less likely to attribute knowledge when the side effect was good. This effect has been widely replicated, suggesting that just like the intentional side-effect, people's prior moral judgments can also play an important role in their knowledge attributions (see Beebe & Jensen, *forthcoming*).

The general phenomenon regarding the influence of morality on intentionality or knowledge judgments is now reasonably well understood. However the psychological explanation for why these effects occur has become the subject of continuing and ever-broadening controversy (see Knobe 2010). Rival explanations are often divided into two broad families of views. One family of views holds that moral considerations *distort* the ordinary application of non-moral concepts like knowledge. Though these views have been developed in a number of different ways, they all posit some sort of additional cognitive process triggered by the presence of moral considerations that biases people's judgments. For instance, Nadelhoffer (2006) argues that moral considerations give rise to certain affective or emotional responses, and these emotional responses in turn bias the judgments participants make. Alternatively, Adams & Steadman (2004) suggest that this effect is due entirely to conversational pragmatics. Still others argue that people only apply the concepts in question because they are looking for a way to justify their desire to blame, or hold agents morally responsible for their actions (Alicke 2008; Malle & Nelson 2003).[61]

The other family of views holds that morality is an important part of people's core conceptual competences. As opposed to distortion accounts, *competence explanations* claim that

[61] For evidence against these three accounts see Young et al. (2006), Knobe (2004) and Guglielmo & Malle (2010), respectively.

there is no additional cognitive process distorting ordinary judgments. Under this type of explanation, the influence of morality is said to reflect the genuine use of the concepts in question (Knobe 2010). For example, one prevalent theory along these lines by Halpern & Hitchcock (2011) suggests that moral judgments play an important role in counterfactual reasoning, and that these counterfactuals in turn underlie the correct application of the concepts in question. Similarly, Schaffer & Knobe (2010) argue that moral considerations fix the class of alternative outcomes people consider relevant, and these contrast classes affect the truth of knowledge ascriptions. Generally though, the basic idea is that moral considerations are an important part of the genuine use of concepts like intentionality or knowledge.

Yet despite the proliferation of data for and against theories of both sorts, no single theoretical proposal in either family has emerged the clear victor (see Knobe et al. 2012). It is fair to say that this debate has reached something of an impasse. The impasse has led some researchers to reconsider whether experimental philosophers actually possess the necessary tools for classifying psychological effects in terms of competence or performance at all (Alexander et al. 2010a, b). Some have posited that there are stable individual differences responsible for the effects in question (Pinillos et al., *forthcoming*; Nichols & Ulatowski 2007). Still others argue that there is unlikely to exist any one unified explanation for these phenomena, and that many overlapping factors are involved in the impact moral judgment has on the ordinary application of concepts like intentionality or knowledge (Phelan 2011).

Research into the best explanation for the role of morality in knowledge attribution is ongoing, and debate is sure to persist among philosophers and cognitive scientists. Our interest presently is in the lessons that epistemologists can learn from the moral component of knowledge attribution in the meantime. It is argued that at the current state of research, the general

correctness of *either* competence or performance style explanations each may have important implications for epistemologists interested in the conditions under which knowledge is ordinarily ascribed. To help frame these implications, and to further explore the role of morality in people's epistemic judgments, we revisit the basic effect discovered in ESEE. What we find is that ESEE case results can be expanded to significantly influence central thought experiments of the philosophical tradition: Gettier cases.

4.2. Gettier Cases in Epistemology

In contemporary epistemology, many refer to Gettier cases as standard counterexamples to the previously received justified-true-belief account of knowledge (1963). More specifically, Gettier cases are thought experiments aimed at eliciting the following intuition. If an epistemic agent in a Gettier case does not know a true proposition (that she believes and is justified to believe), than this is evidence against the view that knowledge is justified true belief. Consider the following example:

> Sam hears on the news that a beneficial new chemical has found its way into the town reservoir. As the water pump operator, Sam continues his duties believing that pumping the water will help the local crops. Sure enough, the crops thrive. So Sam's belief is true. However, now imagine everyone was wrong about the beneficial chemical, and instead a mysterious fungus was responsible for the positive results. *Did Sam know the crops would thrive?*

In the thought experiment above, Sam has a certain belief that his actions will affect the crops, and that belief is supported by relatively good evidence. Sam's belief is justified, but his belief is only true by luck. The target intuition of the Gettier case is that no matter the evidence or justification of the putative knower, Sam did not know that by pumping the water the crops would thrive. Typically, this intuition is taken to be unanimous by contemporary epistemologists.[62] The content of the intuition is then taken as obviously true, and imported as strong evidence against the position that knowledge consists of justification, truth, and belief alone.

But while a significant amount of theoretical work has been done regarding the best philosophical response to Gettier problems, very little has been said about the psychological origins of the intuition itself (see Lycan 2006).[63] More often, debates in epistemology regarding Gettier proceeded from the assumption that the intuition exists, to the conclusion that JTB analyses of knowledge must be false. Presumably though, an investigation of the mechanisms underlying these kinds of intuitions may help us gain insight into the Gettier intuition's evidentiary status, and in turn, deeper philosophical significance. Specifically, perhaps the study of Gettier cases can also benefit from the experimental techniques being used by experimental epistemologists to discover the factors involved in knowledge attribution.

Returning now to our previous discussion concerning the moral component of knowledge attribution, might normative evaluations be playing a role in typical Gettier case intuitions? In a recent article, John Turri (2010) considers that precise possibility. Citing the basic psychological effect found in ESEE, Turri suggests that one might be more willing to attribute knowledge that

[62] Though see evidence that this intuition may be culturally local (Buckwalter & Stich 2011; Weinberg 2001).
[63] In fact, see Zagzebski (1994) for the "inescapable nature" of Gettier problems.

p in Gettier conditions if the moral valence of *p* is manipulated. Specifically the suggestion is that people's prior moral judgments may lead to attributions of knowledge when Gettier agents are judged to be morally bad. This of course, would be very shocking, since it would mean that the moral component of knowledge attribution completely overturns the widely accepted verdict in epistemology that Gettier agents do not know that *p*. To put Turri's suggestion to the test, the following three experiments were conducted to investigate this question: can the epistemic side-effect effect be Gettier-ized?

4.3. GESEE Experiments

The following three sets of experiments are designed to provide initial evidence that the moral valence of a particular Gettier situation significantly influences ordinary knowledge attribution to Gettier subjects.

4.3.1. GESEE Experiment 1 (pump case)

To investigate the relationship between the moral component of knowledge attribution and Gettier intuitions let us begin by revisiting the above pump example. In a between-subjects experimental design, participants in Experiment 1 (N = 86, mean age = 36, 52 female, 87% Caucasian) were provided with one of two vignettes closely resembling cases originally used by Ulatowksi (2012) in the study of action individualization.[64] The two resulting cases are shown below:

[64] All studies reported were run online using Amazon Mturk and Qualtrics. Because Weinberg et al. (2001) found cross-cultural differences in epistemic judgments in prior work, the country of origin for participants was restricted to the United States.

> Sam's job is to pump water into the cistern, which then supplies the water to the farms owned by several families in the community. One day, as Sam operates the pump, he hears a broadcast on the radio. The radio report says that local officials suspect a new chemical from a nearby factory, chemical X, may have found its way into the local reservoir, and that there is a chance it will be very *[beneficial/ poisonous]* to all the local townspeople's crops. Sam thinks to himself, "I don't care about their crops; I just want to earn my pay," and continues pumping the water. Sure enough, the crops started *[thriving/dying]*. It turned out that the local officials were completely wrong about the chemical in the water. After analyzing the water, they found no trace of chemical X. Scientific reports later confirmed that the crops were all *[thriving/dying]* because of a fungus that had been secretly growing inside Sam's pump.

These stimuli describe the conditions of an agent who is indifferent towards bringing about either a good or a bad result. Typical of standard Gettier cases, the agent believes that the result will transpire, and that belief is supported by good evidence. While the belief is in fact true, the way the agent formed and justified the belief is only connected to a state of affairs in the world that makes the belief true by luck. Thus the standard prediction among epistemologists is that no knowledge will be ascribed in either case.

After seeing either the good result (the crops thrive) or bad result (the crops die) conditions of the vignette above, participants were asked to make judgments about the Gettier agent's knowledge. Specifically, participants were asked whether they agreed or disagreed with the statement, "Sam knew that by pumping the water, the townspeople's crops would [*thrive/die*]." Responses were collected on a 7-item scale anchored by positive and negative

agreement terms (1 = strongly disagree, 4 = neither agree nor disagree, 7 = strongly agree). For the hypothesis that knowledge attribution in Gettier conditions is sensitive to moral factors to be confirmed, we would expect people to make asymmetrical judgments depending on case valence.

This is exactly what was found. When pumping the water produced a good result (Help, $M = 3.05$, $SD = 1.59$), participants tended to give the standard philosophical answer that the agent did not have knowledge. However, participants said that Sam did know that his actions would bring about the particular result when that result happened to be bad (Harm, $M = 4.86$, $SD = 1.7$). In order to test the hypothesis that moral judgment affects knowledge attribution in Gettier cases, an independent samples t-test was conducted. As predicted, there is a significant difference between knowledge judgments made between these two groups, $t\ (84)= 5.04$, $p < 0.01$, $d = 1.10$. While in both Gettier cases the agent's epistemic position is the same (in that the relevant justified belief is only true by luck) the valence of the relevant action has an important impact on what Gettier agents are said to know. Such findings begin to suggest that moral judgments have a strong impact on people's intuitions about attributing knowledge in Gettier cases.

