AMORALISTS, INVERTED COMMAS, AND THE PUZZLE OF MORAL INTERNALISM: AN ESSAY IN EXPERIMENTAL METAETHICS

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

KENNETH WESLEY SHIELDS

Dr. Philip Robbins, Dissertation Supervisor

July 2016

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

AMORALISTS, INVERTED COMMAS, AND THE PUZZLE OF MORAL INTERNALISM: AN ESSAY IN EXPERIMENTAL METAETHICS

Presented by Kenneth Wesley Shields

A candidate for degree of Doctor of Philosophy, and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Philip Robbins

_____

Dr. Laura King

_____

Dr. Joshua Knobe

_____

Dr. Robert Johnson

_____

Dr. Peter Vallentyne

ACKNOWLEDGEMENTS

Without the help and love of many people, I would not have completed this degree. There were many times along the way where I couldn't carry on: it is only because of these people that I did. I would like to thank my first clarinet instructor, Mr. Charles Lintz, who helped me develop self-discipline in both music and life. I would like to thank my high school band directors, Mr. Jack McElhannon and Dr. Brack May, who not only helped me survive one of the lowest points in my life, but who also served (unofficially) as my first philosophy instructors. I would like to thank my college clarinet instructor Dr. Steve Becraft, who also helped me through difficult times and served as a philosophical sparring partner. I would like to thank my college band director Mr. David Rollins, who taught me and supported me in spite of my sophomoric attitude at the time.

A very important thank you goes to Dr. Kevin K. J. Durand. He introduced me to philosophy in a summer introductory course during my undergraduate career, and from that point on I don't believe I left his side for more than a few days. He graciously allowed me to sit in on every philosophy class he taught; I showed up to every weekly philosophy club meeting; and I'm sure I followed him to his car multiple times to continue the philosophical discussion for just a bit longer. And when I realized that I was no longer willing to be a secondary school music teacher, he sat with my wife and I at a Waffle House (probably at 2am) and helped me discover what I really wanted to do. I remember him asking me: "So what do you think about what I do?" And I said, "Teach philosophy at a university? Of course I want to do that, but that isn't possible this late in the game, right?" I am so grateful that he believed in me enough to help me begin what has been a decade-long quest to do what I really wanted to do: teach philosophy.

I would like to thank everyone in the Mizzou philosophy department—it was through their personal attention that I was able to complete this degree. In particular: Dr. Claire Horisck, who worked hard to get me to Mizzou at the last minute, and who helped me work through my first-year insecurities; Dr. Andrew Melnyk, who introduced me to the inner-workings of contemporary, analytic philosophy through his proto-seminar; Dr. Peter Vallentyne, whose attention to clarity and precision forced me to develop more strength and clarity in both my prose and my thought; Dr. Robert Johnson, who introduced me to metaethics and prominent metaethicists in my field (like Dr. Michael Smith, who I would also like to thank for some great discussions and words of encouragement!); Dr. Peter Markie, who helped me develop my confidence as an instructor and a graduate student; Dr. Kenny Boyce, who gave me generous portions of his time (and patience) in order to further my grasp of philosophy of religion; Dr. André Ariew, whose door was always open; and of course the dedicated assistance of Mrs. Jonni Paxton and Mrs. Laural Youmans. I would also like to thank my fellow graduate students for putting up with me for so long, particularly Isaac Wagner, Josh Smart and Peter (Angelo) Graf for sticking by me.

I would like to thank my advisor and mentor, Dr. Philip Robbins, for encouraging and supporting me throughout the program (Philip also introduced me to Dr. Joshua Knobe, whose support was essential to beginning my career as a professional philosopher). I feel extremely lucky to have had Philip as my advisor; I am the philosopher and professor I am today in a big part because of my attempts at emulating his example.

I would like to thank my parents, Gary and Veronica Shields. I could not have achieved anything without their love, guidance and support.

I would like to thank Dr. Jake Wright, who is a close friend and who served as my graduate student mentor when I first arrived at the Mizzou philosophy program. He always knew the right thing to say to calm my nerves and self-doubts (and he still does!).

I would like to thank Travis Fox, without whom I would have dropped out of the philosophy program my second semester. The anxiety and stress that comes with working through a graduate program, while also being the father of newborn twins, would have been too much for me to bear if it was not for Travis.

I would like to thank my closest friend, Ryan Dickson. Only Ryan would stay up countless nights with me to see how deep the philosophical rabbit hole goes. I'm very lucky to have a friend like him.

Finally, I would like to give a very important thank you to my wife, Nicole Shields. I could not have imagined where we would be when we first began dating seventeen years ago. Nicole believed in me long before I ever did, and I could not have pursued this degree, much less completed it, without her. I am so lucky to have met her; this is as much her achievement as it is mine.

TABLE OF CONTENTS

List of Figures

Chapter 1: Introduction

In its broadest sense, metaethics is philosophical reflection about ethics. Metaethical questions arise in every branch of Western analytic philosophy: whether there are any moral facts (metaphysics), how moral knowledge is possible (epistemology), what distinguishes moral judgments from other kinds of normative judgments (philosophy of mind), what does it mean to say that something is morally wrong (philosophy of language), and so on.[1] Recently, practitioners of a relatively new movement within philosophy—known as experimental philosophy—are applying the tools of the social sciences to some of these metaethical questions. Part of the purpose of my dissertation will be to investigate, clarify and critically assess the relationship between the nature of certain metaethical questions and the methods employed by experimental philosophers to address those questions.

The central question addressed in this dissertation is whether one must have some degree of motivation to comply with their moral evaluation in order to count as genuinely making a sincere moral judgment. Those that view motivation as intrinsic to moral judgment (internalists) grant this condition on moral evaluation, while those that take motivation to be extrinsic to such judgments (externalists) deny this condition. The traditional dispute between internalists and externalists has centered around thought experiments devised to test the coherence of scenarios involving an agent that genuinely makes moral judgments while being entirely unmotivated by them–an individual called the amoralist. Recently, experimental methods have been employed to determine whether

[1] Miller (2013), p.2.

non-philosophers find amoralist scenarios coherent. This dissertation is concerned primarily with addressing two open questions regarding this recent experimental research: (1) what is this research really tracking in terms of folk psychology, and (2) what impact does this research have on the traditional philosophical dispute over moral internalism. I address (1) by presenting new research showing that amoralist scenarios seem more coherent in factive contexts (e.g., understands that X is wrong) but less coherent in non-factive contexts (e.g., believes that X is wrong). I call this the *Factivity Effect*, and I argue (via experiments) that it is likely a feature of our cognitive architecture concerning morality. I address (2) by arguing that empirical investigation of our shared concepts impacts metaethical questions–particularly the traditional dispute over moral internalism–in a way that is arguably unique to this branch of analytic philosophy. In short, moral psychology is vital for metaethics.

This dissertation will be structured as follows. In chapter 2, I review the traditional debate on moral internalism and amoralist skepticism. In chapter 3, I critically examine the experimental research on moral internalism and present my initial research as well, showing how each has recently culminated into the discovery of the *Factivity Effect*. In chapter 4, I present a variety of studies testing different explanations of the *Factivity Effect*. None of these studies found evidence that the *Factivity Effect* is an experimental artifact, but one study provided some evidence for the idea that the *Factivity Effect* reflects a feature within the structure of folk cognition itself. In chapter 5, I give a more theoretical examination of the relationship between these experimental studies and the traditional debate over moral internalism and amoralist skepticism. I consider two

different approaches for experimental philosophy to contribute to metaethics, and I defend this relationship from objections. In chapter 6, I give some concluding remarks.

Chapter 2: Moral Internalism and Amoralist Skepticism

1. Introduction

Suppose you tell me that you're a sucker for anything that is green: green shirts, green hats, green coffee mugs, etc.[2] So one day we are walking in a mall and you spot a green coffee mug in the window of a store. You stop and say, "Look, we should buy that because it's green!" And I say, "While I agree that it's green, I have no interest whatsoever in buying it." It doesn't seem surprising that I can see that the mug is green even though I have no desire to purchase it. But now let's change the story a bit. Suppose we continue to walk through the mall when we spot a mugging taking place in a store. You stop and say, "Look, let's notify the security guard or something because we should help stop that mugging!" And I say, "While I agree that we should help stop that mugging, I have no interest whatsoever in notifying the security guard or doing anything about it." Does my lack of interest in this case seem no less surprising than in the former? Is it just as clear that I can see that we should help stop the mugging as it is that I can see that the coffee mug is green? Or does it seem a bit surprising that I could really agree that we should stop the mugging if I have absolutely no intention to do so?

Let's consider another example. Suppose your friend tells you that, after thinking long and hard on the matter, she no longer thinks it's okay to eat meat. She now tells you that she thinks eating meat is morally wrong. Yet when you two go out to eat for lunch that day, she calmly orders a bacon cheeseburger. You might say, "but I thought you now think eating meat is morally wrong!" And suppose she replies, "I do think eating meat is

[2] I'm indebted to Paul Bloomfield for this example.

morally wrong; I just don't care if I do things that are morally wrong." Might you begin to question whether your friend really understands what it means to judge some behavior as morally wrong? Or might you wonder how you should interpret her expression of apathy concerning morality?

Philosophers within metaethics have taken an interest in what explains our surprise in these cases. They seem to have a ring of paradox to them, akin to Moore's paradoxical example of someone that asserts that it's raining outside while also claiming that they don't believe it.[3] And one way to explain this surprise is to locate the paradoxical tension within the concepts being employed. So in Moore's example, one might suggest that it's simply part of what it *is* to make a genuine, sincere assertion that it's raining that one believes it. If one doesn't really believe that it's raining, then either they didn't really make a genuine assertion to that effect, or their claim to not believing it's raining is in some way insincere or not genuine. Something similar can be said for the above cases. In the mugging case: perhaps it should be said that person either doesn't really think that they should help stop the mugging, or if they do think this, then perhaps their declining to help is in some way disingenuous (e.g., perhaps it reflects an inner struggle between a motivation to do the right thing and a motivation to avoid confrontation). In the cheeseburger case: perhaps it should be said that your friend doesn't really think that eating meat is morally wrong, or if she does think this, then perhaps her claim that she's apathetic when it comes to moral matters is in some way disingenuous (e.g., perhaps it reflects an inner struggle between a motivation to do the right thing and a motivation to do whatever she presently most wants to do).

[3] Baldwin (2010).

At the core of these accounts seems to be a connection between genuinely judging that some behavior is morally right or wrong, on the one hand, and the presence of some motivation to behave accordingly, on the other. Indeed, some philosophers have suggested that this connection is *conceptual*, which would then seem to clearly explain our surprise concerning stories that involve breaking this connection. We can capture this conceptual connection with the following simple conditional, which we can call (*strong) moral internalism*:

> (*strong) moral internalism*: if someone judges that they should ϕ, then they are motivated to ϕ.

Moral internalism is so called because it identifies moral motivation as being *internal* to the judgment itself.[4] The 'strong' qualification here is meant only to denote that this simple version of moral internalism doesn't allow for weakness of will; but as we will see in later sections, there are many versions of moral internalism.

For now, I want to emphasize how moral internalism is meant to explain our surprise in the above cases. Moral internalism expresses a conceptual connection between making a moral judgment about what one should do and one's motivation to comply. In this way, moral internalism describes what kinds of judgments *count* as moral judgments. It's important to note that moral internalism is not based on some inductive argument that cites all of our encounters with moral judgments up to the present day. In other words, it's not itself an empirical claim to be tested via the scientific method. Rather, it's more akin to the philosopher's standard example of an *analytic* claim: if someone is a bachelor, then that person is an unmarried male. Just as this claim gives us a necessary condition to

[4] For a discussion of what I'm calling moral internalism and its role in the metaethics literature, see Rosati (2008); Johnson (1999).

be fulfilled before someone can *count* as a bachelor—that the person be an unmarried male—so too does moral internalism provide us with a necessary condition to be fulfilled before someone can *count* as making a moral judgment—that the person be motivated to comply with said judgment.

But one might immediately note: while the condition on who counts as being a bachelor seems obviously true, moral internalism doesn't seem nearly as obvious. Why do some philosophers think that there is such a condition on what counts as a moral judgment? To answer this, it may be helpful to look back to the beginnings of analytic metaethics, which arguably takes us back to G. E. Moore's *Open Question* argument.[5]

2. Moore's Open Question

While metaethical questions were certainly addressed by philosophers before the 20th century, what has come to be recognized as analytic metaethics has its origins in the philosophy of G. E. Moore.[6] Specifically, Moore's *Principia Ethica* (1903) focused philosophers' attention on questions about the nature of morality and moral language, rather than on what has now come to be called first-order moral theorizing.[7] It was in this work that Moore first developed what has come to be called the *Open Question* argument against attempts to give naturalistic accounts of moral properties.[8] Briefly, the argument is that, for any proposal that takes moral properties to be identical with natural properties, it can always be asked—reasonably and without confusion—whether the identity actually holds. The standard example is to consider a simple hedonist view of value that identifies

[5] I follow Darwall (1992) on the history of this debate within metaethics.
[6] DeLapp (2016); Hurka (2015).
[7] DeLapp (2016).
[8] The rest of this paragraph draws on Hurka (2015).

moral goodness with pleasure. According to Moore's *Open Question* argument, if moral goodness really is identical to pleasure, then we should find the claim 'pleasure is good' to be no more informative than the obvious tautology 'pleasure is pleasure.' But we don't find the former claim to be as conceptually *closed* as the latter, which reveals (Moore argued) that goodness is not really identical to pleasure after all.

To be sure, Moore's *Open Question* argument has been criticized quite thoroughly, with perhaps the most notable criticism being that Moore simply confused concepts with properties.[9] While I think Moore has the resources to reply to some of these criticisms (which I will discuss when I return to this argument in chapter 5), my main reason for discussing this argument is not to defend it. Rather, I mention it because of the role it has played in the history and development of moral internalism. According to Stephen Darwall, moral internalism (what he calls judgment internalism)[10] was seen by a number of philosophers (most notably, Charles Stevenson and R. M. Hare) as the best way to explain the openness in Moore's open question.[11] If part of what it *is* to judge something as morally good is to be *motivated to pursue it*, then when we are confronted with, say, the hedonist proposal, our seeing that something is pleasure will fall short of our judging that it is good without the necessary motivation. In other words, when we make a moral judgment (or, in this case, a judgment about moral goodness), we appear to be doing more than simply *describing* some state of affairs (recall the example with

---

[9] Ibid.

[10] As Darwall (1992) explains, there are actually two kinds of internalism: existence internalism and judgment internalism. My dissertation is focused entirely on what Darwall calls judgment internalism. But because philosophers in the experimental literature refer to judgment internalism as moral internalism, I have chosen to employ their label instead of Darwall's.

[11] Ibid., p.160.

which we began this chapter: between our judgment that a coffee mug is green versus our judgment that we should help stop a mugging). These philosophers claimed that we seem to be expressing our *favor/disfavor* or *commendation/condemnation* when we make moral judgments—and it is this feature of moral judgment-making that they thought explained why Moore's open question seems "so difficult to close."[12]

It's also worth noting that some philosophers have taken something like moral internalism to be a *methodological* constraint on any conceptual analysis of the concepts employed in moral discourse. Among Stevenson's three criteria for investigating the moral sense of 'good', he cites the "magnetism" of goodness, describing this magnetic feature in terms similar to what I'm calling moral internalism:

> [. . .] "goodness " must have, so to speak, a magnetism. A person who recognizes X to be "good " must *ipso facto* acquire a stronger tendency to act in its favor then he otherwise would have had.[13]

So it seems that moral internalism is at least as old as analytic metaethics itself, and further it has served as part of the phenomena *to be explained* as much as it has served as an explanation itself. In the next section, however, I will introduce one of the main challenges to moral internalism: amoralist skepticism.

3. Amoralist Skepticism and Moral Externalism

'Skepticism' in this context doesn't refer to a questioning of how we come to have moral knowledge or whether there is even any moral knowledge to be had. Instead, *amoralist skepticism* is the view that one might grant that there are moral facts, and grant that we

[12] Ibid., p.160-161. It's worth noting that Moore himself may have not actually endorsed moral internalism, despite the role his philosophy has played in its discussion in the literature – see Hurka (2015).
[13] Stevenson (1937), p. 16.

have knowledge of such facts, but then go on to question "why we should *care* about these facts."[14] Thus, when we speak of the *amoralist*, we don't mean someone that lacks a moral conscience, or even someone who has a radically different moral outlook relative to most people. The amoralist is instead supposed to be an individual that makes the same moral judgments as the rest of us (e.g., judges that killing innocent people for fun is morally wrong), yet—unlike us—is entirely unmoved by such judgments.[15]

But now we have a dilemma: either amoralist skepticism is *in*conceivable or moral internalism, understood as a conceptual truth, is false. If internalism is true—if it's simply part of the very concept of a moral judgment that some motivation follow from moral judgment-making—then the amoralist is not only an unlikely individual: the amoralist must be an *inconceivable* individual. If one could imagine, without any conceptual confusion, an individual that makes genuine, sincere moral judgments without the slightest motivation to comply with those judgments, then this would serve as a fatal counter-example to moral internalism. Thus, some philosophers have devised what we can call *amoralist scenarios* with the aim of demonstrating that moral internalism is false. While we will examine one amoralist scenario from the traditional literature in detail, it's important to note that all amoralist scenarios involve at least the following two features: (1) a depiction of an individual as either knowing that some behavior is morally right or wrong, or as judging to that effect, and (2) an illustration of this individual's *complete* lack of motivation towards complying with what they know, or have judged, to be morally right or wrong to do in the circumstances. This is to be expected, of course, given that moral internalism is itself a conditional claim: it is meant to capture a connection

[14] Brink (1989), p.46. Emphasis in original.
[15] Ibid.

between genuinely making moral judgments and there being some motivation to comply. Thus, feature (1) is designed to ensure that the antecedent is satisfied—a moral judgment has been made—while feature (2) is designed to ensure that the consequent fails to be satisfied—no motivation is present.

Let's look at an amoralist scenario from the traditional literature. In her article, "Moral Cognitivism and Motivation," Sigrún Svavarsdóttir argues that moral internalism should no longer be a methodological constraint on the investigations of metaethicists into the nature of moral judgment and morality generally. Part of her argument relies on an amoralist scenario she describes involving an amoralist individual named Patrick. Since we will just be discussing this one amoralist scenario in detail, I think it is worth presenting it in its entirety:

> *The Example of Patrick*: Virginia has put her social position at risk to help a politically persecuted stranger because she thinks it is right thing to do. Later she meets Patrick, who could, without any apparent risk to himself, similarly help a political persecuted stranger, but who has made no attempt to do so. Our morally committed heroine confronts Patrick, appealing first to his compassion for the victims. Patrick rather wearily tells her that he has no inclination to concern himself with the plight of strangers. Virginia then appeals to explicit moral considerations: in this case, helping the strangers is his moral obligation and a matter of fighting enormous injustice. Patrick readily declares that he agrees with her moral assessment, but nevertheless cannot be bothered to help. Virginia presses him further, arguing that the effort required is minimal and, given his position, will cost him close to nothing. Patrick responds that the cost is not really the issue, he just does not care to concern himself with such matters. Later he shows absolutely no sign of regret for either his remarks or his failure to help.[16]

Note that Svavarsdóttir's Patrick example meets both of the features I described above. First, she depicts Patrick as making a genuine moral judgment—"Patrick readily declares that he agrees with her moral assessment"—and second, she depicts Patrick as being

[16] Svavarsdóttir (1999), p. 176-7.

entirely unmotivated to comply with his moral judgment—"Patrick responds that the cost is not really the issue, he just does not care to concern himself with such matters." And it's important to note that Svavarsdóttir does not describe Patrick as having made a moral judgment or as having moral knowledge. As she points out, "[n]othing has been said about Patrick's mental states."[17] It would clearly be question-begging to simply assert that an individual has made a moral judgment while nevertheless remaining entirely unmoved. Amoralist scenarios are instead designed to present, as plausibly as possible, a situation where it seems that the best explanation of the agent's behavior involves attributing a genuine moral judgment to the individual while, at the same time, granting that the individual lacks any degree of motivation to comply with this judgment. After offering a bit more information "about how Patrick has behaved in the past," Svavarsdóttir suggests that we should say this about her amoralist Patrick: "he knew what was right to do in the circumstances, but could not have cared less."[18] If this is the best explanation of Patrick's behavior, then we seem to be left with a clear counter-example to moral internalism.

Amoralist skepticism serves as part of the motivation for the antithesis to moral internalism: *moral externalism*. According to moral externalism, the connection between moral judgment-making and motivation is *not* necessary; rather, it is contingent upon the existence of something else that, in effect, *glues* the two mental states together.[19] Some moral externalists think the glue is our normal feelings of sympathy for others, while

---

[17] Ibid., p.177.
[18] Ibid.
[19] Philosophers that have endorsed some form of moral externalism include Brink (1989), Svavarsdóttir (1999), Milo (1981), Boyd (1988), Railton (1986) and Stocker (1979).

others (such as Svavarsdóttir) suggest that the glue is our general desire to be moral.[20] But whatever the glue turns out to be, what defines moral externalism is that it be something which some people may, in fact, lack while nevertheless retaining their capacity to make genuine moral judgments. Considering the amoralist Patrick, a moral externalist may describe Patrick as realizing that the morally right thing to do is to help the politically persecuted stranger, but, supposing we take Svavarsdóttir's favored account, what explains Patrick's failure to be motivated by his moral knowledge is the unfortunate fact that Patrick lacks a desire to be moral. It is this contingent feature of Patrick's psychological disposition that explains why he fails to be motivated by his moral judgments.