4.3.2. GESEE Experiment 2 (mayor case)

Of course, we would not want to draw conclusions about the general effect morality might have on Gettier judgments based only on a single experiment. After all, there may be a number of concerns about the specific stimuli used in Experiment 1. For instance, environmental harm is a very sensitive issue. So perhaps participants' personal views concerning the environment are somehow impacting their answers in this particular case, and do not reflect their actual Gettier judgments when moral factors become salient more generally.

So experiment 2 sought to replicate the same asymmetry found in the first experiment using a completely different case that did not involve mention of the environment. Participants (N = 78, mean age = 36, 46 women, 76% Caucasian) were presented with vignettes about a mayor of a small town who has formed the belief that taking a certain action will either create or cut jobs for members of the local community:

> The mayor of a small town is trying to decide whether or not to sign a new contract with a local corporation. The math is all very complex, but all his economic strategists think that there's a relatively good chance that one outcome is that it will [*create/cut*] jobs for workers in the community. The mayor says, "All I really care about is campaign contributions, not people's jobs, and I am sure to get millions from the corporation if I agree." So, he decides to sign their contract. The corporation, however, didn't take any chances. They secretly switched the contract with a totally different one right before the mayor signed it. By changing all the fine print, in some cases the opposite of what the mayor thought he was signing, the corporation could be sure it got what it wanted. Sure enough, shortly after the mayor signed the contract, a number of members of the community [*got/lost*] jobs, and the mayor received a huge donation to his reelection campaign.

Just like the pump case, the mayor has a strong justification for true beliefs. Then, Gettier information is added to the story so that the result that actually transpires has nothing to do with the mayor's evidence. As in Experiment 1, participants were given either the helping or harming version of this scenario and then asked if they agreed or disagreed that "the mayor knew that by

signing the contract he would [create/cut jobs]." Responses were again collected on the same seven-item scale anchored by positive and negative agreement terms.

Though means were slightly higher in Experiment 2, this result replicated the same basic finding as the pump cases.[65] Participants were more likely to say that the epistemic subject knew that the certain result would take place when that result was bad (Harm, $M = 6.05$, $SD = 0.94$) then when it was good (Help, $M = 4.11$, $SD = 1.86$). An independent samples t-test again shows this difference to be significant, $t\ (76) = 5.92$, $p < 0.01$, $d = 1.32$. These results suggest that the effect of moral valence on participants' epistemic judgments is not restricted to cases about the environment. The same pattern of responses appears when participants are presented with the mayor case.

4.3.3. GESEE Experiment 3 (third-person mayor case)

So far we have seen that the moral component of knowledge attribution plays a role in shaping intuitions about Gettier cases. But one immediate objection to these findings is that perhaps people only attribute knowledge in the pump and mayor vignettes because they are looking for a way to hold someone responsible for the outcome of these cases, and not because they believe those agents really have knowledge when the results are bad.

This idea is perhaps best illustrated with an old example from Saul Kripke.[66] Here, Kripke asks us to imagine that we have just heard the news about President Nixon's possible involvement in Watergate, and everyone is trying to figure out what Nixon knew about the

[65] As opposed to Experiment 1, mean judgment in the help cases of Experiment 2 were not significantly above or below the midpoint, suggesting that participants were largely unsure whether or not the Gettier agent knew.

[66] I am grateful to Keith DeRose and Michael McGlone for this reference to Saul Kripke's 1985 "Nozick bashing lectures" at Princeton.

cover-up. A trained epistemologist might tell us that technically Nixon did not have knowledge of the crimes because he *merely had a true belief* about the wrongdoings. Nevertheless few in the 1970s would consider this response satisfactory. Instead, we find ourselves wanting to attribute the relevant knowledge to Nixon. But our knowledge attribution to Nixon in this case would be driven—perhaps not by our properly attending to the justification Nixon had for his beliefs—but rather only as a means to justify our desire to hold him responsible for the Watergate scandal.

So the objection—what we might call the *Kripkean objection* to experiments 1 and 2—is that the data concerning the moral component only appear to reveal participants' real views about knowledge attribution. But instead, people are saying that the agents had knowledge only as a means to justify their desire to hold those agents morally responsible for the bad results. And in fact, as we saw in Section 2, researchers have argued that the effect that morality has been shown to have on the application of different non-moral concepts is not primarily due to people's underlying understanding of those concept, but rather due to this kind of distortion regarding the desire to blame (Alicke 2008).

While such explanations remain controversial, it might be thought that something similar could provide a straightforward explanation of the present data concerning knowledge ascription in Gettier cases seen in Experiment 1 and 2. Thus Experiment 3 was designed to put this idea to the test. Specifically, given the example above by Kripke, what we would like to know is whether or not people are attributing knowledge only as a means to hold agents responsible, and not because their underlying concept of knowledge is actually sensitive to the relevant moral evaluation.

To test whether people are genuinely attributing knowledge to Gettier agents independently of their desire to justify their blame judgments, participants in Experiment 3 (N =

85, mean age = 34, 50 women, 81% Caucasian) were presented with vignettes that very closely resembled the mayor cases from Experiment 2. The only difference was that they both included an additional character such that the epistemic agent in the story was not the same agent whose actions brought about the good or the bad result. This was accomplished by adding the following two sentences directly after those describing the mayor's indifference towards the community, but right before the Gettier information was presented in the previous versions:

> James the office secretary overheard everything, and is appalled by what the mayor said. Nonetheless, the mayor decides to sign the contract.

Given these third-person mayor cases, participants were then asked on the same seven-item scale as before whether they agreed or disagreed that "James the office secretary knew that members of the local community would [get/lose] jobs.[67]"

As predicted, when asked about James' knowledge, the same basic asymmetric pattern of knowledge attribution from Experiment 1 and Experiment 2 was found. Participants exhibited higher levels of agreement regarding James the office secretary's knowledge when the result of the case was bad (Harm, $M = 4.98$, $SD = 1.72$) than when it was good (Help, $M = 3.95$, $SD = 1.48$). But importantly, the mayor, and not James the office secretary, is to blame for causing the bad result.[68] We may even find ourselves empathizing with James as a victim of the mayor's

[67] We might note one important epistemic difference between this case and Experiment 2; perhaps James has less evidence than the mayor for either the good or the bad result.

[68] These findings do not rule out blame entirely. Nonetheless, this case still replicates the target effect when, presumably, the desire to blame is greatly minimized. These results are also consistent with previous research by Machery (2008) and Uttich & Lombrozo (2010) involving similar cases in which intentionality is ascribed to agents whose actions are not explicitly blameworthy.

appalling actions. Nonetheless, the effect of moral judgment on knowledge attribution in Gettier cases appears to persist even when the agent in question is not straightforwardly blameworthy. An independent t-test reveals a significant difference between these two groups t (83) = 2.94, $p <$ 0.01, d = 0.64.

Thus it seems unlikely that the effect reported across the three experiments above could be entirely mediated by the Kripke objection. Third-person mayor cases display a similar moral effect on knowledge attribution, even when participants are asked about the knowledge of a completely different agent than the one who brings about the bad results. But if there is little reason to hold James responsible for the layoffs, then it is doubtful whether or not a sole desire to justify blame judgments can entirely explain this effect. Of course, Experiment 3 does not conclusively prove that this distortion in epistemic concepts plays absolutely no role in people's judgments.[69] Nor does it rule out the possibility that some other kind of distortion theory can account for the present set of data. However, these results do suggest that further experimental evidence is required before dismissing the knowledge ascription found in Gettier cases out of hand. These results across experiments 1-3 are represented in Figure 4.1 below:

[69] In fact, the smaller effect size found in Experiment 3 might suggest that the desire to blame was playing at least some (though not exclusive) role in Experiments 1-2.

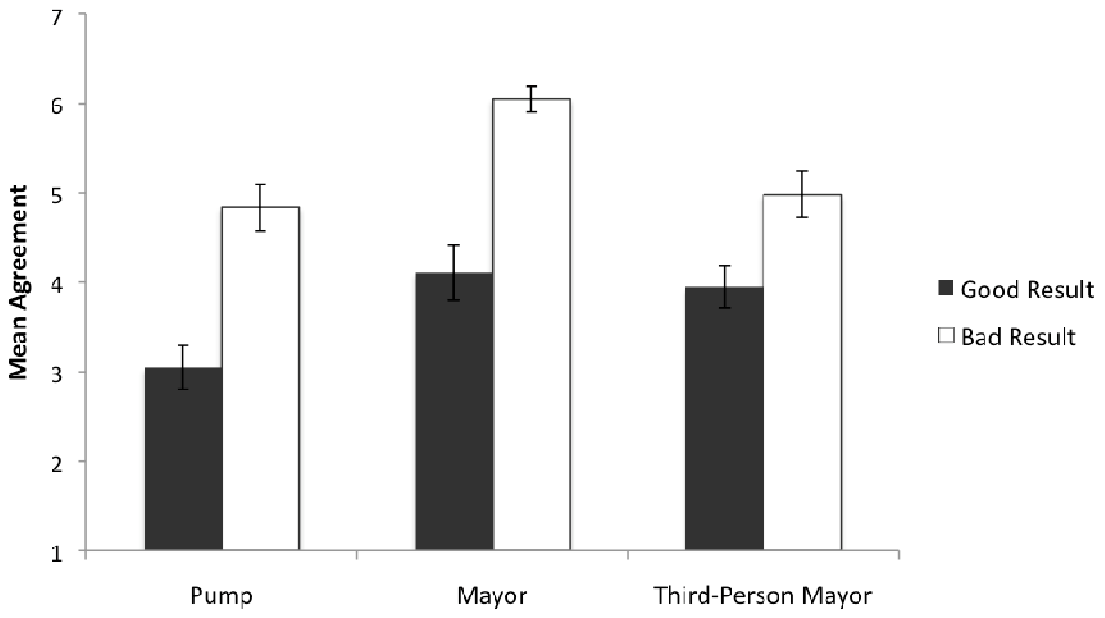

Figure 4.1. GESEE Results. Depicted are participants' ratings of agreement with knowledge attribution to Gettier agents in GESEE Experiments by outcome valence. All scales ran 1-7.