Moral internalists have traditionally responded to amoralist challenges in at least three ways[21]: either (1) they re-calibrate the moral internalist thesis to include a *condition* constraint (e.g., that moral internalism holds under certain conditions, such as when the individual is psychologically healthy and fully rational), or (2) they concede that amoralists are conceivable but only because of a hidden deference to a "background internalism," or (3) they move up a level and offer a debunking explanation of what we

[20] Brink (1989); Svavarsdóttir (1999), p. 170.

[21] There is a fourth way not discussed in this dissertation which involves abandoning the claim that moral internalism is a conceptual truth while instead taking the connection between moral judgment and motivation to be more akin to an *a posteriori* identity, like the connection between water and H^2O (See Björnsson (2002)). This view has the advantage of retaining the necessity of motivation following from moral-judgment making without the baggage that comes with going the conceptual route. However, it's not clear to me that one can gather the kinds of uncontroversial evidence one would need to support this view as one can for the water/H^2O identity claim. It should also be noted that this is a minority position within this area of research.

can call *externalist intuitions* regarding amoralist scenarios. Let's look at this last way of responding first.

4. The Inverted Commas Response

One way internalists traditionally respond to amoralist challenges is to give what has now come to be called the *inverted commas response*.[22] This response involves the claim that externalists have simply confused two different kinds of individuals: one is perfectly conceivable and consistent with moral internalism whereas the other is neither. The perfectly conceivable individual is one that merely utters words like "it is morally wrong to ϕ" while not actually asserting them, thus using the phrase "morally wrong" in an inverted commas sense (hence the response's name). Traditionally, the conceivable individual is presumed to be referencing the moral views of others in her community when uttering words like "it is morally wrong to ϕ" while not endorsing these views herself.[23] But if amoralist scenarios seem coherent simply because externalists are conceiving of *this* individual—the individual that uses moral terms in an inverted commas sense—then (as the response goes) such amoralist challenges fail to establish the

---

[22] This response seems to have been first proposed by Hare (1969), p.164. See also Hare (1981), p.183. Hare allows for an agent to make the moral judgment that everything is permissible – that this is logically consistent and conceivable. What is of interest is whether someone can make the moral judgment that something is forbidden, or something is obligatory, while remaining entirely motivationally indifferent in regards to their own judgment. It is this agent that Hare takes to be inconceivable.

[23] Here is a good explication of Hare's inverted commas response from Milo (1981): "What he believes (or judges) is merely that X is generally held (but not by him) to be wrong. And, if so, he will be making, not a genuinely evaluative moral judgment, but a merely descriptive moral judgment which simply states that not doing X is required in order to conform to the moral standards that people generally accept." (p.380).

conceivability of amoralists. Remember that an amoralist is not someone who happens to have different moral views than the rest of society. Internalism is consistent with the existence of iconoclasts or any individual that has moral views that differ, even drastically, from what's average. What's at issue is whether there can be an individual who may well share the very same moral views as others, but is nevertheless entirely unmoved by their own moral judgments. The inverted-commas response is the charge that externalists are simply confusing the former kind of individual with the latter. On this response, the presence of this confusion better explains why amoralist scenarios seem coherent even if moral internalism is, in fact, true.

The inverted commas response provides a kind of error-theory for externalist intuitions in amoralist scenarios. Of course, one problem with the 'inverted commas response' is that externalists will presumably deny that they are making any such error.[24] We will revisit the inverted commas response numerous times in the coming chapters as it plays a key role in the experimental work on moral internalism and amoralist skepticism.

5. Conditional Internalism

The second way internalists traditionally respond to amoralist challenges is to claim that these amoralist scenarios only show that the internalist thesis fails unless certain *conditions* are met. Perhaps the amoralist seems conceivable only because certain defects of psychology, rationality or moral perception have not been ruled out. Once the individual is sufficiently idealized (so this response goes), the amoralist scenarios may no

[24] Shaun Nichols (2004b) raises this problem for the 'inverted commas response', p.73.

longer seem coherent. So even if it is admitted that amoralist scenarios show that *strong* moral internalism is false, the aim is to provide a more plausible version of internalism that accounts for the likely psychological or rational failings that would underlie amoralism while still maintaining a conceptual connection between moral judgment and motivation in these idealized situations. With this caveat, we get a relatively weaker internalist thesis that we can call *conditional internalism*:

> *Conditional internalism*: necessarily, if a person judges that she morally ought to ϕ, then she is (at least somewhat) motivated to ϕ, if she satisfies condition C.

There have been a number of suggestions about how to fill out condition C. Some have defined the condition in terms of *psychological normalcy*, such that only amoralists that happen to be suffering from some psychological malady – such as depression, exhaustion, or acedia – are conceivable outright.[25] Others have defined the condition in terms of *moral perception*, such that only amoralists that happen to embody some defect in their sensitivity to moral features and properties are conceivable.[26] Finally, some have defined the condition in terms of *practical rationality*, such that only amoralists that happen to embody some defect in their rationality – such as weakness of will – are conceivable. Michael Smith's practicality requirement is a particularly well-known

[25] See Smith (1994), Shafer-Landau (2003).

[26] Björklund et. al. (2012) cite McDowell (1978), but McDowell seems to be defending a form of existence internalism in this article, not judgment internalism. Still, it may be possible to construe existence internalism in terms of judgment internalism. Suppose we said that existence internalism is merely simple internalism coupled with a success condition, such as in the following formulation: "If one judges that one morally ought to X, and if it is true that one morally ought to X, then one is motivated (at least to some degree) to X." Perhaps this is enough for the variant of conditional internalism that Strandberg and Björklund have in mind?

instance of this kind of conditional internalism in both the traditional and experimental literature:

> *Practicality requirement*: If an agent judges that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational.[27]

This requirement amounts to the claim that a practically rational amoralist is inconceivable. As we will see in the next chapter, Smith's practicality requirement, like the inverted commas response, has served as a catalyst for experimental research in this area.

6. Deferred Internalism

The final way internalists traditionally respond to amoralist challenges is to charge that externalist intuitions are "parasitic" upon a "background connection between ethics and motivation."[28] According to this response, as Simon Blackburn describes it, amoralist scenarios are "cases in which things are out of joint, but the fact of a joint being out presupposes a normal or typical state in which it is not out."[29] So the claim is essentially that externalist intuitions themselves must rely on some deference to a kind of background internalism.

This position has been dubbed *deferred internalism*.[30] There are two varieties of deferred internalism: one in which the relevant deference takes place diachronically along the individual's own past moral judgments; the other in which the relevant deference

[27] Smith (1994), p. 65.
[28] Blackburn (1998), p. 61.
[29] Ibid.
[30] Björklund et. al. (2012), p.4-6.

takes place synchronically along the moral judgments made by those in the individual's community:[31]

> *Deferred Individual Internalism*: necessarily, if a person judges that she morally ought to ϕ, then some of her past moral judgments were accompanied by some degree of motivation.

> *Deferred Communal Internalism*: necessarily, if a person judges that she morally ought to ϕ, then she is a member of a community whose member's moral judgments are generally accompanied by some degree of motivation.

On the individual variety, any kind of amoralist is conceivable as long as moral internalism correctly captures the connection between judgment and motivation concerning some of the amoralist's past moral judgments. So too on the communal variety: any kind of amoralist is conceivable as long as moral internalism captures the connection between judgment and motivation concerning the moral judgments made by the other members of that amoralist's community.[32]

Now it might seem that once internalists begin qualifying their thesis, as they have done with the conditional and deferred varieties of internalism, they have simply given the game away to the externalists. After all, externalists grant that there is a very strong connection between moral judgment-making and motivation.[33] They merely deny that such a connection is *necessary*. In fact, it seems that externalists may grant that most (or even all!) actual amoralists are likely to be of the kind that the conditional and deferential internalists bracket off. So what's the big deal?

---

[31] Ibid.
[32] What deferred communal internalism does rule out is the conceivability of an entire community of amoralists. For discussion on this point, see Tresan (2009), Bedke (2009), Lenman (1999).
[33] Svavarsdóttir (1999), p. 196.

The big deal, I take it, is what externalism seems to imply about what we are doing when we make moral judgments. After determining that some action available to us is morally wrong, it seems incoherent (to internalists at least) to then go on to ask "but *what reason do I have* to avoid doing actions I've determined to be morally wrong?" If the judgment that the action in question is morally wrong hasn't provided you with even the slightest motivation to avoid the action, there seems to be some wonder about what exactly the judgment was about in the first place! Wasn't judging that the action is morally wrong itself a judgment about what action you should avoid? It's likely for concerns like these that internalists insist that some kinds of amoralist are, as a matter of conceptual coherence, *inconceivable*. It's not that we just happen to lack any experience with such individuals (as the externalist might be willing to grant). Rather, we *could never* meet such an individual. Like the individual that is completely red all over and completely green all over, certain kinds of amoralists are conceptually inconceivable, according to the internalist.

To see this, consider Smith's practicality requirement again. On this account, it is granted that some amoralists are conceivable. Specifically, the conceivable amoralists are those individuals that happen to be suffering from some lapse in their practical rationality. But once it's clarified that the individual in question suffers from no such lapse—that this individual is ideally practically rational—then Smith is committed to saying that *that* individual, construed as an amoralist, is simply inconceivable. The externalist, on the other hand, will want to say something much weaker: that we are unlikely to run into such an individual, and that perhaps our own concepts were shaped

within an environment where there weren't any individuals like that, but nevertheless such individuals are conceivable just the same.

7. Conclusion

While one could trace the philosophical roots of moral internalism as far back as Plato's *Republic,* it seems that the idea of a conceptual connection between moral judgment and motivation wasn't subjected to analytic scrutiny until the 20th century with the advent of modern-day metaethics. And while some philosophers have viewed moral internalism as not only true but itself a starting point for any investigation into the nature of morality and moral discourse, others have raised what we have been calling amoralist challenges to the internalist thesis. If a coherent story can be told about an individual that genuinely makes moral judgments but is nonetheless completely unmoved by these judgments—what we have been calling an amoralist scenario—then it seems the connection between moral judgment and motivation cannot be a conceptual one. Moral externalism is committed to the connection between motivation and moral judgment being contingent upon features of our psychology: we our motivated to behave in accordance with moral judgments only insofar as we happen to have certain desires to achieve certain goals (e.g., a desire to be moral, or a desire to do the particular action one's moral views require, or a desire to behave in an impartial manner).[34] We have seen that internalists have attempted to buffer their thesis from amoralist challenges by either weakening the thesis or by offering an error-theoretic account of externalist intuitions in these amoralist scenarios. But despite the multiplicity of internalists theses that has resulted from this kind of arms

[34] See Svavarsdóttir (1999), Brink (1989), and Railton (1986) respectively.

race between externalists and internalists—one side crafting more encompassing amoralist scenarios while the other side crafts more refined internalist positions—one thing all of these varieties of internalism have in common is that they deny that the connection between moral judgment and motivation is contingent. And as we will now see in the next chapter, it is partly the conceptual nature of moral internalism that has initially attracted experimental philosophers to this traditional debate within metaethics.

Chapter 3: Experimental Philosophy and Moral Internalism

1. Introduction

In the previous chapter, it was shown how much of the traditional dispute over moral internalism and amoralist skepticism has centered on our intuitions about thought experiments. Externalists press amoralist scenarios, like Svavarsdóttir's amoralist Patrick, and internalists have countered with their own thought experiments, like Dreier's *Sadist* community.[35] This is not particularly out of the ordinary, of course. Much of traditional philosophy is carried out by way of intuitions and thought experiments. Yet philosophers and cognitive scientists have recently raised concerns about the nature of intuitions and how they are related to thought experiments.[36] Around the beginning of the 21st century, these concerns and questions coalesced into a general research program called *experimental philosophy*, given its use of statistical research methods to explore philosophical questions.

Experimental philosophy traditionally involves constructing surveys whose vignettes and questions are designed to bear on certain philosophical issues. One of the first surveys of this kind (and perhaps the most well-known) comes from Joshua Knobe's research on the impact of moral judgments on people's application of the concept of behaving intentionally.[37] In Knobe's pioneering study, participants were given one of the two following vignettes and test questions:

> [vignette (a)] The vice president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us

[35] Dreier (1990).
[36] Knobe et al. (2012)
[37] The following is from Knobe et al. (2012).

> increase profits, and it will also help the environment." The chairman of the board answered, "I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program." They started the new program. Sure enough, the environment was helped.
>
> [test question (a)] Did the chairman intentionally help the environment?
>
> [vignette (b)] The vice president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, and it will also harm the environment." The chairman of the board answered, "I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program." They started the new program. Sure enough, the environment was harmed.
>
> [test question (b)] Did the chairman intentionally harm the environment?

Participants famously gave asymmetric responses to these two scenarios: those given (a) denied that the chairman helped intentionally, whereas those given (b) claimed that the chairman intentionally harmed the environment. This surprising result has been replicated many times and across different cultures.[38] While different explanations of this result continue to be proffered, this experimental research method has been utilized within other areas of philosophy: free will and determinism, knowledge, moral responsibility, phenomenal consciousness, ethics and metaethics.[39]

Experimental research into moral internalism and amoralist skepticism is still in its infancy. Although it began around the advent of experimental philosophy itself (see Nichols (2002)), only a handful of research has been published in the area to date (See

[38] Rakoczy et al. (2015); Mallon (2008); Knobe and Burra (2006).
[39] *Free will and determinism*: Baumeister et al. (2009), Feltz and Cokely (2009), Nahmias and Murray (2010); *knowledge*: Murray et al. (2013), Myers-Schutz and Schwitzgebel (2013), Rose and Schaffer (2013), Beebe and Buckwalter (2010); *moral responsibility*: Nahmias et al. (2007), Roskies and Nichols (2011); *phenomenal consciousness*: Robbins and Jack (2006); *ethics*: Appiah (2008), Schwitzgebel and Cushman (2012); *metaethics*: Sarkissian et al. (2011), Goodwin and Darley (2008), (2010), (2012), Wright et al. (2013).

Strandberg and Björklund (2012); Leben and Wilckens (2015), and Björnsson et al. (2015)). Nevertheless, what has been published seems to initially suggest that most non-philosophers find amoralism *coherent* (notwithstanding Björnsson et al. (2015)). But a recent focus on how these studies were carried out – in particular, whether amoralists were described as having moral *knowledge* or just moral *belief* – has raised questions about these initial externalist-friendly findings.

In section 2, I explain how this early research seemed to favor folk externalism – participants seemed to have no qualms with attributing genuine moral judgments to amoralists. But in sections 3 and 4, I examine recent concerns over these early studies, with a particular focus on amoralist scenario construction. In section 3 I address the use of charity transgressions within amoralist scenarios, and in section 4 I address the use of factive terminology (e.g., she *understands* that X is morally wrong) within amoralist scenarios. Finally, in section 5 I show how investigating concerns like the above led to a surprising discovery: participants lean externalist when amoralists are described as *knowing* or *understanding* that *X* is morally wrong (i.e., when the amoralist's judgment is described *factively),* whereas participants lean relatively internalist when amoralists are described as *thinking* or *believing* that *X* is morally wrong (i.e., when the amoralist's judgment is described *non-factively*). I call this the *Factivity Effect* or *FE*. Far from offering further support for folk externalism, the *Factivity Effect* is arguably more congenial to folk internalism, or it might even undermine the traditional discussion over moral internalism altogether. The next chapter is devoted to research and discussion on what best explains the *Factivity Effect* and what philosophical import different

explanations have on the traditional metaethical debate over moral internalism and amoralist skepticism.

2. Early Evidence: Folk Externalism

As we've seen, contemporary discussion over the plausibility of either moral internalism or amoralist skepticism has generally come down to competing thought experiments. Thought experiments favoring internalism are intended to illustrate the incoherence of someone genuinely judging something morally wrong to do while simultaneously having no motivation to avoid performing the action. On the other side, thought experiments favoring externalism are aimed at revealing the coherence of just such an individual, the amoralist. But before Nichols (2002), no one had ever tested to see if non-philosophers were pre-theoretically disposed to either account. There had been claims from the armchair such as Michael Smith's claim that moral internalism is a folk platitude.[40] And insofar as internalism is construed as a conceptual truth, presumably those competent with the relevant concepts must then be committed to the internalist account.

So in an effort to put claims like Smith's to the test, experimental philosophers carried out studies to see if non-philosophers seemed more inclined towards internalism or externalism. And the early evidence seemed to point to externalism. As researchers presented amoralist scenarios to participants using a variety of different methods, and testing various formulations of internalism, the results seemed to support a kind of *folk externalism*.

[40] Smith (1994).

Shaun Nichols pioneered this research with a study testing whether people viewed psychopaths as making moral judgments. Nichols presented participants with the following vignette:

> John is a psychopathic criminal. He is an adult of normal intelligence, but he has no emotional reaction to hurting other people. John has hurt, and indeed killed, other people when he has wanted to steal their money. He says that he knows that hurting others is wrong, but that he just doesn't care if he does things that are wrong. Does John really understand that hurting others is morally wrong?[41]

Significantly more participants (nearly 85%) held that John really understood that hurting others is morally wrong.[42] Nichols cites the results of this study as evidence that people are more inclined to externalism than internalism. *Contra* the inverted-commas hypothesis (i.e., that amoralists make mere sociological judgments about the moral consensus, not genuine evaluative judgments), the results suggests that participants view John the psychopath having genuine moral judgment, despite not caring in the slightest about acting accordingly. If this is the best way to interpret these results, then it seems judgment internalism is not a conceptual commitment, at least among non-philosophers.[43]

Roughly a decade after Nichols' study, Caj Strandberg and Fredrik Björklund (2012) carried out a study testing multiple formulations of internalism. And for each

---

[41] Nichols (2002), p.289. The focus of Nichols' study was intended to be a version of conditional internalism, where the condition is related to practical rationality. Specifically, the focus was on Michael Smith's "practicality requirement" which states that, necessarily, if one judges that she ought to Φ, then either she will have some motivation to Φ or she is practically irrational. There are concerns about whether Nichols' study, as constructed, actually addresses Smith's conditional internalist thesis. See Joyce (2008).

[42] Ibid, p.289.

[43] Ibid, p.75. Nichols concludes that "it seems to be a platitude that psychopaths [who are thought to have no motivation to behave morally] really make moral judgments."

formulation, the results seemed to provide further evidence for folk externalism. Strandberg and Björklund presented participants with the following vignette:

> Anna is watching a TV programme [sic] about a famine in Sudan. In the TV programme [sic], it is shown how the starving are suffering and desperately looking for food. At the same time, Anna is not motivated at all, not to any extent, to give any money to those who are starving.[44]

Participants were then asked: "could it be the case that Anna thinks she is morally required to give some of her money to the starving even if she is not motivated at all to do so?"[45] 76% of respondents answered in the affirmative. On it's face, this would best be explained by a kind of folk externalism: participants find no problem in imagining that Anna takes herself to have a moral obligation to help the starving, even though she is entirely unconcerned with such an obligation.

While the above results seem to (at best) put pressure on a kind of strong moral internalism, one might wonder if these results raise problems for the more sophisticated internalists theses. For example, couldn't a conditional version of internalism allow for Anna to genuinely make moral judgments, provided that she is better explained as suffering from some psychological malady (e.g., depression)? And wouldn't versions of internalism that construe the motivational component as a deference to a *community* of typical moralists, rather than in any single individual, clearly allow for Anna as well? Strandberg and Björklund agree, which is why they also added further dimensions to their original setup in order to test these more sophisticated internalist theses. While retaining the vignette itself, Strandberg and Björklund added additional information about Anna. For example, one add-on describes Anna as being "deeply depressed":

---

[44] Strandberg and Björklund (2012), p.323.
[45] Ibid.

Depression

> Anna is deeply depressed. Most of the time she is sad and tired. She has also difficulties concentrating and is not interested in doing the things that use to appeal to her. As a result of her depression, Anna is not motivated at all, not to any extent, to give any money to those who are starving.

The other add-ons included *Apathy*, *Psychopath*, and *Normal Functioning*. Yet for each add-on (except *Psychopath*, interestingly enough), participants seemed to continue expressing a kind of folk externalism (*Normal Functioning* 79%; *Apathy* 60%; *Depression* 79%; *Psychopath* 40%).[46] Finally, to adjust their study to capture the communal varieties of internalism, Strandberg and Björklund devised a separate vignette and test question:

Community

> Imagine a society X that in most respects is similar to ours. The citizens of X look roughly as we do, behave roughly as we do, and like pretty much the same things as we do. Citizens of X know that there are other people who are starving and every now and then they watch on TV how these people are suffering and desperately looking for food. None of the citizens of X is ever motivated, not to any extent, to give any money to those who are starving.
>
> *Question*: could it be the case that citizens of X think that they are morally required to give some of their money to the starving even if none of them is motivated to do so?[47]

But even in this communal version of their study, 72% of participants granted that these citizens could be making the moral judgment, thus seemingly providing more evidence for what appears to be a pretty thorough folk externalism.[48]

---

[46] Ibid., p.324.
[47] Ibid.
[48] Ibid., p.330.