4.4. GESEE Discussion

Across three different types of vignettes we find this same basic effect. When Gettier cases are *made ESEE*, morality plays an important role in knowledge attribution to Gettier agents. The effect was found in cases that involved different contexts (the economy and the environment), different kinds of evidence (a radio report, a subordinate's testimony, and eavesdropping), and even persisted where epistemic evaluations were made about completely different agents (responsible or not responsible for bringing about the outcomes in question).

Expanding ESEE to GESEE also demonstrates something important about the original effect. Again, consider the chairman case. One deflationary hypothesis is that the reason why people attribute knowledge in the harm but not the help case is because of the chairman's

evidence. A general assumption often made about corporations is that they are more likely to cause environmental harm than good. So the same evidence that makes true beliefs count as knowledge in the harm condition, may be insufficient in the help condition. Perhaps then, ESEE judgments do not occur because knowledge attribution has a moral component, but simply because participants require agents to possess a greater amount of justification in the good cases in order to know. Yet the presence of the same effect in Gettier cases seems to rule out this hypothesis. It's true that in GESEE, Gettier agents in help and harm conditions across Experiments 1-3 have the same evidence. But in each case this evidence is completely disconnected from the state of affairs in the world that makes their beliefs true. It seems unlikely then that the effect in question arises because participants think knowing requires different amounts of justification across the various conditions. After all, according to the standard philosophical intuition, the very structure of a Gettier case is supposed to show that *no amount of justification* will lead a Getter agent to possess knowledge.

So expanding ESEE to Gettier case intuitions in the experiments above, and perhaps even in professional philosophers' own reflections on these cases, gives us even more evidence for the moral component of knowledge attribution.[70] However, the deeper question now remains to explain why moral evaluations are having this particular effect on knowledge attribution. As mentioned in section 2, there is extensive debate among researchers as to whether competence or performance-style explanations provide the best interpretations of the current findings. These are very difficult questions. As this research continues, no assumption is made presently concerning either the correctness of implementing competence or performance distinctions in experimental

[70] It remains an open empirical question whether or not professional philosophers will display the same sensitivities to moral features when attributing knowledge. However, see Weinberg et al. (2010) for evidence that philosophers are often just as sensitive as non-philosophers to many psychological effects when reflecting on thought experiments.

philosophy, or classifying the data in hand according to one of these two frameworks. However, supposing that these explanations might be extended to also account for knowledge judgments in Experiments 1-3, let's examine how the correctness of either approach might reveal important epistemological implications when it comes to Gettier intuitions.

Beginning with distortion-style explanations, suppose that people are attributing knowledge in Gettier because the presence of morality is somehow biasing or distorting the way they ordinarily make knowledge judgments.[71] Under this kind of explanation the attributions in the cases above can be seen as precisely the kind of performance error that properly controlled experimental methods should weed out. The fact that people attribute knowledge in some Gettier cases but not others leads to a tension to simply discard intuitions about knowledge in the vignettes with "bad" results. Then, even if these cases do tap into a heuristic we use concerning conditions of harm, one can still reject the content of the resulting intuition as informing philosophical theories about what one actually knows.

But such a finding also provides evidence that Gettier intuitions are *case-dependent*. And this result is not without precedent in epistemology. For instance, Gendler & Hawthorne (2005) argue for a similar conclusion regarding the purportedly standard philosophical intuitions in fake barn cases (see Goldman, 1976). By appraising a series of different fanciful barn case variations, their suggestion is that, "the concept of knowledge, prior to its being fashioned and molded by certain philosophical traditions, never offered any stable negative verdict in the original fake barn case" (p. 25).[72] Likewise, it looks as though GESEE provides the beginnings of some

[71] While Experiment 3 is somewhat suggestive, it has not of course, conclusively ruled out this hypothesis.

[72] Also see experimental evidence suggesting that the standard philosophical intuition in barn cases is not unanimous (Colaço, Buckwalter, & Stich, *manuscript*).

evidence that intuitions in Gettier cases may also vary from case to case. This time however, it is because of the distorting effect that *moral considerations* have on epistemic judgments.

But suppose that the experiments above really do point to a factor that systematically biases people's judgments in Gettier cases. If one thinks that GESEE results are due to some kind of distortion or error in performance, one ought to decrease credence in the default position that intuitions in Gettier cases really count as good evidence for or against traditional theories of knowledge. For not only are widely accepted Gettier intuitions easily overturned when moral considerations are added to the vignette. But furthermore, this effect prompts the following question. Just what other sorts of tacit factors *are* responsible for our judgments about knowledge ascription in these theoretically crucial cases? Without further research into this question, as well as the other potential biasing effects that may be present in Gettier judgments, these data begin to question the practice of using Gettier intuitions as decisive evidence for or against theories of knowledge.[73]

Alternatively, suppose that a conceptual competence-style explanation is correct. In which case, the presence of morality would not give rise to any additional cognitive process that biases people's judgments. Instead, knowledge attribution would be said to genuinely fluctuate in the Gettier cases above due to the underling concept of knowledge that people tacitly hold. Furthermore, suppose again that one endorses the popular position in epistemology that judgments of ordinary knowledge ascription should count as good evidence for a correct theory of knowledge. Data strongly suggest that the *salience of moral factors* can be crucial to

[73] Instead, the moral component of knowledge would join a chorus of ongoing research in experimental philosophy suggesting that intuitions are highly susceptible to a host of factors irrelevant to the content of the judgment in question, such as order (Swain et al., 2008), framing (Uhlman et al., 2009), and environmental effects (Helzer & Pizarro, 2011).

judgments people make about what Gettier agents know. Pace the standard philosophical intuition, people do think that Gettier agents have knowledge under these conditions.

These results appear to indicate that just like practical interests or the salience of error possibilities heavily discussed in connection with pragmatist or contextualist theories in epistemology, morality is an equally important factor that influences ordinary ascription. Thus if one thinks that theories of knowledge should include a correct account of the underlying ordinary concept, then one should adapt extant theories to do moral judgment justice. In other words, if a conceptual competence explanation turns out to be correct, then epistemologists attempting to capture the way knowledge is ordinarily ascribed can no longer afford to neglect the moral component of knowledge attribution in their epistemic theorizing.

As we have seen, the correctness of a competence or performance-style explanation leads to different recommendations regarding the moral component of knowledge in epistemic theory building. But now, continue to suppose that ordinary intuitions about 'knows' really do count as important sources of evidence about *knowledge*. Notice that by accepting this principle, either explanation yields a similar result about the typical use of the Gettier case intuition. Namely, the data suggest that one should be *much less confident* about what Gettier agents do and do not actually know. If it's distortion, one should be less confident simply because the evidence demonstrates the ease to which Gettier judgments can be manipulated. If it's competence, one might argue that if Gettier agents in harm conditions have knowledge, and Gettier agents in help conditions have the exact same evidence, then Gettier agents in help conditions must also have knowledge.[74] Of course, neither of these arguments conclusively shows Gettier agents have

[74] For detailed account of this kind of argument see Turri (2010).

knowledge. However, they should decrease credence in the widely accepted view that Gettier intuitions provide insurmountable evidence again JTB analyses of knowledge.

Lastly, the current findings may point to a potential avenue for future research in experimental philosophy regarding the competence and performance distinction. It has been argued that GESEE results have important implications for epistemology. But presumably, the correct psychological explanation for the moral component of knowledge attribution is likely to also apply to the moral effects cognitive scientists have found for other non-moral concepts like intentionality. Yet unlike intentionality, attributions of knowledge are not limited to first-person cases. Specifically, questions regarding what an agent intended always involve judgments about the agents who are in some way causally responsible for an action (e.g. the agent *who intentionally harms* the environment or *unintentionally creates* jobs). As we saw in Experiment 3, knowledge is different. Questions about what an agent *knows* can be asked independently of causal responsibility. Give that GESEE has been shown to persist in these kinds of third-person cases, pursuing experiments regarding the influence of morality in the epistemic domain may provide a promising new way to test competing explanations for these effects.

4.5. Conclusions for the Moral Component of Knowledge

We began with a discussion of a popular new approach in epistemology focusing on ordinary language practices. We saw that that experimental philosophers have discovered that ordinary knowledge attribution is sensitive to some pretty surprising factors. These factors include the influence of people's prior moral evaluations on their epistemic judgments. But current research has also shown the moral component of knowledge attribution does not only arise in the isolated cases of experimental philosophy, but are also extended to influence standard philosophical

intuitions in Gettier cases. And, Gettier case intuitions have been instrumental in philosophical argumentation. We have also seen that there has been debate about the best interpretation of the kind of effect shown in GESEE. This debate questions whether distortion or competence based accounts better explains morality's influence on knowledge attributions. Though disagreement continues with little end in sight, it has been argued that this debate is extremely relevant to epistemologists interested in providing epistemic theories in accord with the way people ordinarily attribute knowledge.

Epistemologists are faced with two equally important outcomes. If as Gettier case data seem to suggest, moral judgment does play an important role in people's underlying conceptual competences of knowledge, then this feature of folk judgments should be accounted for by epistemic views claiming to capture the way people ordinarily attribution knowledge. Conversely, if moral judgment is shown to systematically distort ordinary knowledge judgments in Gettier cases, then perhaps the intuition that Gettier agents do not have knowledge is simply not as strong of evidence for a particular analysis of knowledge as epistemologists have assumed. And either way, one should be much less confident when generally invoking the "Gettier intuition" as evidence. Lastly, the possibility of third person testing in the epistemic domain may provide a novel new approach for future experiments regarding competence and performance explanations.[75]

[75] Many thanks to Jonathan Adler, Jessie Prinz, Jonathan Schaffer, Stephen Stich and Yale Experimental Philosophy Lab for many helpful comments on previous drafts; and to Joshua Knobe, without whose continued support this work would not have been possible.

Chapter 5. Methodology in Epistemology

In Chapters 2 through 4 we have seen that experimental methods have been essential in discovering the factors underlying ordinary knowledge ascription. And in turn, understanding these factors has helped us shed light not only on the nature of the ordinary concept of knowledge that people tacitly hold, but also in addressing specific theoretical disputes between contextualists and advocates of interest-relative invariantism, or in giving us a fresh perspective on new avenues of research regarding Gettier intuitions. It seems experimental methods have been essential in proposing and testing specific theories of knowledge that make substantial claims appealing to commonsense judgment about knowledge, or in addressing perennial problems in epistemology like Gettier.

However, in addition to helping us learn more about the factors that influence mental state attributions, or adjudicating between extant epistemic theories, these data in experimental epistemology may also raise important methodological concerns about the use of empirically *un*-examined intuitions as evidence in epistemology. This is because factivity, error salience, accommodation, and morality are not the only factors discovered that influence our knowledge judgments. Indeed, intuitions—espoused by both philosophers and non-philosophers alike—have also been shown to be highly susceptible to both order effects and performance errors, as well as highly variable by demography. In this chapter we review these latter types of effects that experimental epistemologists have uncovered, and then pause to consider the important implications these effects may have for the use of untested armchair intuitions as evidence in epistemology.

5.1. Performance Errors

As we saw in Chapter 4, when attempting to isolate the factors that influence ordinary attribution of a concept, it is often difficult to separate those that reflect some feature of the concept at issue from those that are *performance* based, or reflect features that only serve to bias or distort the underlying concept.[76] While philosophers have drawn this distinction in a number of different ways, one factor that seems to fall straightforwardly into the "performance" category is effects that arise due to the *order* in which vignettes are presented.[77] However researchers have found that participants' epistemic intuitions can vary according to which thought experiments they read first (Swain et al. 2008).[78] In these experiments, participants were presented with a version of Lehrer's famous Truetemp case, often thought to represent an important objection to reliablism (1990):

> One day Charles was knocked out by a falling rock; as a result his brain was "rewired" so that he is always right whenever he estimates the temperature where he is. Charles is unaware that his brain has been altered in this way. A few weeks later, this brain rewiring leads him to believe that it is 71 degrees in his room. Apart from his estimation, he has no other reasons to think that it is 71 degrees. In fact, it is 71 degrees.

[76] For a related competence vs. performance debate concerning causal intuitions see Knobe 2010; Alicke et al. *forthcoming*. But also see Alexander et al. 2010a, b for a discussion of the difficulties faced by this kind of distinction.

[77] Contextualism may provide one important exception to this classification. These effects might not be considered performance errors if order is included under a particular contextualist theory as one of the factors said to *rightly* influence an ascriber's speech context.

[78] For similar order effects discovered in moral intuitions see Lombrozo, 2011; Liao et al. *forthcoming*. Order has also been shown to effect the way professional philosophers form moral intuitions (Schwitzgebel & Cushman *forthcoming*). Also see recent follow-up work to Swain et al. by Wright 2010.

What Swain et al. found was that attribution of knowledge to Charles varied greatly depending on what case participants had considered prior to reading about Charles. Participants were much less likely to agree that "Charles knows that it is 71 degrees" when the case was preceded by what they consider a clear case of knowledge, than when it was presented after a clear case in which the protagonist did not have knowledge. These authors argue that such results demonstrating that intuitions vary "according to factors irrelevant to the issues thought-experiments are designed to address" may provide a serious methodological challenge to those armchair philosophers who insist on relying solely on empirically unexamined intuitions.

Research concerning the role of performance-based factors in ordinary knowledge attribution is still in its infancy. However future work in epistemology may draw on the discovery of a number of similar effects by experimental philosophers in neighboring domains. In the experimental literature on judgments about free will, Schulz et al. (*forthcoming*) show that certain types of *personality traits* may predict the intuitions that philosophers have about moral responsibility. Gonnerman, Reuter, and Weinberg (2011) report that intuitions in similar sorts of cases are also significantly affected by *the font style* in which vignettes are presented. More generally, intuitions have also been shown to vary as a result of various *environmental factors* (Helzer and Pizarro, 2011). It remains to be seen if these kinds of effects will also influence people's judgments about knowledge specifically.

But neighboring work in experimental philosophy more generally suggests that the prognosis is grim. For instance, recent research by Tobia, Buckwalter, and Stich (2012) studying professional philosophers directly has found that both non-philosophers and philosophers are subject to a type of framing effect, the actor-observer bias, when making judgments of moral

permissibility and moral obligation. When both groups were presented with the famous "Jim and the Natives" case originally proposed by Smart and Williams (1978) or standard Trolley problems, TB&S found that the intuitive judgments both groups gave when they are the actor (or in response to cases framed in the first-person) differ from the intuitive judgments given when they are the observer (or cases framed in the third-person). And of course, whether an action in a moral scenario is framed in first or third person terms is almost always irrelevant to a moral judgment about the action in question.

TB&S use this result in a metaphilosophical argument against something called the "expertise defense" (see Williamson, 2005, Hales, 2006, Ludwig, 2007) or the idea that the intuitions of philosophers are more reliable than those of non-philosophers. Conversely, these data begin to point toward the opposite conclusion. It does not seem that training and ability has inculcated philosophers' moral intuitions from the sorts of irrelevant factors said to only influence the intuitions of philosophically unsophisticated experimental participants. Of course, advocates of the expertise defense need not, and typically do not, insist that philosophers' intuitions are entirely immune from framing effects or that they are never influenced by other factors that are irrelevant to the truth of the intuition. Their claim is only that philosophers' intuitions are *substantially less likely* to be influenced by such factors. But at the very least, such results begin to question whether or not this can be assumed without the need for more empirical evidence concerning the intuitions of professional philosophers.

For our present purposes, we need not engage directly in debates about the "expertise defense" since it is already widely accepted in epistemology that it is useful to draw on the ordinary concept or ordinary usage of 'knows' as evidence when engaged in first-order epistemic

debates.[79] Instead, what we might conclude from the TB&S data is that, assuming epistemic intuitions are sufficiently similar to moral intuitions, we should expect some of them, and perhaps a significant number of them from both philosophers and non-philosophers, to be highly sensitive framing effects. And of course, the only way to discover which intuitions in which cases are prone to display these effects is to study the epistemic judgments made by philosophers and non-philosophers in highly controlled and experimentally regimented settings.

5.2. Demographic Variation

In addition order effects and framing bias, experimental epistemologists have also discovered that epistemic intuitions are influenced by a number of different demographic variables. This research was largely inspired by the work of social psychologist Richard Nisbett and his collaborators who found differences between East Asian (Chinese, Japanese and Korean) and western (European Americans) participants in a wide range of areas including perceptual processing (Nisbett & Miyamoto 2005), and preferences for rule-based or association based reasoning strategies (Nisbett 2003; Norenzayan et al. 2002). In light of these important differences in cognitive processing, a number of experimental philosophers hypothesized that adult mental state attributions of knowledge might also vary from culture to culture (Weinberg et al. 2001; Nichols et al. 2003; Buckwalter & Stich 2011).[80]

In a series of well-known experiments, Weinberg et al. (2001) found cultural differences in a number of influential philosophical thought experiments including Truetemp, Gettier, and

[79] However its defenders should welcome the empirical method I am proposing, since only after we find out what the ordinary ascription practices are, can we make claims about whether the intuitions of trained philosophers are any different—or any better—than ordinary ones.
[80] For similar work on semantic intuitions cross-cultures see Machery et al. 2004, 2009.

Zebra (for more on zebra cases see Dretske 1970).[81] In one of these experiments they used the following Gettier case:

> 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?

They found that while Westerns typically gave the philosophical standard answer that Bob only believes that Jill drives an American car, the majority of East Asian participants said the opposite –East Asians were far more likely than Westerners to attribute knowledge to Bob. If this pattern of attribution is robust, it would seem to suggest that unlike philosophers, East Asians simply do not share the intuition that certain Gettier subjects lack knowledge.[82]

Recent research has also demonstrated that intuitions about knowledge attribution vary dramatically based one's *linguistic background*. In two experiments, Vaesen and Peterson (2011) found that in a number of cases, knowledge judgments of participants whose native language is English are significantly different from those of participants who are native Dutch, German, or Swedish speakers. Here is one of the vignettes they used:

---

[81] Weinberg et al. 2001 not only detects cultural variation in epistemic intuition among Indian, Asian and Western undergraduate students, but also significant variation according to *socioeconomic status*, suggesting that "high SES subjects accept much weaker knowledge-defeaters than low SES subjects."