With these early results, it initially seemed that externalism had won the day: people appeared to be quite comfortable attributing moral judgment to the amoralist. People had no problem saying that John the psychopath really does understand that hurting and killing people is morally wrong, despite his complete lack of concern for behaving in this way. In fact, at least for one team of researchers, the evidence favoring folk externalism seemed strong enough to justify moving on to testing possible error-theoretic explanations for the minority of internalist responses received in these early studies.[49] But as we shall see, the results of these early studies are arguably hiding a more complex phenomenon. And far from giving direct support to folk externalism, much less externalism proper, we will see that this phenomenon may call for a fundamental reframing of the traditional debate concerning moral internalism and amoralist skepticism.

3. Against Folk Externalism: Charity Cases

There are essentially three main studies the results of which appear to suggest a kind of folk externalism: the psychopath study from Nichols (2002), the uncharitable Anna study from Strandberg and Björklund (2012), and the uncharitable community study from that same publication. I will argue that all three studies embody methodological flaws which, when taken into account, should lead us to question how well each supports folk externalism. In this section, I will focus on the uncharitable Anna study and the uncharitable community study. I will devote the next section to Nichols' psychopath study.

[49] Leben and Wilckens (2015)

While there are a number of specific concerns one may raise regarding Strandberg and Björklund's uncharitable Anna and uncharitable community studies, I intend to focus my concern on a feature that plays a crucial role in each: the use of charitable giving as the moral judgment in question. Let's begin my focusing on the uncharitable Anna study.

Given the plausible assumption that most people view charity as either morally supererogatory (i.e., good to do but not bad not to do), or simply morally optional, it's not unlikely that participants might assume that Anna shares this view as well.[50] But if participants were merely granting that Anna could think that giving to charity is morally right to do in this supererogation sense, such responses wouldn't conflict with moral internalism proper, traditionally construed as an account of judgments about moral *obligation* (i.e., it would be morally wrong to not perform the action).[51]

The context of charity makes key issues for testing moral internalism opaque. For instance, how can we tell if participants believe Anna thinks she's doing something morally wrong by not giving her money to charity? Surely most individuals don't believe that this would be morally wrong, but they would need to think that Anna departs from this majority view in order for their responses to bear on moral internalism. Moreover, how can we tell if participants believe Anna isn't motivated to give to charity *later in life*? As Kant might say, giving to charity is an example of an *imperfect* duty; one is obligated to perform it, but the time, place, manner, and *recipient* are all at your

[50] This assumption was in fact empirically confirmed by Leben and Wilckens (2015).
[51] Gunnar Björnsson has cited this worry as one reason for his team constructing a charity case where there would be "no cost involved in the charity." (email correspondence). I think charity cases should simply be avoided given the reasonable (and *empirically* supported—see Leben and Wilckens (2015)) expectation of participants conflating moral supererogation with moral obligation when these cases are used.

discretion.[52] Thus, if participants viewed Anna as seeing charitable giving as, in some sense, like an imperfect duty, then their affirmative responses to Strandberg and Björklund's test questions would not directly conflict with moral internalism. There are plenty of possible moral transgressions that do not raise such conflation worries (i.e., transgressions where it's quite obvious that moral obligation is at issue; for example: unnecessary stealing). Moreover, simply having participants evaluate whether the amoralist personally thinks some behavior is morally *wrong* would easily have sidestepped this worry.

But what about Strandberg and Björklund's uncharitable community study? This second study was designed to test communal internalism, which takes there to be a necessary connection between moral judgment and the presence of motivation to comply among others in the society of which the judgment-maker is a member. Admittedly, since an entire community is the object of evaluation, concerns about whether Anna's particular view of charitable giving is in-line with that of the participant's may be less pressing than in the uncharitable Anna study. Even so, it's not clear whether participants are conceiving of the transgression as being supererogatory in nature (despite the word "requirement") or perhaps just an instance where the citizens of this community aren't currently carrying out a Kantian imperfect duty they take themselves to have regarding the starving. And for what it's worth, later researchers that employed charity-based vignettes took pains to emphasize the zero-cost of the charitable behavior, perhaps attempting to evoke something like the Minimally-Decent-Samaritan principle (i.e., that one is morally obligated to help another person when the benefit is morally great and the

[52] Robert Johnson (personal correspondence). Also see footnote 68.

cost is virtually insignificant).[53] I personally think charity cases should simply be avoided given the evidence that people are likely to conflate moral supererogation with moral obligation.

4. Against Folk Externalism: Factivity

Nichols' study has received criticism from researchers in both traditional and experimental philosophy. I would like to focus on a particular concern here: Nichols' use of *'understand'* in his descriptions of the amoralist. 'Understand' is a *factive* term. In this context, a factive is a term that semantically implies that the judgment is true. For example, if I say that Mary *realizes* that the bank closes at four, I attribute a judgment to Mary (the judgment that the bank closes at four) but I also imply (semantically or pragmatically) that the bank actually does close at four. So for Nichols to describe the amoralist as *knowing* or *understanding* that hurting others is morally wrong is to imply that the amoralist's judgment accurately captures the moral facts. Non-factives, however, lack this implication. Non-factive judgments may address a domain without accurately capturing a fact within that domain. For example, if I say that Mary *believes* that the bank closes at four, I merely attribute a judgment to Mary (the judgment that the bank closes at four).

But why would it matter that Nichols used factives instead of non-factives in his amoralist scenario? Moreover: how is this factive/non-factive distinction relevant to moral internalism and amoralist skepticism? You might think that nothing really turns on whether we describe an amoralist as knowing or just believing in some moral proposition

[53] Gunnar Björnsson (email correspondence).

– what matters is simply whether or not she genuinely makes the moral judgment at all. In other words, you might think that describing the amoralist as having moral *belief* is equivalent to describing the amoralist as having moral *knowledge,* in the sense that these doxastic states can be interchanged without loss (or gain) in the scenario's overall conceptual coherence. Let's call this assumption *Descriptive Equivalence*, or *DE*:

> **Descriptive Equivalence** (*DE*): describing the amoralist as having moral *belief* is equivalent to describing the amoralist as having moral *knowledge*, in the sense that these doxastic states can be interchanged without loss (or gain) in the scenario's overall conceptual coherence.

DE likely reflects the standard view of most metaethicists. Of course, we can make *DE* more general by including all factives and non-factives like so:

> **Descriptive Equivalence** (*DE*): factive descriptions of the amoralist are equivalent to non-factive descriptions of the amoralist, in the sense that they can be interchanged without loss (or gain) in the scenario's overall conceptual coherence.

This broader formulation of *DE* still seems to capture a basic assumption held by most metaethicists. In fact, it's instructive to see how *DE* should work by considering some examples of amoralist scenarios in the traditional literature. A brief consideration of four characterizations of amoralism–two from prominent externalists and two from prominent internalists–will help reveal this tacit commitment to *DE*. It will also help introduce the difference between factive and non-factive descriptions of the amoralist.

Consider the externalist David Brink's introduction of amoralist skepticism:

> But another traditional kind of skepticism accepts the existence of moral ***facts*** and concedes that we have moral ***knowledge***, and asks why we should care about these ***facts***. Call this *amoralist skepticism*. Amoralists are the traditional way of representing this second kind of skepticism; the amoralist is someone who ***recognizes*** the existence of moral considerations and remains unmoved (1989, p.46).[54]

[54] Bold italics mine; non-bold italics in original.

Notice that Brink relies exclusively on *factives* to describe the amoralist's moral judgment. Now if *DE* is true, then the coherence of Brink's characterization shouldn't be affected by replacing his factives with non-factives. So Brink's 'someone who *recognizes* the existence of moral considerations and remains unmoved' can become 'someone who *believes* in the existence of moral considerations and remains unmoved.'

For another example, recall Svavarsdóttir's favored description of her amoralist Patrick: "He *knew* what was right to do in the circumstances, but could not have cared less" (1999, p. 178). If *DE* is true, then the coherence of her descriptions shouldn't be affected by the following non-factive equivalent: "he *believed* [that this] was right to do in the circumstances, but could not have cared less." If *DE* is correct, these translations are equivalent to their original factive versions relative to overall conceptual coherence.[55]

But now notice the internalist Jamie Dreier's use of non-factives when referencing an alleged amoralist politician:

> And suppose he [the politician] now says, "What my friends believe is wrong: not individualism but a life in the service of others is really good." But the politician has no inclination to serve the less fortunate; instead he advances the cause of self-reliance whenever he can [. . . .] It seems to me that what we want to say about the new politician is that he is using the word "good" either insincerely or incorrectly. We will not take his assertion at face value and attribute to him the ***belief*** "Life in the service of others is good" (1990, p.13).

Contra Brink and Svavarsdóttir, Dreier characterizes amoralism (or, more accurately, it's impossibility) through non-factives. But just as *DE* allows converting factive descriptions into non-factive descriptions, *DE* allows converting non-factive descriptions into factive descriptions. Dreier's assessment of the apathetic politician now becomes 'we will not

[55] For a third externalist example, see Boyd (1988, p. 216).

take his assertion at face value and attribute to him the *knowledge* "Life in the service of others is good."'

Finally, recall Michael Smith's practicality requirement: "If an agent *judges* that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational" (1994, p.65). If *DE* is true, then Smith's requirement has the following factive equivalent: "If an agent *knows* that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational." So, if *DE* is correct, these translations are equivalent to their original non-factive versions relative to overall conceptual coherence.[56]

Aside from this tacit commitment, there's a standard view concerning the relationship between knowledge and belief that seems to provide strong support for *DE*: the view that knowledge entails belief. This view has come to be called the *entailment thesis*.[57] It's likely (and plausibly) assumed that moral knowledge and moral belief would fall within the scope of this thesis. If this is right, and if the entailment thesis is true, then part of *DE* seems to follow: if an amoralist is described as knowing that *X* is morally wrong, then that same amoralist must also believe that *X* is morally wrong. But recent experimental results seem to cast doubt on even this part of *DE*.

So if *DE* is correct, then it seems reasonable to predict that Nichols' results would have remained unchanged if the factive terms were replaced with non-factive counterparts (e.g., 'He believes that hurting others is wrong…' and 'Does John really believe that hurting others is wrong?). Roughly a decade after Nichols' preliminary study,

---

[56] For an example where non-factives are used when describing internalism but factives used when describing externalism, see Korsgaard (1986, p.8-9).
[57] See chapter 4, section 3.2, for more discussion of the entailment thesis.

Björnsson, Eriksson, Strandberg, Olinder & Björklund (2015) ran follow-up studies testing this prediction. The results: no – participants were actually significantly more inclined to report relatively *internalist* intuitions when confronted with amoralist scenarios described using non-factives.

This effect—where people's intuitions lean externalist when the amoralist's moral judgment is described *factively* (e.g., understands or knows that she shouldn't do *X*) but then lean relatively internalist when the amoralist's judgment is described *non-factively* (e.g., believes or thinks that she shouldn't do *X*)—has been replicated numerous times both in Björnsson et al. (2015) and in my own research. For ease of reference, let's call it the *Factivity Effect* or *FE*:

> **Factivity Effect** (*FE*): people's intuitions lean externalist when evaluating factive amoralist scenarios, whereas people's intuitions lean relatively internalist when evaluating non-factive amoralist scenarios.

On the face of it, *FE* seems to be in tension with *DE*, thus challenging a likely foundational assumption within the traditional debate over moral internalism and amoralist skepticism. I now turn to the evidence for *FE*.

5. The Discovery of the Factivity Effect

The *Factivity Effect* (*FE*) was independently discovered by myself and by Björnsson et al. (2015). In both cases, we had designed more sophisticated versions of Nichols' *Psychopath* in order to account for a variety of different concerns, one of which was Nichols' use of factives instead of non-factives in his vignette and test question. And in both cases, we found that people's intuitions lean externalist when the amoralist's moral judgment is described *factively* (e.g., understands or knows that she shouldn't do *X*) but

then lean relatively internalist when the amoralist's judgment is described *non-factively* (e.g., believes or thinks that she shouldn't do *X*). I'm calling this the *Factivity Effect*, or *FE*. To see the evidence for *FE*, let's begin by looking at a study from Björnsson et al. (2015).

5.1 A New Psychopath Study

After failing to get Nichols' results from replicating his original psychopath study, Björnsson et al. (2015) constructed a new psychopath study that was designed to account for a variety of issues. They replaced Nichols' moral transgression with a much less violent wrongdoing (in an effort to block participants from granting understanding to amoralists merely out of a desire to hold them accountable). They also added a brief introduction about how people normally classify some actions as morally right and others as morally wrong. And, crucially, they varied their test question from factives to non-factives. Their choice to include non-factives was out of concern that Nichols' original, factive question "does not test internalism, as usually understood."[58] In their view, 'belief' seemed to better capture internalism's focus on moral *judgment*, whereas 'understand' seemed to downplay the judgment aspect.

Participants in their study received a somewhat lengthy vignette. The first part described our ordinary practice of classifying actions as morally right or morally wrong. The second part introduced the reader to the amoralist Anna, a woman who "classifies actions using expressions like 'morally right' and 'morally wrong,'" but, it's stipulated,

[58] Björnsson et al. (2015), p.720.

her classifications don't "influence her choices."[59] Finally, the vignette presents the reader with a scenario where Anna must choose one of two cell phones. While identical in type and price, only one includes a free donation of $20 to help address starvation in Sudan. While Anna does classify purchasing the non-donation phone as "morally wrong," and the donation-phone as "morally right," Anna's classification does not influence her choice of the phones in the slightest. Each participant was then randomly assigned to one of three test question variations:

> We have seen that Anna classifies some actions as "morally wrong". But because she lacks compassion and is strikingly egoistic, this never makes her even the least inclined to avoid these actions. We saw this indifference when she chose her cell phone. In light of this, would you say that she [understands/believes/herself thinks] that it is morally wrong not to choose the left phone?[60]

Participants given the *understands* variation responded similarly to how Nichols' participants responded—76% granted Anna understands; but in both the *believes* and *herself thinks* variations, the responses leaned internalist, relative to the understand variation—only 46% granted Anna believes, and only 49% granted she herself thinks.[61]

Replacing factives like 'understand' with non-factives like 'believe,' while leaving everything else in the study unaltered, significantly changed how subjects responded. Participants would lean externalist when the amoralist was described as *understanding* that *X* is morally wrong but then lean relatively internalist when described as *believing*, or *thinking*, that *X* is morally wrong. This is the *Factivity Effect* (*FE*).

Initially, Björnsson et al. (2015) viewed this result as inconsequential to the traditional debate because (a) it gives no reason to think that either side has stronger

---

[59] Ibid., p.721.
[60] Ibid., p.722.
[61] Ibid.

theoretical biases and (b) both sides will still need to explain the overall data, which appears to involve participants both granting and denying a doxastic state to the amoralist. Yet their attempts at testing possible explanations of *FE*—their *inverted-commas hypothesis* and their *blaming hypothesis*—didn't bear fruit (as we will see in the next chapter). Part of the thesis of this dissertation is that *FE* may be getting at something much more fundamental about the way in which the traditional debate is conceived, and thus it deserves further scrutiny. The final two sections of this chapter are devoted to presenting more evidence for *FE,* this time coming from my initial exploratory study and my follow-up 2x2 factorial study.

5.2 Exploratory Study – Revisiting Nichols' Psychopath *Study*

Two years before Björnsson et al. (2015), I ran an exploratory follow-up study to Nichols' *Psychopath*. I identified four features of Nichols' study that could possibly explain Nichols' externalist responses *without having to attribute* a kind of folk externalism to the participants. We've encountered some of these worries already, but reconsidering them will help clarify the structure and content of the exploratory study. First, *Psychopath* employs an extremely small sample size. Any effects found in such samples may wash out when statistical power is increased (and this is *in fact* what seems to have happened when Björnsson et al. (2015) ran their replication of *Psychopath*). Second, *Psychopath* involves a relatively violent protagonist and transgression. It's not unreasonable to suspect that such violence might provoke a retributivist streak within participants, leading them to grant John's grasp of the wrongness of his actions out of a desire to see John punished for his crimes (though, as discussed in the next chapter,

Björnsson et al. (2015) failed to confirm a similar hypothesis). Third, while *Psychopath* focused on amoralist judgments about what's morally *wrong* to do, other studies have focused on what's morally *right* to do. Moreover, the studies carried out by Strandberg and Björklund involved vignettes that focused entirely on a charitable giving scenario. Perhaps cases like charitable giving lead participants to evaluate the judgment as being about a morally *optional* action instead of a morally obligatory (or forbidden) action, preventing straightforward interpretations of externalist responses as supporting folk externalism. And finally (and importantly) *Psychopath* relies entirely on *factive* characterizations of the amoralist.

From these worries, I devised three hypotheses concerning what best explains Nichols' results:

> (1) *Attitude hypothesis*: externalist responses are partly due to an exclusive use of factive attitudes (e.g., 'understands') to describe the amoralist. If this hypothesis is correct, then replacing these factives with non-factives (e.g., 'believes') would be enough to decrease the prevalence of externalist responses.
>
> (2) *Valence hypothesis*: responses may be affected by whether the judgment is of moral rightness as opposed to moral wrongness. If this hypothesis is correct, then replacing wrongness judgments with rightness judgments could be enough to affect the direction of responses (i.e., towards externalism for rightness, internalism for wrongness).
>
> (3) *Folk Moral Internalism hypothesis*: externalist responses in Nichols' *Psychopath* are best explained by (1) and (2) and not a folk commitment to externalism, *contra* Nichols.

Methods

Participants

432 students taking introductory philosophy courses completed a questionnaire during their respective classes. Participants received bonus credit towards their course grade for the class as compensation for their participation in this study.

Procedures and Materials

Subjects were split into four groups. Participants in two of the groups were assigned a Nichols-inspired vignette (psychopath) with one of two *attitudes* (understands or believes). To illustrate, participants assigned to the psychopath believes group received the following vignette:

> John is an adult of normal intelligence. He doesn't care whether hurting or killing people is morally wrong. Indeed, he hurts and even kills people when he wants their money. John has a long history of violent behavior, dating back to childhood and continuing to the present day. Most recently, he shot a woman when she refused to hand over her wallet during a robbery. Still, John says, "I believe that hurting and killing people is morally wrong."

Those assigned to the psychopath understands group received an identical vignette, except the word 'believe' was replaced with the word 'understand' in the final sentence. After reading their vignette, participants in these two groups were presented with a statement concerning whether the amoralist agent in the story really does understand/believe that the transgression is morally wrong. Specifically, participants were asked to indicate to what degree – on a scale from 1 (definitely no) to 7 (definitely yes) – they agreed or disagreed with the following statement:

> Does John really understand/believe that hurting and killing people is morally wrong?

Participants also received four reading comprehension questions: [62]

> Is it stated in the paragraph above that John says he understands/believes that hurting and killing people is morally wrong? Y/N
>
> Is it stated in the paragraph above that John understands/believes that hurting and killing people is morally wrong? Y/N
>
> Is it stated in the paragraph above that John has never hurt or killed anyone? Y/N
>
> Is it stated in the paragraph above that John doesn't care whether hurting or killing people is morally wrong? Y/N

Participants in the other two groups were assigned one of two vignettes (environment and charity). For these two groups, *attitude* was fixed (believes), though the *valence* of the judgment was altered (believes it is *wrong* to hurt the environment; believes it is *right* to give to charity). To illustrate, participants assigned to the environment group received the following vignette:

> Dave is an adult of normal intelligence. He doesn't care whether harming the environment is morally wrong. Indeed, he commutes a long distance to work in a car that gets very poor gas mileage. He realizes that this practice is wasteful and harmful to the environment, and he can easily afford to buy a much more fuel-efficient car. Still, Dave says, "I believe that harming the environment is morally wrong."

Those assigned to the charity group received a similar vignette, except the environmental harm transgression was changed to a case of Chris spending his money on luxuries instead of giving it to charity, and the valence of the judgment was changed to morally

---

[62] The original intent of adding these questions to each study was to cut participants from the study that gave a 'no' to the first question and/or a 'yes' to the second question. While this was carried out for the pilot study, there was an odd trend in responses to the first two questions in the studies following the pilot study. While almost all participants would give the correct "yes" response to the first question, hardly any participants were giving the correct "no" response to the second question. Since counter-balancing didn't correct this trend, it was later determined that this feature of the study design was flawed and should thus not be factored into the results.

right (e.g., Chris says, "I believe that giving my extra money to charity is morally right"). After reading their vignette, participants in these two groups were presented with a statement concerning whether the amoralist agent in the story really does believe that the transgression is morally wrong/right. Specifically, participants were asked to indicate to what degree – on a scale from 1 (definitely no) to 7 (definitely yes) – they agreed or disagreed with the following statement:

> Does Dave/Chris really believe that harming the environment is morally wrong/giving his extra money to charity is morally right?

Participants also received four reading comprehension questions analogous to those given to the participants in the psychopath understands/believes groups. All participants were invited to submit comments after completing the study.

Results

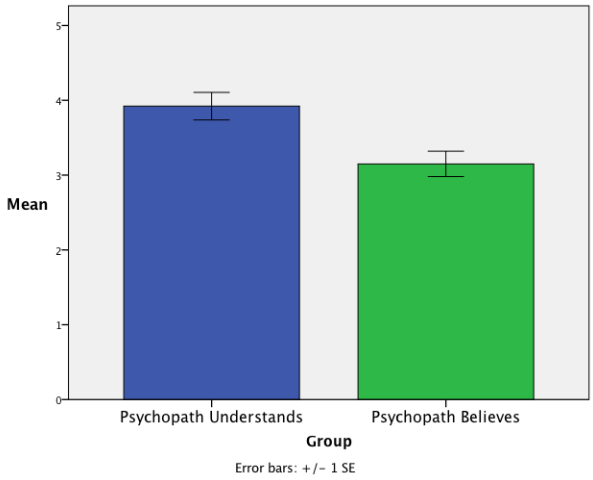
(1) Attitude hypothesis

The culled responses for *psychopath understands* and *psychopath believes* came to 131 responses total *(N=73* and *N=58* respectively).[63] Now if Nichols' results were not affected by the exclusive use of factives within his amoralist scenario, then there shouldn't be any significant change in participant responses when this factive terminology is replaced with non-factive terms. Although the culled responses did not reach the desired significance (*p = .10*), there was still a trend towards the effect predicted on the assumption that the *factive/non-factive hypothesi*s is true. When all

[63] Culled responses are those that came from participants that answered all reading comprehension questions correctly.

responses are taken into account, the effect reaches significance at alpha .05 (*N = 222, p = .001, Cohen's D = 0.41*), as illustrated in the graph below: [64]

Figure 3.1 – Unculled Responses



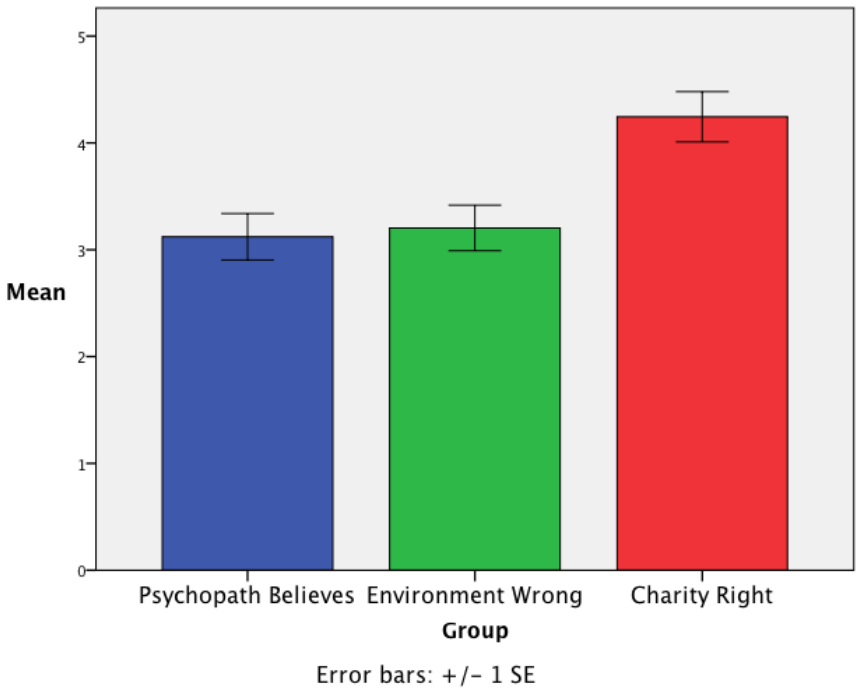
(2) *Valence hypothesis*

The culled responses for *psychopath believes*, *environment wrong*, and *charity right* came to 157 total (*N=58, N=54* and *N=45* respectively). According to the *valence hypothesis*, the direction of responses should be affected by the valence of the moral judgment (e.g., moral rightness judgment in place of moral wrongness judgment). This hypothesis was confirmed; there were two main effects. One effect occurred between *psychopath believes* and *charity right* (*p = .001, Cohen's D = 0.69)*. The direction of the effect is represented by a lean towards internalism in *psychopath believes* (wrongness valence) relative to the externalist-leaning results in *charity right* (rightness valence). Another effect was found between *environment wrong* and *charity right* (*p = .003, Cohen's D =*

[64] A follow-up study conducted on M-Turk did in fact reach significance with culled responses (*N=106, p = .02*). This effect appears again in the 2x2 factorial study (discussed in the next section).

0.66). The direction of the effect seems identical to the previous effect, with responses leaning internalist in *environment wrong* relative to *charity right*. Both effects are illustrated in the graph below:

Figure 3.2 – Culled Responses



(3) *Folk Moral Internalism hypothesis*

The aim of this third hypothesis was to see if (1) and (2) accounted for externalist responses, such that the remaining evidence would suggest a kind of folk internalism, *contra* Nichols. While admittedly this exploratory study isn't specifically suited to test this hypothesis, it was thought at the time that the present results favored folk moral internalism. Over 70% of the responses for *psychopath believes* and *environment wrong* consisted of 3 or lower (i.e., denying the amoralist actually believes) for the test question, as well as over 56% for *psychopath understands*. Intriguingly, only about 34% of respondents gave these internalist-leaning responses in *charity right*. But this seemingly

externalist-favoring result can be plausibly explained in a manner consistent with the folk moral internalism hypothesis, as will be discussed in the discussion section.

Discussion

Though the evidence favoring the *attitude hypothesis* was not particularly robust, there was clearly some evidence that the difference in wording (e.g., replacing 'understand' with 'believe') had an effect on participant responses. It should be recalled that Nichols' *Psychopath* relied entirely on factive attitude descriptions of the amoralist. Given that participants leaned more towards internalism when the amoralist was described with the non-factive attitude of belief, it follows that Nichols was too quick to conclude that his results serve as evidence against moral internalism. At any rate, this attitude effect receives greater scrutiny in the 2x2 factorial study (next section).

The main effects confirming the valence hypothesis are interesting for two reasons. First, as we've seen, both studies in Strandberg and Björklund (2012) and some of the studies in Björnsson et al. (2015) employed charitable giving vignettes.[65] If this valence effect is robust, this may complicate straightforward interpretations of the results of those studies. Second, there happens to be a rather simple explanation for this valence effect that is consistent with folk moral internalism. As we've already discussed, the scope of moral internalism extends only to judgments about what one is morally

[65] This is exactly what Strandberg and Björklund (2012) appear to have done. Each study they present focuses exclusively on a judgment concerning moral *rightness,* a charity case similar to *charity right.*

obligated to do (or *forbidden* to do).[66] But one of the most commonsense examples of *non*-obligatory moral action is charitable giving. So perhaps the reason participants leaned externalist in *charity right* was because they likely conceived of the judgment in terms of supererogation as opposed to moral obligation[67].

Interestingly enough, some comments from participants in the *charity right* group appear to confirm this idea that the folk conceive of moral rightness (or, at least, judgments about giving to charity) primarily in supererogation terms. Here is one comment from someone who correctly answered all comprehension questions and gave an externalist-leaning response to the test question in *charity right*: "It seems that Chris is aware of the strong possibility of giving to charity is morally right. [sic.] But that doesn't mean he has to, or feels obligated to act on it." [68]

Because this exploratory study arose out of a response to Nichols' *Psychopath*, much of its construction is shaped by the setup of *Psychopath*. While this allows its results to connect up nicely with the issues raised against *Psychopath*, its narrow scope makes drawing more general conclusions difficult. So although this study did confirm

---

[66] At least, it is judgments concerning moral obligation that seem to be at issue within the traditional metaethical dispute over moral internalism and amoralist skepticism.

[67] And as mentioned earlier in this chapter, there is now evidence that people are likely to conceive of it this way.

[68] Another possibility is that there is a folk distinction between perfect and imperfect duties, as per Kantian ethics. While a perfect duty must be carried out precisely when it applies, and in the way specified by the duty, imperfect duties are not so time-sensitive (nor specific in how they should be carried out). Thus, one *can* judge themselves to have an imperfect duty even if they aren't motivated at that time to carry it out, but this needn't serve as a counterexample to internalism precisely because the generous time allowances of imperfect duties are *built-in*, as it were. This distinction allows motivation to be temporarily absent despite awareness of one's imperfect duties without conflicting with the internalist thesis. I owe this alternative explanation to David Braun, who suggested it during the Q&A session at the 2012 Buffalo Experimental Philosophy Conference.

both the *attitude hypothesis* and the *valence hypothesis*, such results would be dialectically more powerful if they could be replicated with a non-derivative study. To this end, a stand-alone study was constructed to more clearly test both the *attitude hypothesis* and the *valence hypothesis*. This study is discussed in the following section.

6. 2x2 Factorial Study

Background

The results of the exploratory study seem to suggest that the initial criticisms regarding Nichols' *Psychopath* were basically correct. The more-or-less replication of Nichols' *Psychopath* in *psychopath understand* and *psychopath believe* groups provided some evidence that Nichols' use of the factive 'understand' may indeed have played a crucial role in determining the degree to which his participants granted that his amoralist was making genuine moral judgments. By simply replacing 'understand' with 'believe,' participants in the exploratory study significantly leaned away from ascribing the moral judgment to the amoralist. There was also a concern over how the valence of the transgression might affect participants' intuitions. In particular, Strandberg and Björklund's focus on a case of charitable giving within their studies' vignettes seemed problematic insofar as participants view charitable giving as a paradigm case of supererogation. So in these cases, people might agree that it's nice to help but it's not your moral *obligation* to help. But if this is right, then externalist responses evoked from these cases wouldn't be straightforward evidence for folk externalism, because people could still be internalist regarding moral obligation. This outcome seems to be supported by the results of the exploratory study.

Nevertheless, there's a sense in which the above evidence for the *attitude hypothesis* and *valence hypothesis* are a bit too piece-meal and messy to draw any strong conclusions. What is needed is a study designed specifically to test how attitude ascription (factive/non-factive) and transgression valence (rightness/wrongness) may (or may not) interact and affect folk intuitions regarding amoralists and the nature of moral judgment. To this end, a 2x2 factorial study was designed to test the effect of the two kinds of attitude ascriptions at the two different valence levels. This created four possible scenarios: (a) the amoralist understands that *X* is morally right to do, (b) the amoralist believes that *X* is morally right do, (c) the amoralist understands that *Y* is morally wrong, and (d) the amoralist believes that *Y* is morally wrong. With these four scenarios, the following study was designed and carried out.

Hypothesis

The general hypothesis is that externalist responses track factive attitude descriptions and rightness valences whereas internalist responses track non-factive attitude descriptions and wrongness valences. This can be called the *attitude/valence hypothesis*. If this hypothesis is correct, then the following 2x2 factorial study should reveal a significant difference in participant responses along the lines of the direction found in the exploratory study. Specifically: responses should lean internalist when the amoralist scenario concerns non-factive attitudes (e.g., believes) and moral wrongness judgments, whereas responses should lean externalist when the amoralist scenario concerns factive attitudes (e.g., understands) and moral rightness judgments (no effect is expected for the factive/wrong and non-factive right groupings).

Methods

Participants

288 participants[69] completed a study on Amazon Mechanical Turk, spanning a wide age range. The mean age was 31; 185 female, 103 male. Participants received $0.30 for their participation.

Materials and Procedures

Subjects were split into four groups. Participants in each group were assigned two *vignettes* (video game and headphones – in that order), but only one of two *attitudes* (understands or believes), and only one of two *valences* (wrong or right). To illustrate, participants assigned to the *believes-wrong* group received the following two vignettes:

> Chris is a 21-year-old undergraduate student. He and his friend Bill are in line to buy a video game. At the counter is a charity donation jar with only $30 in it. Chris needs $20 for the game, but Bill argues that taking it from the jar is morally wrong. After listening to Bill's argument, Chris says, "I agree, Bill - I believe that taking money from the jar is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Chris buys the video game with $20 from the jar.
>
> Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - I understand that buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones.

[69] 45 participant responses were cut from the data due to redundancies (i.e., participants taking more than one survey). The sample size for all of the studies carried out was determined by reference to G*Power 3.1.7 software.

Those assigned to the *understands-wrong*, *believes-right*, and *understands-right* groups received similar vignettes, except minor adjustments were made to reflect the change in attitude and valence. After each vignette, participants were presented with a series of statements (counter-balanced) concerning whether the amoralist agent in that story understands/believes that the action in question in the story is morally right/wrong. Specifically, participants were asked to indicate to what degree – on a scale from 1 (definite no) to 7 (definite yes) – they agreed or disagreed with the following statements:

> Chris believes that taking money from the jar is morally wrong.
>
> Chris believes that taking money from the jar is unethical.
>
> Chris believes that he shouldn't take money from the jar.
>
> Jane believes that buying the headphones at the mistaken price is morally wrong.
>
> Jane believes that buying the headphones at the mistaken price is unethical.
>
> Jane believes that she shouldn't buy the headphones at the mistaken price.[70]

Results

[70] Participants also received two reading comprehension questions for each vignette: for video game: 'Is it stated in the paragraph above that Chris says the following: "I believe that taking money from the jar is morally wrong"? Y/N' and 'Is it stated in the paragraph above that Chris believes that taking money from the jar is morally wrong? Y/N'; for headphones: 'Is it stated in the paragraph above that Jane says the following: "I believe that buying the headphones at the mistaken price is morally wrong"? Y/N' and 'Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N' As discussed in footnote 62, the responses to these questions were dropped from the study due to what appeared to be evidence of a flaw in the design of one of the questions. From this point on, while participants continued to answer these comprehension questions in the studies that follow, these responses were not used to cull or adjust the response to the other questions in the studies.

If the *attitude/valence hypothesis* is correct, then participants in *believes wrong* should be significantly more likely to lean internalist relative to those in other groups (i.e., answering 3 or lower for all test questions). More generally, participants in both 'believes' groups should be significantly more likely to lean internalist relative to those in the 'understand' groups. While the attitude aspect of the hypothesis was confirmed for each vignette, the valence aspect was not: (*video game* Items Averaged attitude: $F(1,288)=61.41$, $p<.001$, $\eta^2 = .178$, $\alpha = .68$; *video game* Items Averaged valence: $F(1,288)=1.42$, $p=.23$, $\eta^2 =.005$; $\alpha = .68$; *headphones* Items Averaged attitude: $F(1,288)= 96.18$, $p<.001$, $\eta^2 = .253$, $\alpha = .64$; *headphones;* Items Averaged valence: $F(1,288)= 7.48$, $p=.007$, $\eta^2 =.026$, $\alpha = .64$). There was also no interaction effect between attitude and valence. The Items Averaged results for both *video game* and *headphones* are illustrated in the graph below:

Figure 3.3 – video game – Items Averaged

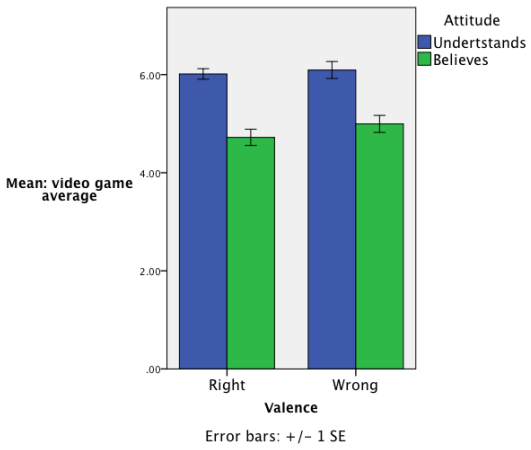
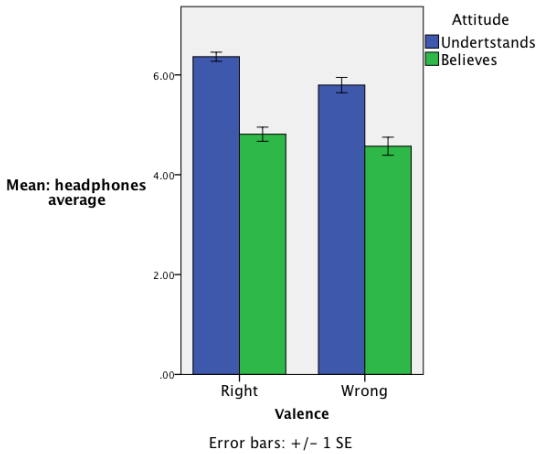
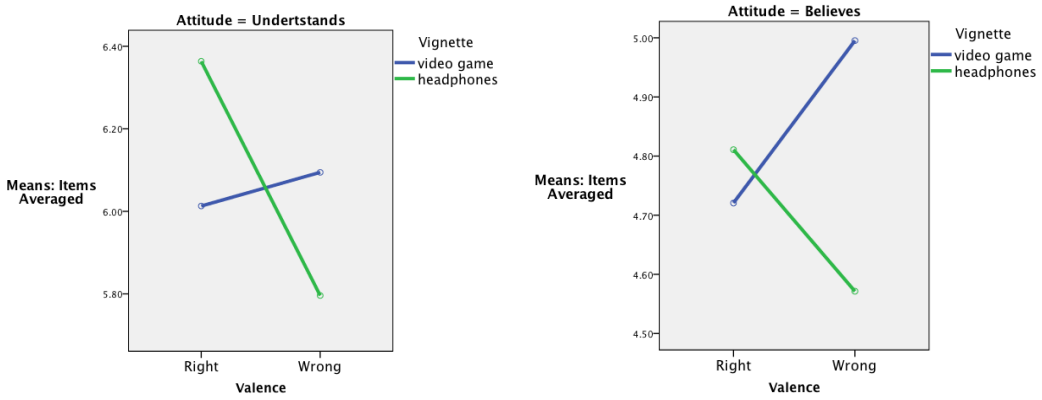


Figure 3.4 – headphones – Items Averaged



There was an interaction effect between vignette and valence, illustrated in the graphs below:

Figure 3.5 – Interaction Effect – Items Averaged



Participants seemed more certain that Chris (the protagonist in *video game*) believed that taking money from a charity to buy a video game was morally wrong than they were that Jane (the protagonist in *headphones*) believed that purchasing the headphones at the mistaken price was morally wrong. This interaction effect could be due to a gender confound, given that the video game vignette involved only males while the headphones vignette involved only females. Another possible explanation for this interaction concerns the folk consensus concerning the moral status of the respective behaviors. In the case of Chris taking money from a charity, it seems likely that the vast majority agree that this is deeply morally wrong, and thus will assume that – all things being equal – Chris would believe it is wrong as well. However, in the case of Jane telling the clerk about the mistaken price, it's not obvious that everyone agrees that her behavior is morally wrong (perhaps they see it as morally optional).[71] Given that the effect size for

[71] There is indeed evidence that people do in fact view one's response to these kinds of mistakes as simply being morally *optional*. The following National Public Radio article highlights this issue: http://www.npr.org/blogs/thetwo-way/2013/12/27/257552179/the-price-is-wrong-and-you-know-it-do-you-buy-that-ticket. The polls at the bottom of the article show that the majority of respondents view the purchasing of *obviously* erroneously cheap flights as morally okay!

the interaction was relatively small, this seems like a plausible account of this interaction effect.

Discussion

The valence aspect of the hypothesis failed to be confirmed, though it's unclear whether or not this failure was due in part to the presence of interaction effects. Perhaps if more carefully constructed scenarios were used, the valence effect would appear, as it did in the exploratory study. But at this point, it seems best to focus on what was confirmed by the 2x2 study—that a change in attitude does impact participant evaluations of amoralists.

The attitude aspect of the hypothesis was confirmed. On one interpretation of the results, participants felt more certain that the amoralist *understands* that the behavior is morally right/wrong than that the amoralist *believes* this about said behavior. This asymmetry in attitude ascriptions to the amoralist occurred in the same direction for all four scenarios:

> *Understands/Believes Right*: participants felt more certain that Chris *understands* that it is right to give to charity while less so about ascribing to Chris the *belief* that such charitable giving is morally right; participants felt more certain that Jane *understands* that she should notify the clerk about the mistaken price while less so about ascribing to Jane the *belief* that she should notify the clerk about the mistaken price.
>
> *Understands/Believes Wrong:* participants felt more certain that Jane *understands* that it is morally wrong to purchase the headphones at the mistaken price while less so about ascribing to Jane the *belief* that this behavior is morally wrong; participants felt more certain that Chris *understands* that taking money from a charity to buy a video game is morally wrong while less so about ascribing to Chris the *belief* that this action is morally wrong.

It appears that people's intuitions lean externalist when the amoralist's moral judgment is described *factively* (e.g., understands that she shouldn't do *X*), whereas people's intuitions lean relatively internalist when the amoralist's judgment is described *non-factively* (e.g., believes that she shouldn't do *X*).

This is the very same effect found in Björnsson et al. (2015)—what I'm calling the *Factivity Effect* (*FE*). That *FE* has now occurred across various vignettes and studies speaks to its robustness. Thus, it seems appropriate to more clearly define *FE* as follows:

> **Factivity Effect** (*FE*): people's intuitions lean externalist when evaluating factive amoralist scenarios, whereas people's intuitions lean relatively internalist when evaluating non-factive amoralist scenarios.

While its robustness has been established, there remains the question of *why FE* occurs. We can distinguish two types of explanations: deflationary explanations and substantive explanations. Deflationary explanations identify features of the study itself as the ultimate cause of the effect. Explaining *FE* as a mere experimental artifact is to give a deflationary explanation. Substantive explanations, on the other hand, identity features of participant thought as the ultimate cause of the effect. Explaining *FE* as a feature of the structure of folk thought is to give a substantive explanation. Why *FE* occurs, and what implications *FE* may have for the traditional debate over moral internalism and amoralist skepticism, is the subject of the next chapter.