[82] For some speculative challenges to these data see Nagel *forthcoming*; for replications of similar effects, see Buckwalter and Stich 2011.

> Boris asks his sister Steffi whether she knows the boiling point of water. Steffi, who has a Ph.D. in chemistry, answers truly: "Yes, I do. The boiling point of water is 100 degrees Celsius at sea level, which equals 212 degrees Fahrenheit."

What Vaesen & Peterson discovered is that Dutch speakers were much less likely than English speakers to agree that the information Steffi provided qualified as knowledge. Equally remarkable, is the fact that this effect was discovered in a population that consisted *entirely of formally trained philosophers*, suggesting that even expert intuitions are susceptible to the effect that linguistic background has on attributor judgments. Dutch speakers, it seems, are less likely to attribute knowledge in cases where the propositions are trivially true, or are part of common knowledge. As Vaesen & Peterson note when summarizing their research, such data may "cast doubt on the common armchair assumption that philosophical theories based on armchair intuitions are valid beyond the linguistic background against which they were developed."

In the same vein, recent work by Colaço, Buckwalter, and Stich (*manuscript*) suggests that intuitions in some central epistemic thought experiments might be variable across the lifespan from participants within certain linguistic backgrounds. In a series of studies, they presented members of the general public with versions of Alvin Goldman's famous "Fake Barn Cases" (1976). The cases they used all began like this:

> Gerald is driving through the countryside with his young son Andrew. Along the way he sees numerous objects and points them out to his son. "That's a cow, Andrew," Gerald says, "and that over there is a house where farmers live". Gerald has no doubt about what the objects are. What Gerald and Andrew do not realize

> is the area they are driving through was recently hit by a very serious tornado. This tornado did not harm any of the animals, but did destroy most buildings. In an effort to maintain the rural area's tourist industry, local townspeople built house façades in the place of destroyed houses. These façades look exactly like real houses from the road, but are only for looks and cannot be used as actual housing.

Then, in order to test how the presence of epistemic defeaters (a factor frequently discussed in connection with barn cases in epistemology) influenced people's judgments, half of the participants were given a low defeater ending to the vignettes that looked like this:

> Having just entered the tornado-ravaged area, Gerald has not yet encountered any house façades. When he tells Andrew "That's a house" the object he sees and points at is a real house that has survived the tornado.

The other half of the participants were given high defeater vignettes which ended like this:

> Though he has only recently entered the tornado-ravaged area, Gerald has already encountered a large number of house façades. However, when he tells Andrew "That's a house", the object he sees and points at is a real house that has survived the tornado.

What CB&S found is that most people considered the barn cases to be cases of knowledge, where the number of defeaters played no role in their ascriptions of knowledge. However, this intuition that fake-barn cases count as knowledge was negatively correlated with age; older participants are less likely than younger participants to attribute knowledge in fake-barn cases. In other words, they found that younger people in their experiments were are far less sensitive to putative epistemic differences between what many epistemologists might consider a clear case of knowledge, and a fake barn case. Participants under 30 exhibited the tendency to treat fake barn cases in almost the exact same way that they treated obvious instances of knowledge.[83] (In fact, younger people seem to have little or no problem counting a fake-barn case as a genuine case of knowledge, regardless of the number of epistemic defeaters in CB&S' materials). Theses results are visualized in the figure below from CB&S, comparing participant scores in the conditions that constituted clear cases of knowledge, the "Combined Cow (Control)" conditions, and the conditions constituting the traditional barn cases, the "Combined House (Test)" grouped by age:

[83] This age was chosen by TB&S because it divided the sample equally in half. Interestingly enough however, this "over 30" category seems to include approximately all authors (until very recently) publishing papers in epistemology on this issue. Thus we might expect to find this result broadly consistent with what one might informally observe in the sample of fake-barn judgments reflected in the epistemology literature.

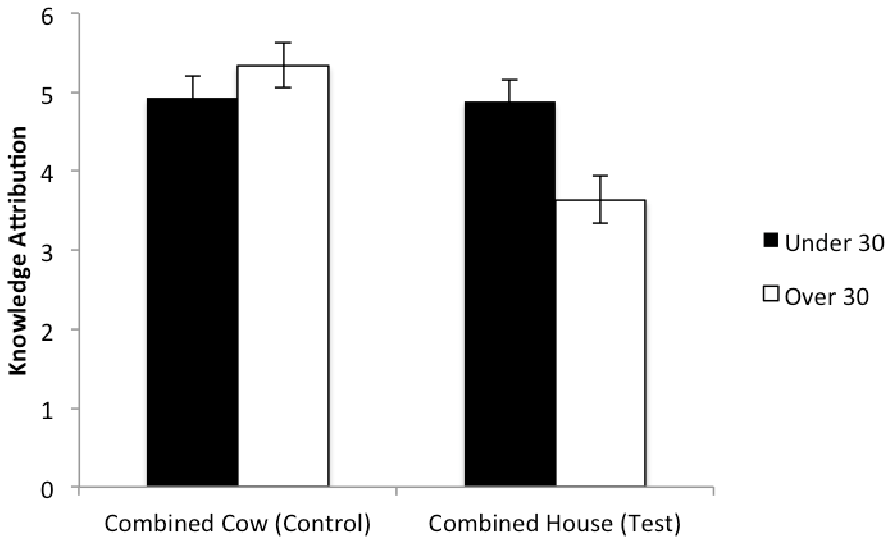

Figure 5.1. Knowledge ratings in barn experiments. Figure represents mean scores in Combined Cow and Combined House by age group. Scales ran from 0 to 6.

Of course, whether similar effects occur with other philosophically important thought experiments is an intriguing topic for further research. But whatever the answer may be, it's clear that philosophers should take note of the correlation between age and attribution of knowledge when they are tempted to employ fake-barn cases in epistemic arguments, and beyond. Or as CB&S put the point, "even if most philosophers have the intuition that fake-barn cases are not instances of knowledge, it is quite possible their students do not!" (11). But again, the only way to know which intuitions are stable across the lifespan, and which judgments about knowledge vary by factors like age, is to look.[84]

[84] Experimental researchers should also take note of these differences, and if robust, begin to take steps towards controlling for and recruiting participants equally distributed across the lifespan when conducting similar research.

5.3. No Turning Back Now

The evidence in experimental epistemology discussed in this chapter has suggested that important epistemic intuitions may be variable *across persons* due to effects of age, culture, and native language. We have also seen that certain epistemic intuitions may be unstable *among persons* due to our susceptibility to a series of cognitive biases like order effects and framing. And lastly, we saw that both sorts of effects reviewed in Sections 5.1-5.2 are not mediated by professional philosophical training. At the very least then, it looks as though we have good reason to expect some of the epistemic intuitions espoused by professionals to display these effects. Before continuing, it should be emphasized again that much more experimental work is needed to ascertain the regularity and robustness of demographic variation in epistemic intuition. Until then, we will not know the precise degree or conditions of such variation. However, this is just to embrace the prescribed method of experimentation in answering these questions.

But that said, how do data suggesting at least prima facie evidence for demographic differences and performance variable effects in our epistemic intuitions affect the practice of using intuitions as evidence in shaping a correct epistemological analysis? The answer, of course, is quite complex, and will depend on the *particular type* of philosophical project with regards to epistemology in which we are engaged. For instance, one thing we may argue from the data is that to the extent to which different demographic groups evince different epistemic intuitions, they also have different epistemologies. So if the philosophical project in which such intuitions are being used is primarily conceptual analysis—whether that be an investigation of what is literally in participants' heads, a characterization of some personal or group state, an analysis of language, or some other favored view about what concepts amount to—then the intuitions garnered from the present studies might serve as rather *good evidence* in such debates.

One might argue that these data begin to show, for instance, that native and non-native Dutch philosophers just hold different tacit concepts of knowledge. In this way the demographic differences data do not undermine projects in epistemology that seek to provide a *conceptual analysis* of the concept of knowledge that certain groups of people tacitly hold.

However, epistemologists are typically not only concerned with the descriptive project of giving a conceptual analysis of knowledge any particular participant or cultural group might share. Instead, a large number of contemporary epistemological projects aim at achieving one of the following non-mutually exclusive goals, which take intuitions as evidence for (1) analyses of knowledge that do not depend on anyone's concepts, but rather characterizes the philosophical phenomenon of *knowledge* itself, (2) normative analyses of how epistemic agents ought to behave, in this case how people should attribute knowledge, and (3) a combination, conceptual analyses of knowledge that deliver conclusions with such metaphysical and normative weight. Following Stich (2008), Weinberg et al. (2001), and Nichols et al. (2003), the question now becomes, do epistemic intuitions also serve as *good evidence* when imported into epistemological arguments whereby the philosophical goals within epistemology are to discover something like the essence of *knowledge*, or when the project of a given epistemological analysis is decidedly normative?[85]

The structure of arguments (1-3) in question are widespread in analytic epistemology, those that take our intuitions as sole evidentiary inputs without argument, and then release metaphysical or normative conclusions about knowledge as outputs.[86] While we might question

[85] See, for example, the debate between Stich (2008) and Sosa (2008) for further discussion about the use of intuitions in analytic epistemology (Williamson 2007) given the evidence that such intuitions are culturally relative.
[86] See Alvin Goldman's *Epistemology and Cognition* (1986) for perhaps the most important example of this kind of project.

these kinds of arguments for lots of different reasons, we are concerned presently with how the multiplicity of inputs in terms of demographic variation bears on the efficacy of such outputs. As we have seen, participants at least appear to have different intuitions in responding to the same epistemically motivated questions depending on their native language, culture, age, etc. Since the input intuitions between groups are significantly different, we have good reason to believe that any conclusions formulated and supported on the basis of such intuitional evidence will also be radically different between groups. Similarly, given intuitional variation, any normative positions generated from these kinds of arguments will also differ radically.

Given these two specific aims in epistemology, where different intuitions yield different evidence for a respective metaphysical or normative conclusion, both groups cannot be right. If the content of intuitions observed in the experiments above associated with native Dutch speakers are right, then those associated with the non-native Dutch speakers are wrong (and vice-versa). It seems relatively clear that in the cases displaying cognitive bias, we have independent criteria for filtering out judgments likely to be false. Yet, what independent argument could there be to privilege the intuitional responses detected from one demographic group in these epistemological thought experiments over the others—in that they better characterize *knowledge*, or generate more virtuous normative claims about the relevant epistemic principles? While there surely may be various different reasons why one might prefer the different metaphysical or normative conclusions a particular intuition generates over another, empirical evidence shows that reliance on intuitions from novel cases in the usual way cannot alone buttress such arguments. Instead, given the variation in epistemic intuition detected here, the burden of proof

rests on those epistemologists who rely on such intuitional evidence to provide other compelling reasons to support a favored analysis.[87]

In the face of such potential variation, if there is no philosophical reason to think that the content of the intuitions from one group are more likely to be true than another, then it is difficult to see why there is any substantive philosophical reason to take one rival intuition as evidence over another.[88] Since such a process would rely heavily on demographic variation, such differences seem to belie genuine metaphysical or normative force of the conclusions we reach on the basis of these kinds of evidence. Lack of consensus detracts from the evidentiary status of those intuitions in deciding the respective conditions for knowledge. That is, such divergence in epistemic intuition questions whether *these* intuitions in *these* cases as good evidence for or against *these* particular views, insofar as the conclusions generated from these intuitions are meant to reveal some metaphysical or normative truth about knowledge.[89]

So which intuitions are specifically vulnerable to bias or variable to demography, and why? The answer to these questions is that we do not know. How often, and in what circumstances, will a philosopher's epistemic intuitions be moderated by her age or native language, and when will they be more likely to generalize across philosophers? The answer is also that we currently do not know. In the words of David Lewis, even "Blind Freddy can see where this will lead" (1996, 549). Not only has it been *useful* to apply the techniques of

[87] Such reasons should explain why so many people in one group typically go "right" while so many other people typically go "wrong." Of course, if an error theory is given in terms of pragmatics, it should be sure to also tell us why those pragmatic features disproportionately affect different groups, and how those features are not themselves a relevant part of that group's epistemology.

[88] Though an open question, see the following papers for compelling arguments that appeal to "philosophical expertise" is not going to be a good reason to prefer philosopher's intuitions over those of non-philosophers in these cases (Weinberg et al. 2010, Machery 2011).

[89] That is, the native language of the philosopher who espouses *p* is typically a factor thought to be irrelevant to the metaphysical truth of *p*.

experimental epistemology as seen in Chapters 2-4, but also it appears as though we cannot afford *not* to apply these techniques. In the absence of further experimentation we cannot discover which intuitions are shared and which diverge between demographic groups, or which are subject to bias. Until then, we simply cannot decide which intuitions are unanimous or free from bias from the armchair, and therefore which epistemological projects are supported by questionable intuitional or ordinary usage data.[90]

[90] And it's perfectly possible that the majority of intuitions are unanimous, we just don't know which until more experimental philosophy is done.

Chapter 6. Conclusions and Future Study

We began in Chapter 1 with what I called the core insight in contemporary epistemology. In the old days, it was typical for philosophers to make claims about their intuitions to support the analyses of knowledge they gave. The ones that were shared by the most philosophers pointed out the clear winner. But as we have seen, given the renewed focus on knowledge ascriptions, the principle thesis of this dissertation has been that to find out who the winners are today—in arguments that make substantial and important use of information about ordinary knowledge judgments, processes and behavior—we simply require empirical data. Only careful experimentation can resolve whether or not the ordinary concept of knowledge is factive, even if it may sometimes appear otherwise. Only these kinds of data can tell us whether stakes, salience, or accommodation influences knowledge ascription, and subsequently, adjudicate disagreement between theories that claim to account for ordinary uses of 'knows' and people's underlying concepts of knowledge that ostensibly support this usage. Only experimental evidence can uncover the moral component for knowledge, reveal performance errors, or demographic variation in our epistemic intuitions. These advances have shown how the methods advocated for here can directly move contemporary debates in epistemology forward in an empirically informed way. And as we have seen, to draw on these factors as evidence for a theory of knowledge, or to debate the further epistemic significance of these factors, one must first understand them. For this, I have argued, experimental epistemology is here to stay.

But in addition to and beyond just the bearing on these particular debates, we have also contributed to a greater scientific understanding the ordinary semantics of 'know' and the concept of knowledge that people hold. In short, we have contributed to a general understanding of who we are qua epistemic agents. In demonstrating the importance of understanding the

psychological factors that underlie epistemic judgment, the evidence presented in the last three chapters has pointed towards a number of important effects influencing the conditions under which we consider others to know: the large impact of *factivity, error salience, accommodation,* and *morality* on ascription practices. And we also saw that although much more research needs to be done than is presently available, sometimes ascription can potentially be quite susceptible to *order effects,* and *demographic variation*. Due to this research, and as these various experimental programs continue to move forward, epistemologists as well as psychologists and cognitive scientists are left with a growing series of important questions apt for future research. New findings in each of the areas presented will undoubtedly continue to advance our understanding of these factors. Future work may focus on the precise role of *attributor erro*r or normative sensitivity in knowledge and its relation to mental state attribution more generally in cognitive science. Alternatively, individual differences psychology may add knowledge ascription to the list of potential new areas of study, along with the robustness of demographic variation in knowledge behaviors, worries about ecological validity across these studies, and our degree of susceptibility to framing, order and environmental effects on epistemic intuitions.

Another area of potential progress involves learning more about what explains the effects that have been discovered that cue ascription, and in turn, more about the deeper, normative value of those ascriptions. For instance, as we saw in Chapter 4, there is currently a growing debate about the best explanation of the influence of normativity, or moral judgment, on knowledge attribution. Some have argued that the best explanation may be due to the way normatively cues counterfactual reasoning about relevant alternatives more generally (Schaffer and Knobe 2010; Knobe 2010). It is an interesting further question for future work whether all the effects we have presently uncovered, like factivity or error salience could also fit under such

an explanation. If true, this may lead to further developments concerning the relationship between knowledge ascription and the vast literature in psychology on the rational imagination, or how people rely on counterfactuals and possibilities when making inferences, plans and goals (see Byrne 2005). Alternatively, if error salience or normativity in knowledge ascription is the result of exclusively system 1 processes, as discussed in Chapters 3-4, then the study of knowledge ascription may constitute and interesting addition to the fast and frugal heuristic program (see Gigerenzer, Hertwig, and Pachur 2011) in larger discussions of bias and bounded rationality.

While more research is required, these potential new avenues of study mark exciting ways in which the study of knowledge ascription in epistemology can be extended to larger debates within psychology and cognitive science. But until then, and in the meantime, if this dissertation has been successful, then you will have begun to see that the empirical methods advocated here need to become an integral and widely accepted part of research in epistemology in order to move forward—not only for debates in epistemology, but for also advancing our understanding of the semantics of 'know', and how the ordinary concept of knowledge can fit within a larger theoretical framework of the study of mental states and human decision making. The hope is that approaching these empirical and theoretical questions in tandem may lead to significant and lasting progress upon which to build a new study of knowledge.

Appendix

Projection Experiment 1 Materials

1. **Everyone knew** that stress caused ulcers, before two Australian doctors in the early 1980s proved that ulcers are actually caused by bacterial infection.

> Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:
> A) Everyone thought they knew
> B) Everyone really did know

2. He figures that anything big enough to sink the ship will be seen in time to make a turn. But when he tried to turn the ship the rudder was too small. What **he knows** about sailing is wrong.

> Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:
> A) What he thought he knew
> B) What he really did know

3. In school **we learned** that World War I was a war to 'make the world safe for democracy,' when it was really a war to make the world safe for the Western imperial powers.

> Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:
> A) They thought they learned
> B) They really did learn

3. Jan told us that during her trip last year, she had trouble breathing, sharp pains in her side, several broken ribs and a partially collapsed lung. She was in the middle of nowhere without any real rescue assets and that was when '**she realized** she was going to die out there.'

> Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:
> A) Jan thought she realized
> B) Jan really did realize

Projection Experiment 2 Materials

1. Two biologists are having a discussion about a crab:

> Dr. Krycek: "Can we be sure that this one is of the genus *Calcinus*?"
> Dr. Scully: "I have my notes right here. **I know** that this is a specimen of *Calcinus hazletti*"

But this crab was actually a completely different crab, *Calcinus obscurus*.

Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:
A) Dr. Scully thought she knew
B) Dr. Scully really did know

2. Two biologists are having a discussion about a crab:

Dr. Krycek: "Can we be sure that this one is of the genus *Calcinus*?"
Dr. Scully: "I have my notes right here. **I learned** that this is a specimen of *Calcinus hazletti*"

But this crab was actually a completely different crab, *Calcinus obscurus*.

Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:

A) Dr. Scully thought she learned
B) Dr. Scully really did learn

3. Two biologists are having a discussion about a crab:

Dr. Krycek: "Can we be sure that this one is of the genus *Calcinus*?"
Dr. Scully: "I have my notes right here. **I believe** that this is a specimen of *Calcinus hazletti*"

But this crab was actually a completely different crab, *Calcinus obscurus*.

Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:

A) Dr. Scully thought she believed
B) Dr. Scully really did believe

4. Two biologists are having a discussion about a crab:

Dr. Krycek: "How did you figure out that this one is of the genus *Calcinus*?"
Dr. Scully: "I was just rereading my notes from class when **I realized** that this is definitely a specimen of *Calcinus hazletti*"

But this crab was actually a completely different crab, *Calcinus obscurus*.

Which of the following do you think best **describes** what is meant in the **BOLD portion** of the above sentence:

A) Dr. Scully thought she realized
B) Dr. Scully really did realize

Bibliography

Adams, F., Steadman, A. (2004). Intentional action in ordinary language: core concept or pragmatic understanding? *Analysis*, 64, 173-81.

Alexander, J., Mallon, R., & Weinberg, J. (2010a). Accentuate the Negative. *Review of Philosophy and Psychology*, 1, 297–314.

---., ---., & ---. (2010b). Competence: What's In? What's Out? Who Knows?, *Behavioral and Brain Sciences*, 33, 329–30.

Alicke, M., D. (2008). Blaming badly. *Journal of Cognition and Culture*, 8, 179-186.

Alicke, M., Rose, D., & Bloom, D. Causation, Norm Violation and Culpable Control. Forthcoming in *Journal of Philosophy*.

Austin, J.L. (1956). A plea for excuses. *Proceedings of the Aristotelian Society*, new series, 57, 1–30.

Beebe, J. R., & Buckwalter, W. (2010). The Epistemic Side-Effect Effect. *Mind & Language*, 25, 474-498.

Beebe, J.R., & Jensen, R. Surprising Connections Between Knowledge and Action: The Robustness of the Epistemic Side-Effect Effect. Forthcoming in *Philosophical Psychology*.

Birch, S. A. J. (2005). When knowledge is a curse: Children's and adults' reasoning about mental states. *Current Directions in Psychological Science*, 14, 25-29.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18, 382-386.

Brown, J. (2011). Experimental Philosophy, Contextualism and SSI. *Philosophy and Phenomenological Research*, Online Publication, 12 JAN 2011 DOI: 10.1111/j.1933-1592.2010.00461.x.

Byrne, R. M. J. (2005). The rational imagination: How people create alternatives to reality. MIT Press.

Buckwalter, W. (2010). Knowledge Isn't Closed on Saturdays. *Review of Philosophy and Psychology*, 1, 395-406.

---. (2012). Non-Traditional Factors in Judgments About Knowledge. *Philosophy Compass*, 7/4, 278–289.

---. The Mystery of Stakes and Error in Ascriber Intuitions. Forthcoming in J. Beebe, (ed.) *Advances in Experimental Epistemology*, Continuum.

---., & Stich, S. (2011). Epistemology and Demography. Invited presentation at the Pre-

Conference Workshop in Experimental Epistemology. American Philosophical Association Pacific Meeting. 20 April 2011.
---., & Schaffer, J. Knowledge, Stakes, and Mistakes. Unpublished manuscript. Rutgers University.

Chatterjee, S., Heath, T.B., & Min, J. (2009). The susceptibility of mental accounting principles to evaluation mode effects. *Journal of Behavioral Decision Making*, 22, 120–137.

Chisholm, R. (1977). *Theory of Knowledge*. Englewood Cliffs: Prentice Hall

Cohen, S. (1988). How to be a Fallibilist. *Philosophical Perspectives*, 2, 91-123.
---. (1999). Contextualism, Skepticism, and the Structure of Reasons, *Philosophical Perspectives*, 13, 57-89.
---. (2004). Knowledge, Assertion, and Practical Reasoning, *Philosophical Issues*, 14, 482–491.

Colaço, D., Buckwalter, W., & Stich, S. Epistemic Intuitions in Fake-Barn Thought Experiments. Unpublished manuscript, Rutgers University.

DeRose, K. (1992). Contextualism and Knowledge Attributions, *Philosophy and Phenomenological Research*, 52, 913-29.
---. (1999). Contextualism: An Explanation and Defense, *The Blackwell Guide to Epistemology*, eds. John Greco and Ernest Sosa: 182-205. Basil Blackwell.
---. (2005). The Ordinary Language Basis for Contextualism and the New Invariantism, *Philosophical Quarterly*, 55, 172-98.
---. (2009). *The case for Contextualism.* Oxford, Oxford University Press.
---. (2011). Contextualism, Contrastivism, and X-Phi Surveys, *Philosophical* Studies 156:1 81-110.

Dretske, F. (1970). Epistemic Operators. *Journal of Philosophy* 67, 1007-1023.

Fantl, J., & McGrath, M. (2002). Evidence, Pragmatics, and Justification, *The Philosophical Review*, 111, 67-94.
---., & ---. (2010). *Knowledge in an Uncertain World*. Oxford: Oxford University Press.

Feltz, A., & Zarpentine, C. (2010). Do You Know More When it Matters Less? *Philosophical Psychology*, 23, 5, 683–706.

Fischer, P., Jonas, E., Frey, D., & Kastenmüller, A. (2008). Selective exposure and decision framing: The impact of gain and loss framing on confirmatory information search after decisions. *Journal of Experimental Social Psychology*, 44, 312-320.

Gendler, T. S., & Hawthorne, J. (2005). A Real Guide to Fake Barns: A Catalogue of Gifts for your Epistemic Enemies. *Philosophical Studies*, 124, 331–352.

Gettier, E. (1963). Is Justified true belief knowledge? *Analysis*, 23, 6, 121–3.

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). Heuristics: The foundations of adaptive behavior. New York: Oxford University Press

Goldman, A. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy,* LXXIII: 20, 771–791.
---. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Gonnerman, C., Reuter, S., & Weinberg, J., M. More Sensitive Intuitions: Print Fonts, Could Choose Otherwise, and Free Will. Society for Philosophy and Psychology and European Society for Philosophy and Psychology, 6-10 July 2011.

Guglielmo, S., & Bertram, M. F. (2010). The timing of blame and intentionality: testing the moral bias hypothesis. *Personality and Social Psychology Bulletin,* 36, 1635-1647.

Grice, H.P. (1989). *Studies in the Way of Words,* Harvard University Press.

Hales, S. D. (2006). *Relativism and the foundations of philosophy*, Cambridge, MA: MIT Press.

Halpern J, Hitchcock C. (2011). Actual causation and the art of modeling. In R. Dechter, H. Geffner, & J. Halpern (Eds.) *Heuristics, Probability, and Causality: A Tribute to Judea Pearl* (pp. 383-406). London: College Publ.

Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research,* 80, 497-522.
---. (2012). Factive Presupposition and the Truth Condition on Knowledge. *Acta Analytica* Online First: http://dx.doi.org/10.1007/s12136-012-0163-3.

Hawthorne, J. (2004). *Knowledge and Lotteries*. Oxford: Oxford University Press.
---., & Stanley, J. (2008). Knowledge and action. *Journal of Philosophy,* 105, 571-590.

Helzer, E., & Pizarro, D. A. (2011). Dirty liberals!: reminders of cleanliness promote conservative political and moral attitudes. *Psychological Science*, 22, 517–22.

Holton, R. (1997). Some Telling Examples: A reply to Tsohatzidis. *Journal of Pragmatics,* 28, 624-28.

Hsee, C.K. & Zhang, J. (2004). Distinction bias: misprediction and mischoice due to joint evaluation, *J. Pers. Soc. Psychol,* 86, 680–695.
---., & ---. (2010). General evaluability theory, *Perspectives on Psychological Science*, 5, 343-355.

Knobe, J. (2003). Intentional Action and Side Effects in Ordinary Language. *Analysis,* 63, 190–4.
---. (2004). Intention, intentional action and moral considerations. *Analysis*, 64, 181-187
---. (2010). Person as scientist, person as moralist. *Behav. Brain Sci,* 33, 315-29.
---., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H., & Sommers, T. (2012).

Experimental Philosophy. *Annual Review of Psychology*, 63, 81-99.

Lehrer, K. (1990). *Theory of Knowledge*. Westview Press.

Lewis, D. (1979). Scorekeeping in a Language Game, *Journal of Philosophical Logic*, 8, 339–59.
---. (1996). Elusive Knowledge, *Australasian Journal of Philosophy*, 74, 549-567.

Liao, M., Wiegmann, A., Alexander, J., and Vong, G. Putting the Trolley in Order: Experimental Philosophy and the Loop Case. Forthcoming in *Philosophical Psychology*.

Lombrozo, T. (2011). The role of moral commitments in moral judgment. *Cognitive Science, 33,* 273-286.