7. Conclusion

The rise of experimental philosophy has led to widespread experimental investigation into the nature of philosophical intuitions across any area of philosophy that relies on, or appeals to, intuitions about cases and thought experiments. The traditional dispute over

moral internalism and amoralist skepticism within metaethical inquiry is one such area. Thought experiments constructed by both externalists and internalists attempt to draw on our philosophical intuitions about whether an entirely unmotivated agent could make genuine moral judgments. However, perhaps unlike other areas of experimental inquiry, experimental research over moral internalism is very much in its infancy. Only a handful of publications make up the extant experimental literature on moral internalism. But while this research has just begun, a particularly interesting effect has been discovered, which I'm calling the *Factivity Effect*: people's intuitions lean externalist when evaluating factive amoralist scenarios, whereas people's intuitions lean relatively internalist when evaluating non-factive amoralist scenarios. Having been replicated numerous times, we have seen that *FE* is a particularly robust effect. Evidence that *FE* reflects a genuine psychological feature of the structure of folk thought has the potential to overhaul how metaethicists conceive of the externalism/internalism debate and the nature of moral judgment itself.

Chapter 4: What Explains the *Factivity Effec*t?

1. Introduction

In chapter 3, it was suggested that the traditional discussion over moral internalism and amoralist skepticism likely involves a tacit assumption to what I'm calling *Descriptive Equivalence*, or *DE*:

> **Descriptive Equivalence** (*DE*): factive descriptions of the amoralist are equivalent to non-factive descriptions of the amoralist, in the sense that they can be interchanged without loss (or gain) in the scenario's overall conceptual coherence.

It was also shown that recent experimental work in this area has uncovered a rather surprising effect that appears to be in tension with *DE*: people's intuitions lean externalist when the amoralist's moral judgment is described *factively* (e.g., understands or knows that she shouldn't do *X*), whereas people's intuitions lean relatively internalist when the amoralist's judgment is described *non-factively* (e.g., believes or thinks that she shouldn't do *X*). I'm calling this the *Factivity Effect* or *(FE)*:

> **Factivity Effect** (*FE*): people's intuitions lean externalist when evaluating factive amoralist scenarios, whereas people's intuitions lean relatively internalist when evaluating non-factive amoralist scenarios.

FE seems to raise a challenge to *DE*, thus challenging a foundational assumption within the traditional debate over moral internalism and amoralist skepticism. In this chapter, I argue that *FE* is unlikely to be explained away as an experimental artifact; as a consequence, the traditional dispute over moral internalism and amoralist skepticism may need an overhaul. The results of studies testing four different explanations of *FE* arguably provide support for this thesis. I conclude this chapter by discussing possible implications of these results for the traditional dispute.

In section 2, I discuss two attempts at explaining *FE*. In section 3, I present three studies that test the notion that *FE* is merely an experimental artifact. In section 4, I present a final study that tests the notion that *FE* reflects a feature of the structure of folk thought. In section 5, I consider three possible philosophical ramifications of *FE* reflecting a division within folk thought. In section 6, I respond to two possible defenses of *DE* against the experimental results. I end the chapter by highlighting some limitations of the present work that could be addressed in future research.

2. Two Attempts at Explaining the Factivity Effect.

One may recall from the previous chapter that one of the initial studies to produce *FE* came from Björnsson et al. (2015). In that study, participants were given a description of how we ordinarily classify actions as morally right and morally wrong, and then they were introduced to the amoralist Anna whom is said to make these ordinary moral judgments but yet those judgments don't "influence her choices." Participants were then given a vignette involving Anna having to choose between two cell phones that are identical in price and type, except one phone includes a free donation of $20 to help address starvation in Sudan. When participants were asked if Anna *understands* that it is morally wrong not to choose the donation phone—despite her choosing the non-donation phone—a large majority said she did. But when participants were asked if Anna *believes* or *herself thinks* that it is morally wrong not to choose the donation phone, less than half of the participants agreed that she did. This is an instance of *FE*—participants are granting the factive mental state (understands) to the amoralist but not the non-factive mental states (believes; thinks). But why is *FE* occurring? Björnsson et al. (2015) tested

two possible explanations for why this effect occurs: the *blaming hypothesis* and the *inverted commas hypothesis*.

2.1 The Blaming Hypothesis

Björnsson et al. (2015) note that there is likely a background connection between moral accountability and moral understanding, given that one's accountability is mitigated if it can be shown that one lacks sufficient understanding of their behavior and its consequences.[72] Given this background connection between accountability and moral understanding, people may be inclined to attribute understanding to those that they wish to hold fully morally accountable. If participants were so inclined concerning Anna, this could explain why they granted that she understands that choosing the non-charity cell phone is morally wrong.

To test this *blaming hypothesis*, the team constructed two scenarios to add to the original study's cell phone scenario. In one scenario, Anna knowingly shoots and kills a random hiker. In another scenario, Anna knowingly shoots and damages a motorcycle. They adjust the test question from the original study to fit these new scenarios:

> We have seen that Anna classifies some actions as "morally wrong." But because she lacks compassion and is strikingly egoistic, this never makes her even the least inclined to avoid these actions. We saw this indifference when she chose her target. In light of this, would you say that she [understands/believes] that it is morally wrong not to choose the cactus?

---

[72] In defense of this assumption, consider that jurors apparently take this into account when trying to determine if an insanity defense seems plausible. See the following NPR article for an instance of this: http://www.npr.org/sections/thetwo-way/2015/07/16/423610640/aurora-colo-theater-shooter-james-holmes-found-guilty?sc=17&f=1001&utm_source=iosnewsapp&utm_medium=Email&utm_campaign=app

If the *blaming hypothesis* is correct, there should be significantly more attributions of understanding to Anna in the hiker scenario when compared to both the original cell-phone scenario and the new motorcycle scenario.

The results did not go as expected: participants ascribed understanding *less* in these two scenarios than they did in the original cell phone scenario, despite these actions being presumably more blameworthy![73] Given this odd result, they wondered if something about the new scenarios may have counter-acted the blaming motivation. They ran another follow-up study to test this concern, this time simply adding a test question to their original study which would rank how blameworthy participants viewed Anna on a 7-point Likert scale '1' = 'not at all' to '7' = 'completely.' They predicted a positive correlation between understanding attributions and higher rankings of blame attributions. However, this prediction failed as well.

2.2 The Inverted Commas Response Hypothesis

Within the traditional literature, some internalist philosophers give an error-theoretic account of externalist intuitions called the *inverted-commas response* (*ICR*). In its traditional formulation, *ICR* involves claiming that alleged amoralists use moral language only in an inverted commas sense. Rather than making an evaluative moral judgment about some behavior, an amoralist is merely making a descriptive judgment concerning the general opinion of that behavior. On this account, for example, if the amoralist utters the sentence 'torturing babies for fun is morally wrong,' the amoralist is merely referring

[73] Ibid, p.726.

the *general consensus* that torturing babies for fun is morally wrong, *not* the moral fact itself. *ICR* can thus be summed up as two claims:

> (1) Amoralists can't really make evaluative moral judgments.
>
> (2) When using moral language, amoralists make descriptive judgments about the general moral consensus (thus, externalist intuitions must be tracking this, if anything).

While an argument can be made for weakening (2),[74] let's focus for now on this traditional formulation of *ICR*.

As it happens, *ICR* suggests a ready explanation of *FE*. Confusing descriptive judgments regarding the general moral consensus with evaluative moral judgments proper might be what leads participants to grant understanding to the amoralist, even if their conception of moral judgment is thoroughly internalist. In more detail, *FE* can be explained by the following set of claims:

> (a) the folk conception of moral judgment is internalist.
>
> (b) given *ICR*, factives like 'understand' in experimental studies *mis*lead participants by confusing knowledge of the general moral consensus with the amoralist's evaluative moral knowledge.
>
> (c) Non-factives like 'believe' *correctly* lead a substantial number of participants[75] to deny genuine moral judgment to amoralists.
>
> (d) (Thus) *FE* is an experimental artifact; factives elicit artificial externalist responses due to the ambiguity between the two different senses of 'understand' in this context.

---

[74] See the discussion section in 3.1 for one way to do this.

[75] There is, of course, the question of what to make of the sizable minority of participants that *do* ascribe belief to the amoralist. Presumably all experimental studies within any domain are going to have some unexpected results. When we're reasonably optimistic about the study, these unexpected results are traditionally called *noise* – unanticipated variations in the data that do not necessarily tell against the effect borne out by the data. Thus, a first pass at explaining the divergent responses in these studies is to attribute such responses to noise.

This explanation of *FE* can be called the *ICR-hypothesis*.

So if we recall the original Anna study, one straightforward explanation of *FE* is that participants who grant understanding and belief to Anna are merely granting an "inverted-commas" sense of these mental states and not the real deal. In effect, they may be simply granting that Anna understands or believes that *others view* choosing the non-charity cell phone as morally wrong, though she herself doesn't really think this.

To test the *ICR-hypothesis*,[76] Björnsson et al. (2015) replaced their characterization of Anna as a psychopath with a description of Anna as someone that only uses moral language in an inverted-commas sense. To achieve this, they added the following to their original setup:

> When she [Anna] classifies acts into "morally right" and "morally wrong" she does so only on the basis of how most other people in her society use these expressions. For example, she classifies murder and theft as "morally wrong" only because she thinks that most other people in her society do so.[77]

Given that Anna's inverted-commas sense use is now made explicit, they predicted that almost every participant should ascribe *this* sense of understanding quite easily (it's stipulated, after all!) while then going on to deny—in greater numbers—that Anna believes or herself thinks that purchasing the non-charity cell phone is morally wrong. Essentially, they predicted that the original effect (*FE*) would become even more pronounced given this inverted-commas manipulation.

---

[76] Björnsson et al. (2015) call it the inverted commas hypothesis, but to reduce confusion, I'm using my label for their study as well.
[77] Ibid., p.723.

But the results didn't go as they predicted. The difference between understanding responses and believe responses was "reduced to non-significance."[78] Participant responses that granted understanding to Anna were significantly lower than in the first study. To explain this, Björnsson et al. suspect that making the inverted-commas sense explicit produced the *opposite* effect on participant evaluations of Anna. While making Anna's inverted-commas sense more clear, their manipulation may have inadvertently presented a clearer contrast with *genuine* understanding, making the real deal more salient to participants. As a result, participants on the whole simply denied that Anna has this genuine sense of understanding. To add to this explanation, they note the significantly lower responses they received for the "herself thinks" variation relative to their original study.[79] They speculate that the inverted-commas manipulation again made the real deal more salient by contrast, leaving participants to think that perhaps Anna is just looking to others when she judges rather than thinking for herself. The above explanation of these results presupposes a kind of folk internalism; the suggestion is that participants deny Anna understanding and belief when the genuine (relevant) sense of these states are particularly salient. Thus, Björnsson et al. (2015) take these results, despite failing to confirm their inverted-commas hypothesis, as evidence for folk internalism.

---

[78] Ibid., p.724.

[79] Belief responses, however, remained unchanged when compared with the previous study. They take the lack of movement in belief responses, compared with the movement down for both "understands" and "herself thinks" responses, as evidence that ascriptions of belief are parasitic on something else, perhaps similar the inverted-commas sense. I'm not sure if I understand their interpretation here, but they think this favors internalists accounts of moral judgment.

With neither of their explanations of the *Factivity Effect* confirmed, they offer a final speculation: perhaps belief requires some further mental condition, such as "endorsement" or "mental assent" that isn't required for understanding, making understanding more easily ascribed than belief. Nevertheless, they admit that this hypothesis would be rather difficult to test.[80]

We have seen two deflationary explanations offered for *FE* by Björnsson et al. (2015): the *blaming hypothesis* and the *inverted-commas hypothesis*. In calling these hypotheses *deflationary*, I mean to mark them as explaining *FE* by appeal to some deflationary cause (e.g., an experimental artifact or a performance error). In this case, both hypotheses attempt to identify a feature of the experiment that may have confused participants or primed them to respond in ways that result in *FE*. According to the *blaming hypothesis*, it was thought that a background connection between moral understanding and moral accountability was leading participants to grant understanding in order to hold the amoralist accountable for her transgression. According to the *inverted-commas hypothesis*, the use of 'understand' was thought to have been ambiguous, such that perhaps participants that granted understanding to the amoralist were only meaning to grant that the amoralist understands that *others view* certain actions

[80] Ibid., p.726-27. One possibility is that participants are simply answering randomly, with the results being a kind of fluke or coincidence. Björnsson et al. test this concern by altering their original study to make it obvious that Anna clearly understands and believes that purchasing the non-charity cell phone is morally wrong. To achieve this, they first have the phones be different colors – the charity phone is white and the non-charity phone is black, and then they stipulate that Anna really wants a black phone. Out of weakness of will, she purchases the black, non-charity phone over the white, charity phone. If participants are paying attention – as opposed to answering randomly – then they should ascribe both understanding and belief to Anna in this weakness-of-will scenario. Results go as they predicted: 85% ascribed understanding to Anna and 80% ascribed belief.

as morally wrong or right. In each case, it was thought that controlling for the confusion or priming would reveal *FE*'s dependence on these features of the original experiment. But neither of these hypotheses received confirmation.

In the following section, I present three studies, each testing more deflationary explanations for *FE*. I begin with another attempt at testing the *ICR-hypothesis*.

3. Three Deflationary Explanations of the Factivity Effect.

3.1. Deflationary Explanation #1: The Inverted Commas Response (ICR) Hypothesis

Background

Recall that the *ICR-hypothesis* explains *FE* by treating the denial of non-factive mental states to the amoralist as resulting from a kind of folk internalism, whereas the ascriptions of factive mental states are explained as the result of participants confusing sociological judgments concerning the moral consensus versus genuine moral judgments proper. We have seen one attempt from Björnsson et al. (2015) at testing the *ICR-hypothesis*: explicitly stipulate that the amoralist makes only inverted-commas moral judgments, and then predict that *FE* should be even *more* pronounced, given that now there would be obvious reason to ascribe understanding and deny belief. But as we saw, this is not what happened. Participants instead began denying all three—understanding, belief, and herself thinks—resulting in *FE* disappearing.

So why didn't the results come out how they predicted? The *ICR-hypothesis* seems like a perfectly good explanation of *FE*, particularly because of its origins in the traditional literature on moral internalism and amoralist skepticism. One possibility is that the participants took their inverted-commas stipulation to apply to *all* moral judgments

made by Anna, not just the factive ones like her understanding that choosing the non-donation cell phone is wrong. If participants think they are being told to assume that Anna always considers what the norm is when she is asked about whether she understands, *believes* or even *herself thinks* that something is morally right or wrong to do, then wouldn't this explain why *FE* disappeared?[81] Not in this case, at least. Because if this is what happened, responses to all three variations of the test question should have been significantly higher than in the original study. Yet the opposite result occurred: 'yes' answers to both *understands* and *herself thinks* variations of the test question were significantly lower as a matter of frequency (responses to the belief variation remained the same).[82]

Another possible explanation for why Björnsson et al. (2015) failed to confirm the *ICR-hypothesis* points to their generally direct, explicit approach. If factives are ambiguous in this context – as *ICR* seems to predict – then creating further senses of 'understand' seems unwise. Explicitly stipulating that 'understand' will now refer to the inverted-commas sense seems too demanding (recall that participants must first learn about *ICR* from the study itself). However, this is *not* to say that the folk lack the relevant competence for determining whether amoralists make genuine moral judgments, any more than the denial that the folk have much of an explicit grasp of the nature of grammar must imply that they lack the relevant competence for determining whether

[81] Thanks to an anonymous reviewer of a publication draft of this chapter for this explanation.
[82] Ibid., p.724.

certain sentences in their native language are grammatical.[83] Rather, the concern is simply that the explicit approach may have actually done more to create ambiguity than to clarify it.

An implicit approach would instead eliminate, in the participant's mind, the possibility that the amoralist has any knowledge about the general moral consensus. If successful, this should focus participant attention on moral judgment. If achieved, and if the *ICR-hypothesis* is correct, then participants should deny *both* understanding and belief to the amoralist; *FE* should disappear. Such a result would provide evidence for the *ICR-hypothesis*, and thus go some way towards deflating *FE* and its impact on metaethical inquiry. The following study uses this indirect approach to test the *ICR-hypothesis*.

Two Preliminaries

First: to prevent the participants from thinking the amoralist has knowledge of the general moral consensus, it was important to select actions that currently lack any moral consensus. Two actions were chosen to serve this purpose: ending a pregnancy simply to avoid parenthood and smoking marijuana recreationally. Showing that the amoralist recognizes the lack of a moral consensus on these cases was thought to be sufficient to block participants from attributing the inverted-commas sense of understanding.

Second: because factive terminology suggests (semantically or pragmatically) that something is actually the case, it would have been infelicitous to ask participants whether the amoralist understands that *X* is morally wrong if the participants *themselves* deny that

[83] Note that it needn't follow from this that there is an innate moral grammar analogous to the Chomskian notion of innate grammar. See Alfano and Loeb (2014) for further discussion of this analogy.

X is morally wrong. Thus, it was necessary to devise a sorting procedure to remove participants from the data who did not personally view the behavior in question as morally wrong.

The following study tests the *ICR-hypothesis*. If this hypothesis is right, then *FE* should disappear when participants evaluate an amoralist confronted with an action that (even the amoralist realizes) lacks any moral consensus.

Method

Participants

400 participants completed a study on Amazon Mechanical Turk, spanning a wide age range, 77 of which made the cut (see below).[84] The mean age was 35; 31 female, 46 male. Half of the participants received $0.15 for their participation, the other half received $0.30 (please see footnote for explanation).[85]

Procedures and Materials

[84] Because these studies were between-subject, multiple responses from the same IP address were cut. The cutting procedure was effectively random; the IP addresses for all responses in the study were ordered from lowest to highest, and the first response of a set of identical IP addresses was kept while the others were cut. For this study, 18 responses were cut due to redundant IP addresses.

[85] Given the large amount of responses cut because participants didn't' personally view either of the moral transgressions as morally wrong, the study was run a second time to make up for this dramatic decrease in statistical power. And because participants were, on average, taking longer to complete the survey than anticipated, the compensation was increased on the second run.

Subjects were split into four groups.[86] Participants in each group were assigned one of two *vignettes* (abortion or marijuana) and one of two *attitudes* (understands or believes). To illustrate, participants assigned to the abortion groups received the following vignette:[87]

> Jane is a 24-year-old graduate student. She and her friend Sandy are outside a clinic that performs abortions. Jane intends to end her pregnancy because she doesn't want to be a parent, but Sandy argues that an abortion for that reason is morally wrong. Jane replies that it's controversial whether abortion to avoid parenthood is immoral, and Sandy concedes that there's currently no consensus on the matter, but Sandy continues to make her case. After listening to Sandy's argument, Jane says, "I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane enters the clinic and a doctor ends her pregnancy as she requests.

Those assigned to the marijuana groups received a similar vignette, except the abortion transgression was replaced with a case of recreational marijuana use and a discussion between Chris and Bill on its moral status. After reading their vignette, participants were presented with a series of statements concerning whether the amoralist agent in the story understands/believes that the transgression is morally wrong. Specifically, participants were asked to indicate to what degree – on a scale from 1 (strongly disagree) to 7 (strongly agree) – they agreed or disagreed with the following statements:[88]

---

[86] Note that these studies all employ *between-subject* designs. *FE* also occurred in a within-subject variation of a preliminary study, with test questions counter-balanced to avoid possible order effects.

[87] For the complete materials for this study, please see the appendix.

[88] Participants received two reading comprehension questions: 'Is it stated in the paragraph above that Jane says the following: "I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong"? Y/N' and 'Is it stated in the paragraph above that Jane understands/believes that ending her pregnancy to avoid parenthood is morally wrong? Y/N' The original intent of adding these questions to each study was to cut participants from the study that gave a 'no' to the first question and/or a 'yes' to the second question. While almost every participant gave a "yes" to the first question, hardly any participants gave a "no" to the second

> Jane understands/believes that ending her pregnancy to avoid parenthood is morally wrong.
>
> Jane understands/believes that ending her pregnancy to avoid parenthood is unethical.
>
> Jane understands/believes that she shouldn't end her pregnancy to avoid parenthood.

Finally, participants indicated their agreement/disagreement with a statement concerning their personal view of the transgression in the story:

> In my personal view, it's morally wrong for anyone to end their pregnancy to avoid parenthood.

Responses were cut from the data if the participant gave anything lower than a '5' on this personal view statement – the lowest degree of agreement allowed by the scale.

Results

Out of 400 participants, only 77 gave a response of 5 or above on the personal view statement (marijuana, *N=25*; abortion, *N=52*). If the *ICR-hypothesis* is correct, then these participants should have denied both understanding and belief to the amoralists. If they had done this, *FE* would have disappeared, thus providing some reason to think *FE* is simply an experimental artifact.

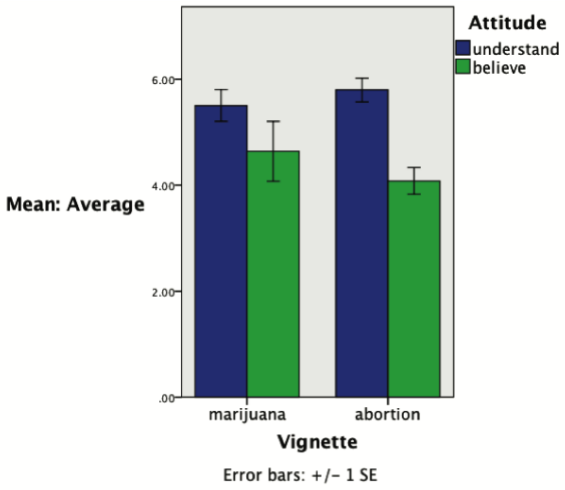
But this is not what happened. Even with the dramatic loss in power resulting from the cut in responses to the personal view statement, *FE* was still detectable (Items

---

question. For instance, only 9 participants answered the second question correctly, whereas only 2 failed the first. Counterbalancing the order of the presentation of the questions did not affect this result. After this occurred for all three studies, it was decided that there was a problem with the question itself (perhaps an ambiguity in the word 'stated'?). For this reason, responses to these questions were ignored for all three studies (though the use of the comprehension question in the moral emotions study was mainly to help focus participant attention on which word – factive or non-factive – was being employed in the vignette).

Averaged: $F(1,77)=16.32$, $p<.001$, $\eta^2 = .183$, $\alpha = .72$). The Items Averaged results are illustrated in *Figure 4.1*:[89]

Figure 4.1 – Items Averaged



Discussion

FE remained despite the manipulations. Participants continued to attribute moral understanding, but deny moral belief, to the amoralist. This result conflicts with the *ICR-hypothesis*, thus providing no support towards a deflationary explanation of *FE*.

One might reformulate *ICR* by removing any positive claims to the nature or content of the judgment the amoralist succeeds in making.[90] The *ICR-hypothesis* could then be reformulated to reflect this more minimal explanation of *FE*: participants mistakenly grant understanding to the amoralist for *some* reason, not necessarily because they confuse it with sociological knowledge of a moral consensus.

[89] While there was no significant difference between the two attitude levels for the marijuana factor, this is likely because the sample size was only 25, which is admittedly quite small.

[90] Here's Smith (1994): "the very best we can say about amoralists is that they try to make moral judgments but fail." p.64.

But what this minimal version gains in plausibility, it lacks in testability. Since nothing definite is said about where the mistake may lie, it's difficult to translate this explanation into a testable hypothesis.[91]

At any rate, this first follow-up study failed to confirm one plausible, though deflationary, account of *FE*. The next follow-up study tests a different deflationary explanation that relies on assuming a kind of folk *externalism*.

3.2 Deflationary Explanation #2: The Dispositional Belief Hypothesis

Background

A deflationary explanation for *FE* can be drawn from recent experimental work on what is called the *entailment thesis*: knowledge entails belief. Myers-Schulz & Schwitzgebel (2013) offer counterexamples to the entailment thesis by way of four studies. One study, called *unconfident examinee*, resulted in a significant majority of participants ascribing to Kate (the examinee) the knowledge that Queen Elizabeth died on 1603 *while denying* that

---

[91] An anonymous reviewer from a publication version of this chapter offers another possibility: the reference of 'morally wrong' in the mouth of Jane needn't be the general consensus on what's morally wrong, but rather what the general population *ought* to believe given society's commitments on other moral matters. For example, suppose Jane knew that the general consensus on infanticide was that it was morally wrong, and then suppose Jane believed that, for people to have *consistent* moral views, they should then view abortion as morally wrong. If participants attributed something like this reasoning to Jane, they may still have interpreted Jane's use of 'morally wrong' in an inverted-commas sense *even if* my attempt at blocking participants from attributing knowledge of a general consensus about abortion to Jane was successful.

One worry about this explanation is that it requires attributing a relatively complex reasoning process to participants. To conclude that Jane may still be using 'wrong' in an inverted-commas sense while granting Jane's knowledge that there's no general moral consensus on abortion, presumably participants must engage in some fairly sophisticated reasoning about Jane's thoughts on the matter. But it seems unlikely that participants would have done this.

Kate believes this proposition. The asymmetry in folk ascriptions in this study looks eerily similar to *FE*. The participants appear to be granting a factive mental state while denying the (traditionally entailed) non-factive mental state.

In response to Myers-Schulz & Schwitzgebel (2013), Rose & Schaffer (2013) argue that none of their four studies present genuine counterexamples to the entailment thesis when conceived in a *dispositional* sense: knowledge entails *dispositional* belief. For instance, concerning *unconfident examinee*, Rose & Schaffer (2013) show that participants were willing to ascribe *both* knowledge and belief to Kate once they were first asked to evaluate a sleeping individual's doxastic states.[92]

The apparent similarity between *FE* and the above asymmetry suggests a common cause. If Rose & Schaffer (2013) have discovered the cause of the asymmetry in Myers-Schulz & Schwitzgebel (2013), then it seems worth testing whether this cause might also explain *FE*. This account presumes a kind of folk externalism, which explains ascriptions of understanding to the amoralist. The reluctance to ascribe moral belief is then explained as a myopic focus on the amoralist's *occurrent* mental states. This can be called the *dispositional belief hypothesis*. If correct, *FE* should then disappear when participants are reminded of dispositional belief.

Method

Participants

[92] Rose and Schaffer (2013), p. S37.

307 participants completed a study on Amazon Mechanical Turk, spanning a wide age range. [93] The mean age was 31; 106 female, 198 male, 2 other, 1 not specified.

Procedures and Materials

Subjects were split into four groups. Participants in each group were assigned one of two *valences* (rightness or wrongness) and one of two *attitudes* (understands or believes).[94] To illustrate, participants assigned to the wrongness groups received the following vignette:

> Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones. When they get back to their apartment, Sandy takes a nap while Jane listens to music with her new headphones.

Those assigned to the rightness groups received a similar vignette, except the wrongness judgment was converted into a rightness judgment (e.g., "telling the clerk about the mistake is the morally right thing to do"). After reading the vignette, participants were presented with a series of statements concerning whether both the sleeping agent and the awake amoralist in the story understands/believes that the transgression is morally wrong/right. Specifically, participants were asked to indicate to what degree – on a scale

[93] 93 responses were cut from an original 400 due to redundant IP addresses (see footnote 84 for explanation).

[94] The focus on valence in this study is a residual from previous preliminary experiments in which it was hypothesized that a *Rightness Effect* analogous to the *Factivity Effect* may also be present in people's responses to amoralist scenarios (see Chapter 3). While this hypothesis did not receive confirmatory results, the presence of interaction effects made analysis difficult.

from 1 (strongly disagree) to 7 (strongly agree) – they agreed or disagreed with the following statements:[95]

> Sandy (despite being asleep) understands/believes that buying the headphones at the mistaken price was morally wrong.
>
> Sandy (despite being asleep) understands/believes that buying the headphones at the mistaken price was unethical.
>
> Sandy (despite being asleep) understands/believes that Jane shouldn't have bought the headphones at the mistaken price.
>
> Jane understands/believes that buying the headphones at the mistaken price was morally wrong.
>
> Jane understands/believes that buying the headphones at the mistaken price was unethical.
>
> Jane understands/believes that she shouldn't have bought the headphones at the mistaken price.

Results

Unlike in Rose & Schaffer (2013), the introduction of a sleeping agent did not alter the results in any significant way. *FE* remained despite this manipulation. Participants granted that sleeping Sandy both understands and believes that *X* is morally right/morally wrong (Sleeping Sandy Averaged: $F(1,307)=.048$, $p=.83$, $\eta^2 < .001$, $\alpha = .7$), but they continued to deny that Jane believes (Amoralist Jane Averaged: $F(1,307)=94.16$,

[95] Participants received three reading comprehension questions: 'Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - buying the headphones at the mistaken price is morally wrong"? Y/N' and 'Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N' and 'Is it stated in the paragraph above that Sandy argues that buying the headphones at the mistaken price is morally wrong? Y/N' All participants failed the third question, whereas only 38 failed the second and 37 failed the first. But see footnote 88.

$p<.001$, $\eta^2 = .237$, $\alpha = .7$). *Figure 4.2* and *Figure 4.3* illustrate the difference in the Items Averaged responses to Sleeping Sandy and Amoralist Jane:

Figure 4.2 – Sleeping Sandy – Items Averaged

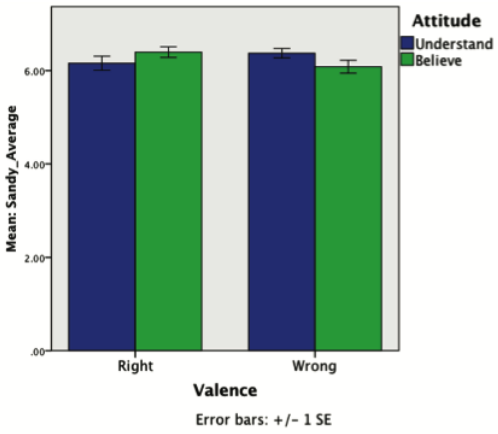
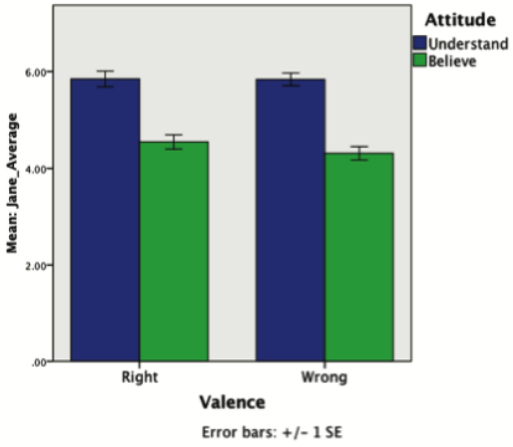


Figure 4.3 – Amoralist Jane – Items Averaged



Discussion

This is a rather surprising result for two reasons. First, it's surprising that the Rose/Schaffer-inspired sleeping manipulation didn't cause *FE* to disappear, given how similar *FE* looks to the effect in Myers-Schulz & Schwitzgebel (2013). Secondly, this result seems to suggest that even the dispositional entailment thesis fails when it comes to moral knowledge and moral belief. A denial of the entailment thesis means that the amoralist might know, but still not really believe, that *X* is morally wrong. What could this mean?

Some might take the above question as part of a *reductio* against viewing *FE* as anything but an experimental artifact. But Buckwalter and Turri (draft) argue for a different interpretation of this apparent failure of the entailment thesis: perhaps the

traditional debate over moral internalism is simply a verbal dispute founded upon a distinction within the folk psychology of belief.[96] Evaluating Buckwalter and Turri's argument will provide further motivation for testing our third deflationary account of *FE*.

Thin **and** ***Thick*** **Belief?**

Buckwalter, Rose and Turri (2013) present evidence for the notion that folk psychology admits of two different kinds of belief: *thin* belief and *thick* belief. *Thin* belief that *p* "involves representing and storing P as information [. . . .] [and] nothing more," whereas *thick* belief that *p* "involves emotion or conation."[97] While they agree that the traditional entailment thesis holds for thin belief, they argue that it does not hold for thick belief. And while they say thick belief entails thin belief, they deny that thin belief entails thick belief.[98]

A key part of Buckwalter and Turri's evidence for this thin/thick distinction comes from follow-up work on studies carried out by Murray et al. (2013). In particular, they focus on a study involving Karen the geocentrist, who, as part of her home-schooling, was taught that the earth is at the center of the universe. But when Karen is given the true/false statement 'The earth revolves around the sun' on an exam as part of her first year at a "prestigious university," she answers it correctly, ultimately making

[96] Their suggestion that the dispute can be viewed as merely verbal can be found on p.7 of their draft. I am not permitted to give direct quotes, given that this paper is in such an early stage.
[97] Buckwalter, Rose and Turri (2013), p.2.
[98] Ibid.

100% on the exam.[99] The participants in this study were asked two yes/no questions (counterbalanced):

Does Karen know that the earth revolves around the sun?

Does Karen believe that the earth revolves around the sun?[100]

Participants who granted knowledge to Karen overwhelmingly denied belief,[101] whereas a slim majority of participants who denied belief to Karen nevertheless granted knowledge.[102] These results appear to suggest that the entailment thesis doesn't always hold.

But when other researchers attempted to replicate this study, Buckwalter and Turri note that the entailment thesis held when participants were instead asked if Karen *thinks* that the earth revolves around the sun, or if "at least on some level," Karen thinks this.[103] Buckwalter and Turri offer their thin/thick distinction as the best explanation for why the entailment thesis fails in the original study yet succeeds when participants are asked the 'think' variation. On their account, while Karen can be said to both know and *thinly* believe that the earth revolves around the sun, she cannot accurately be said to *thickly* believe that the earth revolves around the sun.

With this thin/thick distinction, Buckwalter and Turri argue that the metaethical dispute over moral internalism is (likely) verbal, citing their experimental studies that directly apply their thin/thick distinction to standard amoralist scenarios. Perhaps the strongest evidence comes from their "Michael the liar" case, where Michael begins lying

---

[99] Murrary et al. (2013), p. 94.
[100] Ibid.
[101] Ibid.
[102] Ibid.
[103] Buckwalter and Turri (draft); p.13. They also cite Buckwalter, Rose & Turri (2013) as replicating these new results over a wide range of vignettes.

to his boss about how much overtime he's worked as a result of his frustration over being "passed up for promotion several times."[104] Participants in their *thick* belief group were asked "Does Michael believe that he ought to tell his employer the truth about how much overtime he works? [Yes/No]", whereas participants in the *thin* belief group were asked "At least on some level, does Michael think that he ought to tell his employer the truth about how much overtime he works? [Yes/No]."[105] The results went as they predicted.[106]

I do not believe Buckwalter and Turri's thin/thick distinction is the best explanation for the results they cite.[107] But even if their thin/thick distinction is a real

---

[104] Ibid., p.23. Recall that this paper is in an early stage; studies may appear in different forms, or not at all, in a later draft.
[105] Ibid.
[106] Ibid., p.24. "76% of participants ascribed thin belief [. . . .] [but] only 17% ascribed thick belief." It should be noted again that this paper is currently in a draft stage. This study may not appear in a later stage of the paper.
[107] Consider the Karen/geocentrist study again. While denying belief to Karen likely captures the stipulation that Karen is a geocentrist, granting the knowledge and think ascriptions aren't nearly as transparent. Participants may be struggling with disquotation; perhaps they simply want to say that Karen knows that the sentence 'The earth revolves around the sun' is regarded as true by many people (particularly her teacher) though not by her own lights. On this interpretation, participants are not ascribing thin belief to Karen. And notice that, if they were ascribing thin belief, this would imply (as Buckwalter and Turri define the notion) that Karen represents the world in a heliocentric fashion. But this is patently false of Karen. For participants to assent to the knowledge and think ascriptions, it seems sufficient that they construe Karen as grasping that the sentence 'the earth revolves around the sun' is viewed as "true" according to her instructor (after all, Karen must indicate this on the exam to get credit). This explains the asymmetry in participant responses without invoking a folk psychological distinction between thin and thick beliefs.

I think something similar can be said in response to Buckwalter and Turri's "Michal the liar" study. By assuming their thin/thick distinction, Buckwalter and Turri correctly predicted that participants would (a) grant that Michael thinks, at least on some level, that he ought to tell his employer the truth about how much overtime he works and (b) deny that Michael believes that he ought to do this. First, note that 'that it is wrong to lie' may well be just one of those sentences that people assume most regard – or, more importantly, are *aware* that most regard – as true. But awareness of that does not entail that Michael even thinly believes that he ought

feature of folk psychology, this needn't commit metaethicists to viewing the dispute over moral internalism as (likely) verbal. I will return to this second claim in section 6.2. Nevertheless, the similarity between *FE* and Buckwalter and Turri's results seems uncanny. If we imagine that participants are treating 'thinks' as a kind of *factive awareness*, then we seem to get a fairly straightforward way of accounting for *FE*: *FE* is the result of an inadvertent priming of the thick and thin senses of 'belief' – the factive priming the thin sense and the non-factive (for the most part) priming the thick sense. This account serves as a third deflationary hypothesis for *FE*.[108] And (as luck would have it)[109] the next study tests this hypothesis, which can be called the *alternative factives/non-factives hypothesis*.

3.3. Deflationary Explanation #3: The Alternative Factives/Non-factives Hypothesis

Background

Perhaps the *Factivity Effect* is merely localized to the words 'understand' and 'believe.' If different factive/non-factive words were to be employed, then, according to this explanation, *FE* would disappear. This can be called the *alternative factives/non-factives hypothesis*. As we've seen in the above section, Buckwalter and Turri's account of the

---

to tell the truth in this situation. This seems to raise the same doubts about the alleged evidence for the thin/thick distinction as in the Karen/geocentrist case.

[108] It might seem that, if Buckwalter and Turri are right about the thin/thick distinction, then the hypothesis derived from their account is actually *substantive* because their distinction is meant to mark a division within folk cognition. But, strictly speaking, since *FE* is marking a distinction between factives and non-factives, their account would actually be considered deflationary: if they're right, *FE* is an experimental artifact that arises from the difference in priming involved in using 'understand' and 'believe', having nothing to do with the use of factives or non-factives.

[109] I say this because the below study was carried out prior to any awareness of Buckwalter and Turri (draft).

folk psychology of belief seems to suggest that *FE* will disappear in the direction of externalism if we were to use words that only prime for thin belief, such as 'knows' and 'thinks.' As it turns out, the study below tests this exact replacement of factives/non-factives.[110] On this hypothesis, *FE* should disappear when 'knows'/'thinks' replaces 'understands'/'believes.'

Method

Participants

93 participants completed a study on Amazon Mechanical Turk, spanning a wide age range. [111] The mean age was 28; 32 female, 61 male.

Procedures and Materials

Subjects were split into two groups. Participants in each group were assigned one of two *attitudes* (knows or thinks). All participants received the following vignette:

> Jane and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones.

After reading the vignette, participants were asked to indicate to what degree – on a scale from 1 (strongly disagree) to 7 (strongly agree) – they agreed or disagreed with the following statement:[112]

---

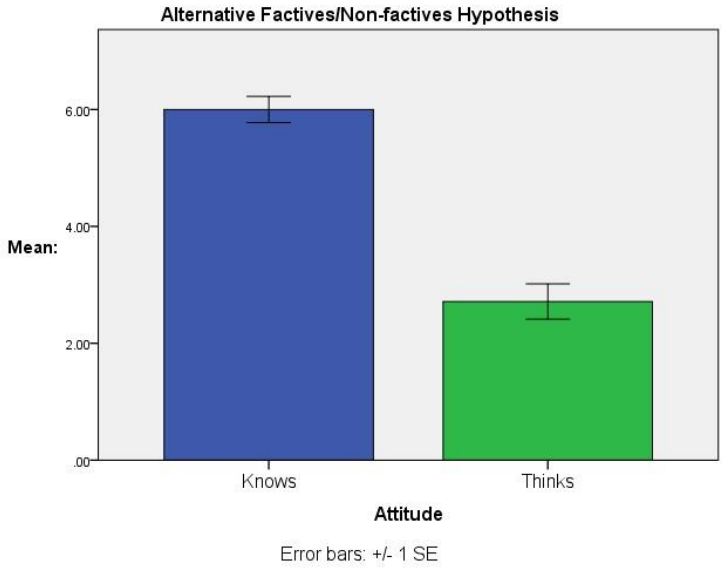
[110] This is again a rather fortuitous coincidence (see previous footnote).
[111] 7 responses were cut from an original 100 due to redundant IPs (see footnote 84 for explanation).

Jane knows/thinks that she shouldn't buy the headphones at the mistaken price.

Results

The hypothesis was not confirmed. *FE* persisted despite the replacement factive/non-factive words *(One-way ANOVA; F(1,93)=73.46, p < .001, Cohen's d = 1.8, knows: N=44, Mean=6.00, S.D.=1.49; thinks: N=49, Mean=2.71, S.D.=2.11)*. *Figure 4.4* below illustrates *FE*'s persistence:

Figure 4.4 - Results



---

[112] Participants also received two reading comprehension questions:
Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - buying the headphones at the mistaken price is morally wrong"? Y/N
Is it stated in the paragraph above that Jane knows that buying the headphones at the mistaken price is morally wrong? Y/N All but 10 participants gave the correct answer to the second question, whereas only 1 failed the first. But see footnote 88 regarding a decision made to ignore these responses given a concern about a possible ambiguity in the wording of these kinds of questions.

Discussion

Perhaps the most interesting thing about this result is that it conflicts with the prediction drawn from Buckwalter and Turri's view regarding the alleged thin/thick distinction within the folk psychology of belief. On their account, words like 'know' and 'think' both prime the thin sense of 'belief.' Since their account predicts externalist responses when the thin sense is primed, if their account is correct, we should have seen participants grant that the amoralist both knows and thinks that the behavior is morally wrong. But this is not what happened. *FE* remained, exactly as it has before, with externalist responses aligning with the factive 'knows' and internalist responses aligning with the non-factive 'thinks.' This seems to be evidence that Buckwalter and Turri's thin/thick account of belief does not carve folk cognition at its joints.

Let's take stock: three deflationary explanations of *FE* failed to be confirmed. The final study we'll consider tests a substantive explanation of *FE*. As will be seen, this is the only study of the three to receive confirmatory results.

4. A Substantive Explanation: The Moral Emotions Hypothesis

Background

Deflationary concerns notwithstanding, the *Factivity Effect* (*FE*) suggests that folk thought regarding amoralist skepticism is divided: amoralism seems more coherent if the amoralist is described as understanding/knowing that the behavior is morally wrong, but less coherent when described as believing/thinking that the behavior is morally wrong. This division could be explained if it reflected a feature of the structure of folk thought *itself*. But what might this feature be?