Ludwig, K. (2007). The epistemology of thought experiments: First person versus third person approaches. *Midwest Studies in Philosophy*, *31*, 128-159.

Lycan, W. (2006). On the Gettier problem problem. In S. Hetherington (Ed.), *Epistemology Futures* (pp. 148-168). Oxford: Oxford University Press.

Machery, E. (2008). Understanding the folk concept of intentional action: philosophical and experimental issues. *Mind & Language*, 23, 165-189.
---. (2011). Thought Experiments and Philosophical Knowledge. *Metaphilosophy,* 42, 191-214.
---., R. Mallon, S. Nichols and S. Stich. (2004). Semantics, cross-cultural style, *Cognition*, 92, B1–B12.
---., Olivola, C. Y., De Blanc, M. (2009). Linguistic and metalinguistic intuitions in the philosophy of language, *Analysis*, 69, 689–694.

Malle, B. F., & Nelson S. E. (2003). Judging mens rea: the tension between folk concepts and legal concepts of intentionality. *Behav. Sci. Law*, 21, 563-80.

Marshall B., J., Armstrong, J. A., McGechie, D.,B., & Glancy, R.J. (1985). Attempt to fulfill Koch's postulates for pyloric campylobacter. *Medical Journal of Australia,* 142, 436-439.

May, J., Sinnott-Armstrong, W., Hull, J. G., & Zimmerman, A. (2010). Practical Interests, Relevant Alternatives, and Knowledge Attributions: An Empirical Study. *Review of Philosophy and Psychology*, 1, 265-273.

Mayseless, O., & Kruglanski, A.W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. *Organizational Behavior and Human Decision Processes*, 39, 162-183.

Myers-Schulz, B. & Schwitzgebel, E. Knowing that P without Believing that P. Forthcoming in *Nous*.

Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: some problems

for jury impartiality. *Philos. Explorations*, 9, 203-20.

Nagel, J. (2010). Knowledge Ascriptions and the Psychological Consequences of Thinking about Error. *The Philosophical Quarterly* 60, 286-306.
---. Intuitions and Experiments: A Defense of the Case Method in Epistemology. Forthcoming in *Philosophy and Phenomenological Research*.

Nichols, S., Stich, S. & Weinberg, J.M. (2003). Metaskepticism: Meditations in Ethno-Epistemology. *The Skeptics*. Ed. S. Luper. Ashgate Press, 227-232.
---., & Ulatowski, J. (2007). Intuitions and individual differences: the Knobe effect revisited. *Mind and Language,* 22, 346-65.

Nisbett, R. E. (2003). *The Geography of Thought : How Asians and Westerners Think Differently...and Why.* New York: Free Press.
---., & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in Cognitive Sciences,* 9, 467-473.

Nolan, D. (2008). Non-Factivity About Knowledge: A Defensive Move. *The Reasoner*, 2.11, 6-7.

Norenzayan, A., Smith, E., & Kim, B. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science,* 26, 653-684.

Phelan, M. (2010). The inadequacy of paraphrase is the dogma of metaphor. *Pacific Philosophical Quarterly*, 91, 481-506.
---. (2011). The Intentional Action Factory. *The Philosopher's Magazine*, 52, 72-77.
---. Evidence that Stakes Don't Matter for Evidence. Unpublished manuscript, Lawrence University.
---., Arico, A., & Nichols, S. Thinking Things and Feeling Things: On an Alleged Discontinuity in Folk Metaphysics of Mind. Unpublished manuscript. University of Arizona.

Plato, ., Cooper, J. M., & Hutchinson, D. S. (1997). *Complete works*. Indianapolis, Ind: Hackett Pub.

Pinillos, N, Á. (2011). Some Recent work in Experimental Epistemology. *Philosophy Compass*, 10, 675–688.
---. (2012). Knowledge, Experiments and Practical Interests. in *New Essays On Knowledge Ascriptions* (Eds. Jessica Brown and Mikkel Gerken) Oxford University Press.

Rose, D., & Schaffer, J. Knowledge entails dispositional belief. Forthcoming in *Philosophical Studies*.

Ross, W.D., (1984). *The Complete Works of Aristotle*. The Revised Oxford Translation, vol. 2, Jonathan Barnes, ed., Princeton: Princeton University Press.

Schwitzgebel, E. & Cushman, F.A. Expertise in moral reasoning? Order effects on moral

judgment in professional philosophers and non philosophers. Unpublished manuscript, University of California Riverside.

Schaffer, J. (2005). Contrastive Knowledge. *Oxford Studies in Epistemology*, 1, 235-71.
---. (2006). The irrelevance of the subject: against subject-sensitive invariantism. *Philosophical Studies*, 127, 87–107.
---., & Knobe, K. (2010). Contrastivism Surveyed. *Nous*. doi: 10.1111⁄j.1468-0068.2010.00795.x. [Online]. Retrieved on 12 January 2011 from: http://onlinelibrary.wiley.com⁄doi⁄10.1111⁄j.1468-0068. 2010.00795.x⁄abstract.

Schulz, E., Cokely, E., & Feltz, A. Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. Forthcoming in *Consciousness and Cognition*.

Schwitzgebel, E. & Cushman, F. Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. Forthcoming in *Mind & Language*.

Shope, R., K. (1983). *The Analysis of Knowing. A Decade of Research*. Princeton: Princeton University Press.

Smart, J.J.C. & Williams, B. (1973). *Utilitarianism: For and Against*. Cambridge University Press.

Smith, S. M., Fabrigar, L, R., & Norris., M.E. (2008). Reflecting on Six Decades of Selective Exposure Research: Progress, Challenges, and Opportunities. *Social and Personality Psychology Compass*, 2, 464-493.

Sosa, E. (2008). A Defense of the Use of Intuitions in Philosophy. In Murphy, D. & Bishop, M. (Eds.), *Stich and His Critics*. Oxford: Blackwell.

Sripada, C., & Stanley, J. Empirical Tests of Interest-Relative Invariantism. Forthcoming in *Episteme.*

Stanley, J. (2005). *Knowledge and Practical Interests*, Oxford University Press.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition* Online First http://dx.doi.org/10.1016/j.cognition.2012.05.017.

Steup, M. (2006). The Analysis of Knowledge. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/knowledge-analysis/#IEX>.

Stich, S. (2008). Reply to Sosa. In Murphy, D. & Bishop, M. (Eds.), *Stich and His Critics*. Oxford: Blackwell.

---. Experimental philosophy and the bankruptcy of the great tradition. In preparation, Rutgers University.

Swain, S., Alexander, J., & Weinberg, J. (2008). The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research*, 76, 138–55.

Tobia, K., Buckwalter, W. & Stich, S. (2012). Moral intuitions: Are philosophers experts? *Philosophical Psychology*. Online First DOI:10.1080/09515089.2012.696327.

Tsohatzidis, S. (2012). How to Forget that "Know" is Factive. Acta Analytica Online First: DOI: 10.1007/s12136-012-0150-8

Turri, J. (2011a). Manifest Failure: The Gettier Problem Solved. *Philosophers' Imprint* 11.8, 1-11.
---. (2011b). Mythology of the Factive. *Logos & Episteme*, 2, 143-52.
---. (2010). Is Knowledge Justified True Belief? Forthcoming in *Synthese*.
---. Knowledge as Achievement, More or Less. Forthcoming in *The Present and Future of Virtue Epistemology*, ed. Miguel Angel Fernandez, press TBD.

Ulatowski, J. (2012). Act Individuation: An Experimental Approach. Review of Philosophy and Psychology, 3, 249-262.

Uhlman, E., Pizzaro, D., Tannenbaum, D., & Ditto, P. (2009). The motivated use of moral principles. *Judgment and Decision Making*, 4, 476–491.

Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116, 87–100.

Vaesen, K. & M. Peterson. The Reliability of Armchair Intuitions. Unpublished manuscript, Eindhoven University of Technology.

Weatherson, B. (2005). Can We Do Without Pragmatic Encroachment? *Philosophical Perspectives*, 19, 417–443.

Weinberg, J. (2011). Out of the Armchair, and Beyond the Clipboard: Prospects for the Second Decade of Experimental Philosophy, Invited speaker, The 103rd Annual Meeting of the Sothern Society for Philosophy and Psychology. March 11, 2011.
---., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, 29, 429–60.
---., Gonnerman, C., Buckner, C., & Alexander, J. (2010). Are philosophers expert intuiters? *Philosophical Psychology*, 23, 331-355.

Williamson, T. (2005). Armchair philosophy, metaphysical modality and counterfactual thinking. *Proceedings of the Aristotelian Society*, *105*, 1-23.
---. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.

Wittgenstein. *Philosophical Investigations,* 4th edition, 2009, P.M.S. Hacker and Joachim Schulte (eds. and trans.), Oxford: Wiley-Blackwell.

Woolley, J. D. & Wellman, H. M. (1993). Origin and truth: Young children's understanding of imaginary mental representations. *Child Development, 64,* 1-17.

Wright, J.C. (2010). On Intuitional Stability: The Clear, the Strong, and the Paradigmatic. *Cognition* 115, 419-503.

Young, L., Cushman, F., Adolphs., R., Tranel, T., & Hauser, M. (2006). Does emotion mediate the effect of an action's moral status on its intentional status? Neuropsychological evidence. *Journal of Cognition and Culture,* 6, 291-304.

Zagzebski, L. (1994). The inescapability of Gettier problems. *The Philosophical Quarterly*, 44, 65-73.