One suggestion is that this feature involves a folk association between moral emotions and moral belief that's absent for moral knowledge.[113] Consider: when we realize we've done something wrong, we normally feel some twinge of guilt or regret. We feel bad about what we did. But an amoralist doesn't feel these feelings of regret and guilt precisely because of her amoralism. If people associate these traditional moral feelings with moral belief, then perhaps this would explain their reluctance to ascribe moral belief to the amoralist. And if people don't associate these feelings with factive mental states like knowledge, then we would seem to have a good explanation of *FE*. On this hypothesis, the folk should be more willing to attribute classic moral feelings to amoralists when they are stipulated as believing, but not when they are stipulated as understanding, that *X* is morally wrong. This can be called the *moral emotions hypothesis*.

Method

Participants

139 participants completed a study on Amazon Mechanical Turk, spanning a wide age range.[114] The mean age was 31; 46 female, 92 male, and 1 other.

Procedures and Materials

[113] Laura King suggested this hypothesis.
[114] 61 responses were cut from an original 200 due to redundant IPs (see footnote 84 for explanation).

Subjects were split into four groups. Participants in each group were assigned one of two vignettes (*$20 bill* or *Headphones*) and one of two attitudes (understands or believes). To illustrate, participants assigned to *Headphones* received the following vignette:

> Unbeknownst to the clerk, the headphones Jane wants to buy mistakenly ring up $15 instead of $150. Jane [understands/believes] that buying the headphones at the mistaken price is morally wrong. Nevertheless, Jane proceeds to pay $15 and leaves with the headphones.

Those assigned to *$20 bill* received a similar vignette, except the headphones transgression was replaced with a case of stealing $20. Notice that the agent's cognitive attitude is explicitly *stipulated* in the vignette. This was an effort to gather ascriptions of the agent's emotional state rather than her cognitive state (unlike in the previous studies).[115] After reading their vignette, participants were presented with a series of statements concerning how the amoralist agent in the story *felt*. Specifically, participants were asked to indicate to what degree – on a scale from 1 (strongly disagree) to 7 (strongly agree) – they agreed or disagreed with the following statements:[116]

> Jane felt bad about buying the headphones.
>
> Jane felt regret about buying the headphones.
>
> Jane felt guilty about buying the headphones.

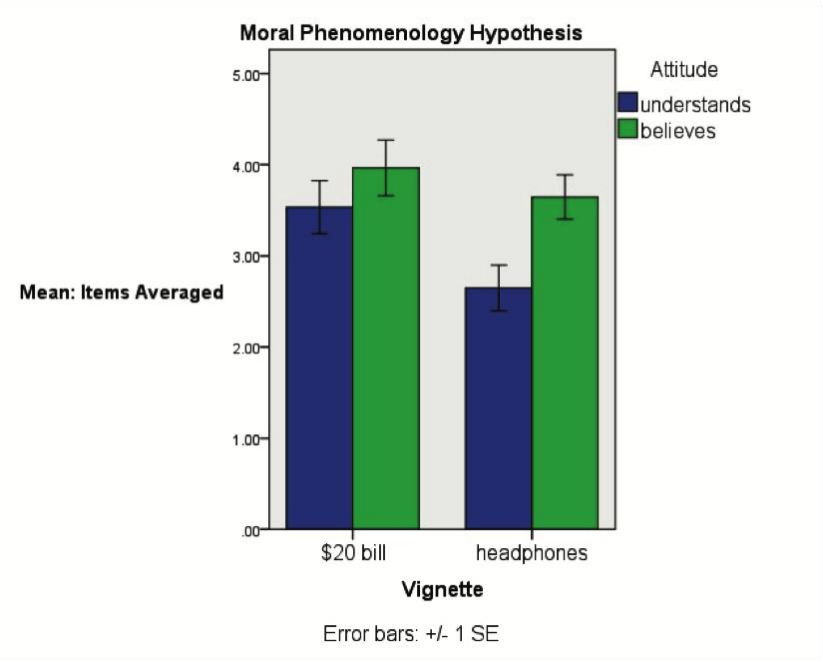
Results

[115] While the moral emotions served as dependent variables in this study, and the cognitive attitude as an independent variable, there wasn't an experimental or philosophical reason for this assignment. The hypothesis and prediction for this study would remain the same if the roles were switched.

[116] Participants also received one reading comprehension question: 'Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N' Only 2 participants failed to give the correct answer to this question.

The hypothesis was confirmed, though the effect size was relatively small (Items Averaged: *(F(1,139)=6.54, p = .01,* η^2 *= .046,* α *= .91). Figure 4.5* illustrates the results of the Items Averaged:

Figure 4.5 – Items Averaged



While an effect appears to be trending in *$20 bill*, *Headphones* shows an effect in the direction predicted by the hypothesis.

Discussion

Unlike the previous studies, this study's hypothesis was confirmed. Although the effect size was relatively small, the fact that this substantive explanation gained some plausibility where the three deflationary explanations failed is intriguing: it seems to suggest that perhaps *FE* cannot be brushed aside as merely an experimental artifact. Such

a notion seems to have serious implications for metaethical inquiry regarding moral internalism and amoralist skepticism.[117]

5. General Discussion

Three conclusions seem supported by the experiments addressed in this paper:

> (a) There is a *Factivity Effect* (*FE*): people's intuitions lean externalist when evaluating factive amoralist scenarios, whereas people's intuitions lean relatively internalist when evaluating non-factive amoralist scenarios.
>
> (b) Three deflationary explanations of *FE* were not confirmed (four when including the *blaming hypothesis* tested by Björnsson et al. (2015)).
>
> (c) There is some evidence that *FE* occurs because people associate moral emotions with non-factive states but *not* factive states.

These experimental results seem in tension with *Descriptive Equivalence* (*DE*):

> *DE*: factive descriptions of the amoralist are equivalent to non-factive descriptions of the amoralist, in the sense that they can be interchanged without loss (or gain) in the scenario's overall conceptual coherence.

Nevertheless, one might respond to these results in one of three ways:

> (1) People *mistakenly* attribute factive states to the amoralist.
>
> (2) People *mistakenly* deny non-factive states to the amoralist.
>
> (3) *P*eople *correctly* attribute factive states while also denying non-factive states to the amoralist.

In this section, I will discuss each way of responding to the experimental results. While the current data admittedly underdetermines all three of these positions, they still seem to deserve attention.

[117] To be sure, it is worth testing whether participants will, *upon reflection*, give up their initial responses for responses that more clearly respect the entailment thesis. Kenny Boyce suggested this deflationary explanation.

5.1 People Mistakenly Attribute Factive States to the Amoralist.

Internalists deny, or qualify, amoralism. So regarding *FE*, they would likely view people's ascriptions of factive states to the amoralist as mistaken. But how might the above conclusions support this view? One way is to show that internalism fits with *FE* better than externalism.

To see this, consider that Björnsson et al. (2015) ultimately view the results of their studies as suggesting a kind of *default internalism* within the folk conception of moral judgment. Essentially, the folk are internalists by default, but they will grant moral judgment without motivation *only if there's a special explanation* for this deviation from the default position. To test this explanation—what they call the *default internalism hypothesis*—they constructed two variations of their original Anna case. In one variation, the Anna scenario simply leaves out the reason why Anna lacks motivation to comply with her moral judgment. In the original scenario, it was stipulated that "she lacks empathy." This first variation simply cuts that explanation from the vignette, which they dub the NO REASON scenario. In their second variation, however, they replace the "lacks empathy" explanation with a different reason for Anna's missing motivation: she is experiencing "temporary listlessness induced by a personal crisis." They dub this the LISTLESSNESS scenario. If their *default internalism hypothesis* is correct, then participants should only be willing to ascribe understanding and belief when a special explanation is present, such as in LISTLESSNESS, but otherwise they will withhold attitude ascriptions where no explanation is found, such as in NO REASON. Thus, they predicted high rates of ascriptions for both attitudes in LISTLESSNESS, whereas they expected low rates of ascriptions for both attitudes in NO REASON.

Their predictions were confirmed: in LISTLESSNESS, 81% granted understanding and 70% granted belief, whereas for NO REASON, only 55% ascribed understanding to Anna and only 36% ascribed belief. They take these results to provide some evidence for a folk default commitment to a conditional form of internalism (they suggest something like Michael Smith's practicality requirement).[118]

Finally, it's also worth noting that the moral emotion study could reasonably be construed as identifying an internalist component within folk cognition. Perhaps some inclination to regret is necessary when one fails to live up to one's own moral beliefs, provided those beliefs are genuine. This seems much more in-line with broadly internalist pictures of our doxastic relationship to morality, at least on the (plausible) supposition that moral emotions are not motivationally inert.

5.2 People Mistakenly Deny Non-factive States to the Amoralist.

To show that people mistakenly deny non-factive states to the amoralist, one might appeal to the *entailment thesis*: knowledge entails belief. So insofar as participants are willing to ascribe certain factive states to the amoralist, they should be willing—on pain of *incoherence*—to ascribe non-factive states to the amoralist.

Now one might question whether the entailment thesis is true, or whether its scope includes any and all factive and non-factive states. But on the strongest reading of the entailment thesis, any factive state that an agent has entails that the agent has a corresponding non-factive state. And it's traditionally held to be more difficult to obtain a

[118] See the previous chapter for Smith's practicality requirement.

factive state than a non-factive, given that presumably *more* must be added to the non-factive state to make it factive.

However, there appears to be some evidence that knowledge is more easily attained – and more easily ascribed - than belief. This is a highly unexpected result if the entailment thesis is true. One would predict that ascribing belief would be quite simple in cases where the agent clearly has knowledge; the entailment thesis allows for belief to be easily read off such scenarios. But current experimental work on this prediction suggests that the matter is more complex.[119] At any rate, it's not clear that an appeal to the entailment thesis alone will be enough to show that people are mistaken to deny non-factive states to the amoralist.

5.3 People Correctly Attribute Factive States while Denying Non-factives States to the Amoralist.

One might think that people aren't making any mistakes: they're right to attribute factive states to the amoralist, and they're also right to deny non-factive states to the same amoralist. As the results from the moral emotion study suggest, people seem to take some moral emotions to be a kind of requirement for non-factive states. However, they don't seem to take these same emotions to be required for factive states. So when presented with an individual that seems to have moral knowledge while lacking this emotional component, people seem to want to describe this person as having moral knowledge without having the associated moral belief.

---

[119] "Knowledge before belief: response-times indicate evaluations of knowledge prior to belief." Conference presentation, *Society for Philosophy and Psychology*, Durham, NC (2015).

But to make sense of this account of moral judgment, we will need to make sense of how non-factive states aren't simply entailed by factive states. A start might be to show how moral knowledge and belief do not, or need not, fall within the scope of the *entailment thesi*s – the view that knowledge entails belief. But since knowledge and belief seem to fit quite naturally with truth-apt, declarative claims, it seems this task will be more challenging for some accounts of morality (e.g., cognitivism) than others.

But perhaps the results from the moral emotion study can be instructive here. Attention to this idea that non-factive states require certain moral emotions, whereas factive states lack this requirement, may reveal a sketch of how the factive ascription to an amoralist can be correct even when the non-factive ascription to that same amoralist is false. Consider again the scenario where the amoralist Chris keeps for himself the $20 bill that he just saw the person in front of him accidently drop. In this case, people are inclined to say that Chris knows that this act of theft is morally wrong, yet they want to deny that Chris really believes, or thinks, that this action is wrong. And as the results of the moral emotion study suggest, people want to say that Chris needn't feel any regret or guilt if he knew that the action was wrong, but he would need to have such states if he believed that the action was wrong.

Maybe what's happening is that some normative evaluations of Chris don't require that Chris feel a certain way about his behavior, whereas others do require that Chris have certain relevant moral emotions. Suppose we wanted to blame Chris for what he did. In that case, Chris's knowledge may be enough for him to be morally blameworthy; it's just not relevant whether Chris feels guilty or not about the behavior (at

least, on the folk conception). And, more importantly, it's irrelevant if Chris doesn't *believe* it's wrong.[120]

However, if it's true that Chris thinks that this theft is morally wrong, then perhaps the normative evaluation shifts from *moral responsibility* to *rationality*: if Chris thinks this, and does it anyway, then he will feel guilty for what he's done (on pain of rationality). On this account, participants find that the only way to really make sense of Chris, if he really believes that this theft is wrong, is to assume he must feel some of the standard moral emotions that come with such moralizing thoughts, insofar as they are genuine and sincere.

But if it's possible to hold someone morally accountable without their having any of the traditional moral emotions insofar as they have knowledge, and if it's not possible to attribute genuine moral beliefs to someone without their experiencing such states, it does seem that the entailment thesis fails in this context. To be sure, metaethicists may wish to offer a revisionist account of some of these concepts – either moral responsibility, or moral belief, or both – that would respect the entailment thesis. But if the results of the moral emotions study suggest something like the sketch above, and if metaethicists are trying to capture commonsense normative concepts and relationships, then perhaps the plausibility of the entailment thesis deserves more scrutiny within moral contexts.

---

[120] Precedence for this idea can be found within Christianity. In Romans 1:18-21, Paul seems to argue that atheists have no excuse because God has made Himself known to them: "[. . .] since what may be known about God is plain to them, because God has made it plain to them" (1:19; New International Version). Another interpretation of Paul here is that he views God as having provided sufficient evidence regarding His existence, such that the atheist's denial of God is either not genuine or the result of *epistemic* negligence.

6. Two Objections

6.1 The Expertise Defense

One might argue that *DE* is still plausibly true because the current research doesn't rule out the expertise defense.[121] The expertise defense is the idea that what matters for many of these philosophical questions is not the untutored, untrained intuitions of the masses, but rather the properly tutored, extensively trained intuitions of the expert. It's one thing to have *FE* occur among a population of non-specialists, but will this effect remain among the specialists?

While I am normally sympathetic to the expertise defense, I wonder how applicable it is to the metaethical question at issue, particularly because the specialists are *themselves* divided over moral internalism and amoralist skepticism. Thus, it's not exactly clear what role expertise would play in this context.

Of course, one might suggest that the expertise in this case is not philosopher *qua* externalist or internalist, but rather philosopher *qua* metaethicist. On this line, *FE* will presumably disappear within each *group*: externalist philosophers are expected to grant both factive and non-factive descriptions of the amoralist (i.e., accepting that the amoralist both knows and believes that *X*), whereas internalist philosophers are expected to deny both descriptions (i.e., denying that the amoralist both knows and believes that *X*).

It would be interesting to see such a study carried out. But even if the results of this study came out as predicted, it's unclear how this result saves *DE* from the above experimental results. In particular, the results from the moral emotion study may suggest

[121] See Williamson (2011) for more on the expertise defense.

a transformation in how we should conceive of the relationship between factive and non-factive states when it comes to moral evaluation and information. Insofar as the metaethicists' responses are couched in the traditional, entailment-thesis-respecting paradigm, such results may not actually tell us much about the plausibility of *DE*.

6.2 Doesn't an Ambiguity in 'Belief' Undermine the Traditional Debate?

As may be recalled from section 3.2, Buckwalter and Turri (draft) argue that the best explanation of the traditional dispute over moral internalism is that it is founded upon a "fault line" within folk psychology between what they call *thin* belief and *thick* belief. If this thin/thick distinction satisfactorily explained *FE,* the plausibility of *DE* would no longer be threatened. Factive descriptions of the amoralist would remain equivalent (in the sense of conceptual coherence) to non-factive descriptions when restricted to *either* the thin notion, or the thick notion, of belief. So on this account, *FE* is only an illusion brought about by an equivocation.

But recall that a prediction borne out by their account was falsified by the *alternative factives/non-factives hypothesis* study. If *FE* were merely the result of an ambiguity in the term 'belief,' it was hypothesized that *FE* would disappear when unambiguous terminology is employed ('know'/'think' instead of 'understand'/'believe'). But *FE* remained, despite using more thin-salient language. Thus, there thin/thick distinction doesn't best explain *FE*.

But suppose (for the sake of argument) that the thin/thick distinction actually does mark a genuine division in folk thought. Does it then follow that the traditional philosophical dispute over moral internalism and amoralist skepticism is (likely) merely

verbal? No, and here's why. At most, Buckwalter and Turri's thin/thick distinction merely reformulates the question into new terminology. The original question—do genuine moral judgments conceptually require some motivational component—has simply been *reformulated* into this question: are genuine moral judgments *thin* beliefs or *thick* beliefs? It cuts no philosophical ice to claim that they are both: the internalist never denied that the amoralist makes *any* kind of judgment. Rather, the claim is that whatever kind of judgment is being made in the amoralist context, it can't be a *moral* judgment. In the next section, I will discuss some limitations of the present research and offer suggestions for future research.

7. Limitations and Future Research

One limitation of this research (and much experimental philosophical research) is the relatively small number of vignettes used for each study. For instance, given the relatively small effect size for the results in the *moral emotions hypothesis* study, questions of whether the effect would survive more numerous stimuli seem appropriate. Increasing the statistical power of these experiments should be a focus of future research.[122]

Speaking of the vignettes, a further limitation of this research concerns the variety of moral transgressions covered. For the most part, the studies in this paper addressed only *fairness*-based transgressions such as stealing money or not paying full-price for headphones (the abortion vignette notwithstanding). Showing that *FE* isn't merely tied to some peculiar feature of fairness-based moral transgressions would speak to its

[122] Thanks to an anonymous reviewer of a publication draft of this chapter for pressing this limitation.

robustness. Thus, another focus of future research might be to employ *harm*-based and *rights*-based moral transgressions in future vignettes.

Finally, there is still some conceptual work to do regarding *how* exactly *FE* conflicts with *DE*. *DE* and *FE* are *not* mutually exclusive, so their conflict is not likely one of logical inconsistency. However, *DE* seems to warrant the prediction that the intuitive reactions of people competent with the relevant concepts will remain *consistent* across the factive/non-factive divide. For instance, assuming *DE* is true, if participants have externalist reactions in factive contexts, one may reasonably expect (*ceteris paribus*) that these externalist reactions will remain in non-factive contexts. But as the studies from myself and Björnsson et al. (2015) seem to show, this prediction failed to be confirmed. Instead, what these researchers found was *FE*. So while *FE* does not contradict *DE*, if *DE* warrants the above prediction, the presence of *FE* should count as *inductive* evidence against *DE* being true. In this way, *FE* seems *surprising* on the assumption that *DE* is true. So if *DE* is true, and *DE* does in fact warrant the above prediction, then *FE* must have some deflationary explanation. Yet as I've shown in this chapter, three initially promising deflationary accounts of *FE* failed to be confirmed. Instead, a *non*-deflationary account of *FE*—the *moral emotions hypothesis*—was the only account that received confirmation. To be sure, the *moral emotions hypothesis* doesn't necessarily conflict with *DE* any more than *FE* does. But non-deflationary explanations of *FE* are arguably less sympathetic to *DE* than deflationary accounts provided that the participants are competent with the relevant concepts.[123, 124]

[123] One might deny that the participants are competent with the relevant concepts. See Kauppinen (2007) for example, but see Nahmias and Nadelhoffer (2007) for a reply to this general concern. While I think this worry likely gets traction with some

8. Conclusion

Both internalists and externalists traditionally assume *Descriptive Equivalence* (*DE*)—factive descriptions of the amoralist are equivalent to non-factive descriptions of the amoralist, in the sense that they can be interchanged without loss (or gain) in the scenario's overall conceptual coherence. However, non-philosophers lean externalist when the amoralist scenario is described using factives yet lean internalist when the same scenario employs non-factives instead. This is the *Factivity Effect (FE)*. Because *FE* seems in tension with *DE*, defenders of *DE* may wish to explain away *FE* as an experimental artifact. But four deflationary explanations of *FE* failed to receive confirmation. A substantive explanation of *FE*—that a feature within folk cognition requires moral emotion for genuine moral belief—was the only hypothesis that received confirmation. Implications of these results for the traditional debate were considered. While some implications seem to leave the traditional debate intact (e.g., externalism may be more theory-driven than internalism), others seem to call for a radically different approach (e.g., the entailment thesis fails for moral knowledge). That the *Factivity Effect* appears to have some crucial implications, and that it doesn't appear (so far) to have a deflationary explanation, seems sufficient to be of interest to those engaged in metaethics and moral psychology.

---

experimental philosophy projects, I think the present project concerning moral judgment is an exception. I revisit this objection in chapter 5.

[124] Thanks to an anonymous reviewer of a publication draft of this chapter for pressing me on the relationship between *DE* and *FE*. The reviewer may have had a stronger worry in mind: that *no* amount of inductive evidence could affect the plausibility of *DE*. While I'm sympathetic to this kind of concern generally, I think there's reason to view some metaethical contexts as exceptions. I revisit this concern in chapter 5.

However, there is presently a general skepticism about how *any* empirical results of the kind above could contribute to *any* philosophical question or debate. While I have made some suggestive remarks about how the present research might contribute to the metaethical debate over moral internalism and amoralist skepticism, a more detailed argument seems needed to make the case that experimental philosophy can contribute to metaethics. In the next chapter, I will make the positive case that experimental philosophy can contribute to metaethics, and I will reply to a battery of objections to the effect that experimental philosophy cannot contribute to any philosophical inquiry, much less metaethics. Part of my response will rely on singling out metaethics as likely unique (relative to other philosophical enterprises) in regards to the possible relevance of experimental research on conceptual commitments and other linguistic dispositions. Another part of my response will rely on the uncontroversial assumption that both philosophers and non-philosophers alike have a folk psychological ability of ascribing moral accountability to others, which I will argue relies on an underlying competence with ascribing moral judgments to others.

Chapter 5: What Experimental Philosophy Can Contribute to Metaethics

I. Introduction

> It is therefore, the 'eye on the whole' which distinguishes the philosophical enterprise. Otherwise, there is little to distinguish the philosopher from the persistently reflective specialist; the philosopher of history from the persistently reflective historian. *To the extent that a specialist is more concerned to reflect on how his work as a specialist joins up with other intellectual pursuits, than in asking and answering questions within his specialty, he is said, properly, to be philosophically-minded* [. . . .] *a philosopher could scarcely be said to have his eye on the whole in the relevant sense, unless he has reflected on the nature of philosophical thinking.* It is this reflection on the place of philosophy itself, in the scheme of things which is the distinctive trait of the philosopher as contrasted with the reflective specialist; and in the absence of this critical reflection on the philosophical enterprise, one is at best but a potential philosopher.
>
> - Wilfrid Sellars, "Philosophy and the Scientific Image of Man"[125]

To this point, we have: (a) examined moral internalism as it arose within traditional metaethics, (b) examined amoralist skepticism as it arose as a challenge to moral internalism, (c) critically reviewed the experimental challenges raised against moral internalism and in favor of amoralist skepticism, and (d) developed an experimental argument for reworking the debate over moral internalism so that it takes into account the surprising *Factivity Effect*: people's intuitions lean externalist when evaluating factive amoralist scenarios, whereas people's intuitions lean relatively internalist when evaluating non-factive amoralist scenarios. Yet as Sellars advises in the quote above, we should now place our "eye on the whole" and begin reflecting upon the nature of philosophical thinking, with a particular focus on the place of both experimental and

[125] Sellars (1962), p.3; my emphasis.

more traditional modes of philosophical investigation. So while the *Factivity Effect* may be intriguing in its own right, an argument needs to be made for its relevance to the traditional question concerning the relationship between moral judgment and motivation. Such an argument will need to show how experimental results could impact metaethical questions. The aim of this chapter is to give this argument and defend it from possible objections.

In section 2, I will give some reasons for why questions about the nature of moral judgment do *not* simply reduce to neurological or psychological questions concerning how our minds work. In section 3, I will examine two general approaches to connecting the results from experimental studies on moral internalism to the internalism debate itself: the dialectic approach and the conceptual analysis approach. In section 4, I develop the conceptual analysis approach by way of defending a metasemantic account of our moral terminology, and I respond to an objection concerning the relevance of linguistic dispositions of non-philosophers. Finally, in section 5, I will defend experimental metaethics from general worries about the philosophical relevance of experimental philosophy.

2. Why Isn't This Just a Job For Cognitive Science?

When describing the debate between internalists and externalists as a debate about the nature of moral judgment, some might wonder why this wouldn't simply be a question to be investigated by cognitive scientists. After all, if a moral judgment is a kind of mental state, and if investigating the nature of mental states is the purview of cognitive science, then it seems to follow that the question of whether moral judgments have a connection

with motivation is straightforwardly an *empirical* question, to be explored (one assumes) by psychologists and neuroscientists. This approach has its advocates. Adina Roskies offers an argument against a strong version of internalism by way of empirical investigations into whether patients with ventromedial (VM) frontal brain damage can be accurately described as having moral beliefs despite their lacking any moral motivation.[126] The peculiar nature of their brain damage appears to leave their judgment (including, allegedly, moral judgment) intact while interfering with physiological responses that are traditionally indicative of motivation or emotion.[127] For this reason, VM patients are sometimes said to be suffering from a kind of "acquired sociopathy."[128] Roskies argues that VM patients are basically "walking counter-examples" to strong internalism. Something like the following reconstruction seems to capture Roskies' approach:

> P1: If (strong) internalism is true, then there cannot be a case of a genuine moral judgment and no corresponding motivation. [the internalist thesis]
>
> P2: If there *is* a case, then there *can be* a case. [if actual, then possible]
>
> P3: There is a case. [Roskie's cites research on VM patients]
>
> P4: Thus, there can be a case. [from P2 and P3; *modus ponens*]
>
> C: Therefore, (strong) internalism is false. [from P1 and P4; *modus tollens*]

How should traditional metaethicists respond to this argument? Some claim that the conclusion of this argument is uninteresting because it is already accepted that the strong

[126] Roskies (2003), (2006), (2008).
[127] Roskies (2003), p.55-57.
[128] Ibid., p.57.

version of internalism is implausible.[129] But can a weaker formulation of internalism still say something important about the nature of moral judgment without collapsing into an externalist account of moral judgment? Roskies doesn't think so.[130] If internalists attempt to weaken their thesis by viewing the connection between moral judgment and motivation as *ceteris paribus* instead of *necessary*,[131] then (as Roskies argues) it's difficult to see why an externalist couldn't grant this while retaining their externalist thesis. This is because *both* sides agree that motivation typically correlates with moral judgment.

I think internalists should agree with Roskies that jettisoning the necessary connection between moral judgment and motivation is to basically endorse externalism. However, retaining this necessary connection is what ultimately explains why moral internalism *cannot* simply be treated as a hypothesis to be empirically tested within cognitive science. As Michael Smith writes:

> [T]he mark of internalism, whether the internalism in question is of the strong kind [Roskies' favored formulation] or of some weaker kind, must surely be that it posits some sort of conceptually necessary connection between moral judgment and motivation. Internalism is, after all, supposed to function as an a priori constraint on what is to count as a moral judgment. The connection between moral judgment and motivation must therefore hold in virtue of the content of the moral judgment itself. It cannot be a connection that we discover empirically by uncovering a contingent psychological law.[132]

Thus, it seems internalists should object to either P1 [the strong internalism thesis], or P3 [there exists an empirical counter-example to internalism], or both, from our

---

[129] Kennett and Fine (2008a), p.178-181; Smith (2008), p.208.
[130] Roskies (2008), p.192-93.
[131] As Kennett and Fine (2008a) appear to do, though see Kennett and Fine (2008b), p. 217.
[132] Smith (2008), p. 210.

reconstruction of Roskies' argument for treating moral internalism as a thesis about how our minds work.

For her part, Roskies thinks internalists must give a "substantive" and "independently motivated" account of the conditions where motivation fails if weaker versions of internalism are to avoid vacuity and *ad hockery*.[133] For example, she argues that Smith's practicality requirement is "too weak to be revealing about the nature of moral judgment" given the lack of any substantive account of "what it is to be practically rational."[134] Once properly worked out as suggested – so the argument goes – this weaker version of internalism will also involve empirical implications *somewhere* down the line.

It's informative to see why some philosophers deny this possible implication (or at least qualify it), so I will consider two responses: one from Smith defending the meaningfulness of his practicality requirement and another from Antti Kauppinen defending the *methodological autonomy* of metaethical inquiry relative to empirical moral psychology.

Smith guesses that what Roskies has in mind by requiring his account of practical rationality to be "substantive" and "independently motivated" is something like providing either an operational definition or a stipulation. But Smith argues that neither of these ways of defining 'practical rationality' will be promising. By analogy, he points to how such a requirement would fail to capture a presumably meaningful and important sense of *theoretical* rationality, illustrated in the following claim:

---

[133] Ibid. We will return to this *ad hockery* charge later in the chapter.
[134] Roskies (2003), p.53-4.

> If someone believes that Socrates is a man and she believes that all men are mortal, then either she believes that Socrates is mortal or she is theoretically irrational.[135]

Let's follow Smith in calling this claim MORTAL. Notice its similarity to Smith's practicality requirement. Yet Smith argues, "[u]nder any plausible operational definition [of theoretical rationality], MORTAL looks bound to turn out false."[136] It seems one would always be able to discover a counter-example to MORTAL once understood operationally. But Smith also argues that a defender of MORTAL needn't simply stipulate what is or isn't going to count as being theoretically rational. Rather, it seems one can point to the contents of the beliefs in question and then explain that the agent isn't being appropriately sensitive to those contents insofar as she genuinely lacks the belief that Socrates is mortal yet genuinely believes that Socrates is a man and all men are mortal. Smith goes on to give his account of practical rationality along similar lines. The takeaway seems to be that there is a perfectly meaningful way to characterize the "exception conditions" of some weaker internalist thesis while (1) retaining a conceptually necessary connection between judgment and motivation and (2) avoiding any straightforward empirically testable implications.

The second reply I wish to consider comes from Antti Kauppinen. Kauppinen argues that neurological and physiological observations are insufficient to determine what psychological state a person is in at any given time. This is because (roughly) *commonsense evidence* is needed to determine whether certain psychological states obtain in any given individual at any given time. By 'commonsense evidence,' Kauppinen appears to mean the pre-theoretical, folk-psychological intuitions that are

---

[135] Smith (2008), p. 212
[136] Ibid.

used by both philosophers and non-philosophers alike in making sense of our behavior (we will revisit this notion of 'commonsense evidence' when developing both the conceptual analysis and dialectic approaches to connecting up experimental philosophy with metaethics). The upshot for the metaethical debate over moral internalism is this: no amount of neurological or physiological observations could even *in principle* falsify the moral internalist thesis since (a) these observations must first be interpreted via our commonsense evidence about what psychological state obtains, and (b) such evidence involves the very same pre-theoretical intuitions that metaethicists make appeal to in order to buttress (or discredit) moral internalism itself. In a phrase: one cannot appeal to neurological or physiological evidence to support either internalism or externalism without essentially *begging the question*.[137]

I think something like Kauppinen's account here is the right thing to say about why Roskies' cognitive science approach is unlikely to succeed.[138] Yet—as one should expect given the title of this chapter—I *don't* think this account successfully blocks *any* kind of empirical research from gaining traction with metaethical questions. To understand how experimental philosophy might find an opening where cognitive science couldn't, consider how Kauppinen qualifies his concession that, of course, if empirical work did provide information of how we actually form moral judgments, then such research "may exert pressure on some metaethical views":

> Even then, though, it seems to be a matter of details: one must already have decided that pedigree matters, for example, before data about what

[137] See Kauppinen (2008) for his defense of these claims. For my purposes, I simply want to get his account on the table.

[138] Others have raised similar concerns, such as citing the plausibility of a Davidsonian mental holism when it comes to interpreting the behavior and mental states of an agent. See Cholbi (2006).

> people actually do give rise to pressure to modify one's view on what kind of processes count as leading to genuine moral judgment. And this already presumes that one believes people actually make moral judgments. If, for example, it tums out that practically nobody reasons in accordance with the categorical imperative, a Kantian either has to lower the bar to count less than impartial judgments as moral or, like Kant himself, conclude pessimistically that morality is a scarce commodity indeed in this world.[139]

Analogous to Kant's pessimism, it may seem open to the internalist to take any alleged "walking counter-example" to internalism and turn it on its head—viewing it instead as evidence that fewer people than we may have thought *really* make moral judgments.

To be sure, this move might initially seem *ad hoc*, but I think it's important to see that one can give examples where this move seems exactly right. Suppose some cognitive scientists are interested in whether the mathematical judgment *addition* fits with what mathematicians call the *plus* function, or whether it's actually more like Kripke's *quus* function (i.e., just like *plus*, except when you get to a certain set of calculations the sum always comes out '5').[140] So they begin putting certain patients with peculiar kinds of brain damage through PET scans, and—lo and behold—they find that these patients appear to be using the *quus* function after all! Now it seems perfectly open to mathematicians to argue that these results have told us nothing about the nature of addition judgments. At most, what it shows is that these patients are not *in fact* doing *genuine* addition when we give them addition problems to solve. And notice that this would seem to be the best response even if the sample was particularly massive. In fact, it may still be the right response even if the cognitive scientist put *every person on the planet* through those PET scans and found that *we are all* using the *quus* function! So as

[139] Kaupinnen (2008), p.23.
[140] Kripke (1982).

it seems with mathematical judgment, it seems metaethicists may wish to reserve genuine moral judgment against all empirical contingencies.

But I think it's important to consider reasons why resistance in the mathematics case seems appropriate. Perhaps it's because we have other evidence available to us about the content of the *plus* judgment (e.g., its apparent employment in scientific calculations, computers, etc.). If this is right, and if there's nothing analogous in the case of moral judgment, then this could raise suspicions about the plausibility of this move in the metaethical context. And as Kennett and Fine explain, there does seem to be a price to this move as greater and greater portions of population are discounted:

> [W]e think that philosophical accounts of moral judgment would be called into question if it turned out that they were highly esoteric. Rationalists like Smith take themselves to be providing an analysis of our folk concepts. The program of conceptual analysis is to refine and systematize the elements of our folk concept. While it thus to some degree idealizes and abstracts away from everyday practice, it simply can't be the case that those engaged in the practice do not for the most part have a mastery of the concepts. [Thus] if it turned out that the folk do not, implicitly or explicitly, take themselves to be making any claims about anyone's reasons for action in making their everyday moral judgments, we would have reason to reject rationalist versions of internalism.[141]

Phillips and Worsnip (unpublished manuscript) also suggest something like this concern when defending the relevance of experimental philosophy to investigating moral internalism:

> […] motivational internalism of the sort that interests us is a conceptual claim [. . .] And we think that conceptual claims should be faithful to how ordinary people understand the concept in question: in this case, this is tested by examining how people ordinarily attribute moral judgments. This fidelity can be thought of as a kind of adequacy constraint on accounts of the conceptual features of moral judgment.[142]

---

[141] Kennett and Fine (2008b), p. 219.
[142] Phillips and Worsnip (unpublished manuscript).

Thus, even if there's no straightforward way for cognitive science research to impact debate over moral internalism and amoralist skepticism, if something like the "adequacy constraint" above is correct, there seems to be a way to get the experimental foot in the traditional door: empirical investigation into the 'commonsense evidence' itself.

3. How Experimental Results Can Impact the Traditional Debate: Two General Approaches

Recent critical discussions concerning how experimental philosophy connects up with traditional philosophy reveal that there are a variety of different kinds of projects that go under the heading 'experimental philosophy.'[143] The upshot (for our purposes) is that an investigation into how experimental philosophy could get traction with traditional philosophical questions will likely come down to the nature of the particular project in question. Moreover, even within a single experimental project, there seem to be a variety of different approaches to connecting up the experimental results with philosophical questions. Regarding experimental metaethics, two general approaches have been suggested in the published literature: one I'm calling the *conceptual analysis approach*, and another I'm calling the *dialectic approach*. We will begin with the *dialectic approach* first.

3.1. The Dialectic Approach

One way to connect up experimental results with metaethical questions is by what I'm calling the *dialectic approach*. On this approach, while experimental results may not

[143] Nadelhoffer and Nahmias (2007).

directly influence the truth of either moral internalism or amoralist skepticism, it can influence the arguments and general moves within the dialectic over these philosophical issues. Questions about, for example—where the burden of proof lies, or how to conceive of the desideratum, or whether there's reason to suspect that one side's intuitions are more theory-driven—have all arisen in the traditional literature on moral internalism. Answering these questions would seem to have an impact on the traditional discussion, thus influencing the philosophical plausibility of the positions in play. So if there's good reason to think that experimental philosophy can help answer these questions, then this would show another way experimental philosophy can contribute to traditional philosophy. Below I discuss two ways this dialectic approach has been developed in the literature: (1) experimental philosophy might help reveal relevant cognitive biases afflicting one side of the philosophical debate, and (2) experimental philosophy might help defend against the charge of *ad hocness.*

Revealing relevant cognitive biases

Björnsson et al. (2015) point out that cognitive biases do not only afflict non-philosophers.[144] So supposing it could be shown that either externalist philosophers or internalist philosophers are myopically focusing on features of thought experiments that tend to confirm their externalist or internalist perspective, this might be evidence that one side is more prone to *confirmation bias* than the other. If it could be shown that, say, externalists seem more prone to confirmation bias than their internalist interlocutors, then the dialectic charge of theoretical bias might weigh more heavily on the backs of

[144] Björnsson et al. (2015), p. 717.

externalist philosophers than internalists. While this isn't itself evidence against externalism, it may undermine certain kinds of evidence used to support externalism, unless externalists can explain away the evidence supporting the theoretical bias charge. Björnsson et al. (2015) claim that the kind of experimental results we've been addressing in the past couple of chapters could serve as theoretically-independent evidence regarding whether one of the sides—externalists or internalists—seem more subject to theoretical bias.[145] They take their own experimental results to provide just such evidence against externalists.[146]

Defending against *ad hockery* charges

Within the traditional literature, externalists have sometimes viewed internalists' adjustments to their internalist thesis as merely *ad hoc*. In a desperate move to save internalism from counter-examples (according to this charge), the internalist philosopher merely adds new conditions that the alleged amoralist must meet (e.g., the individual must be rational (Smith); the individual must be psychologically healthy (Smith and others), the individual must have been motivated by their moral judgment at some point in their life (Dreier), the individual must be part of a community of speakers where most members are motivated by their judgments (Dreier; Blackburn).[147] What can the internalist do to defend themselves against this charge of *ad hockery*? Phillips and

---

[145] Ibid.

[146] Ibid., "[. . .] our studies suggest that the worry about theoretical bias is a more pressing one for externalists, as most subjects seemed to have broadly internalist intuitions, and as there is reason to mistrust or reinterpret the seemingly externalists tendencies that yielded the same attributions of moral belief in the Psychopath and Inverted Commas scenarios" (p. 731).

[147] Phillips and Worsnip. (unpublished manuscript).

Worsnip (unpublished manuscript) argue that survey results could help internalists defend themselves against this worry.[148] If many people *unfamiliar* with the metaethical discussion also deny that the unmotivated individual really makes a moral judgment in amoralist scenarios sensitive to these conditions, then internalists could cite this result as some evidence that they aren't merely fashioning these conditions in order to save internalism. So, if pre-theoretical responses to amoralist scenarios constitute a kind of theory-*neutral* starting place for evaluating moral internalism, and if experimental philosophy is capable of systematically gathering such pre-theoretical responses,[149] then internalists might cite experimental results as evidence that their conditional varieties of internalism need not be *ad hoc* attempts at dodging externalist counter-examples (that is, assuming the results come out sympathetic to internalism). This is then another way in which experimental philosophy can contribute to the traditional philosophical discussion over moral internalism and amoralist skepticism.

3.2 The Conceptual Analysis Approach

Every experimental philosopher working in this area has noted that moral internalism has traditionally been put forth as a *conceptual* thesis about what is to *count* as a moral judgment. As we saw in the previous section, this explains why the debate between internalists and externalists cannot likely be straightforwardly settled by merely deferring to cognitive science or neuroscience. Yet if moral internalism is alleged to be a *conceptual* truth, then it seems we might still derive an *empirical* prediction from this

[148] Ibid.

[149] Kauppinen (2007) argues that they cannot do this. I take up Kauppinen's argument within the next section.

thesis after all: those competent with the concept of a moral judgment will be disposed to act and make judgments in line with moral internalism. So, if carefully-crafted studies could—at least *in principle*—test for evidence confirming or disconfirming this empirical prediction, then this serves as one way for experimental philosophy to contribute to traditional metaethics.

To be sure, how one views concepts, conceptual competence, conceptual content, and so forth, may determine how plausible one finds this approach. But instead of evaluating every account of conceptual analysis in the literature and determining how amenable each is to this approach,[150] let's discuss one account in particular which has already been claimed to be appropriately amenable. This account is sometimes called *Lewisian-style* conceptual analysis or *conceptual role semantics*.[151] Briefly, conceptual role semantics treats the meanings of our concepts as determined by the role they play within our thinking.[152] This account places focus on a kind of pre-theoretical understanding that competent speakers have with their concepts. Michael Smith provides a nice caption describing something like a Lewisian-style conceptual analysis for moral concepts:

> To say that we can analyze moral concepts, like the concept of being right, is to say that we can specify which property the property of being right is [. . .] by reference to descriptions of the inferential and judgmental dispositions of those who have mastery of the term 'rightness' [. . . .][153]

If we apply Smith's characterization of the analysis of the concept of being right to the concept of being a moral judgment, we get the following:

---

[150] Machery (2008), p.170-75 provides a sketch of how the three predominant theories of concepts might account for experimental results of this kind.
[151] The classic expression of this account is found in Lewis (1970).
[152] Greenberg and Harman (2005).
[153] Smith (1994), p.39.

> To say that we can analyze the concept of being a moral judgment, is to say that we can specify which property the property of being a moral judgment is by reference to descriptions of the inferential and judgmental dispositions of those who have mastery of the term 'moral judgment.'

This "reference to descriptions of the inferential and judgmental dispositions" seems interestingly similar to Kauppinen's comments from the previous section about *commonsense evidence*: the pre-theoretical, folk-psychological intuitions we use to explain behavior. Insofar as participant responses in experimental studies are derived from said inferential and judgment dispositions, then it seems an argument can be made for taking these responses as evidence for how we should fill out the descriptions of those dispositions. And since the descriptions are what philosophers cite (at least on this account) to determine which property is to be the property of, in our example, being a moral judgment, then we seem to get a relatively straightforward way of connecting up experimental results to the plausibility of moral internalism as a thesis concerning that very property.

Shaun Nichols—who first developed what I'm calling the conceptual analysis approach to connecting up experimental philosophy to the traditional debate over moral internalism—cites Michael Smith's affinities for this Lewisian-style conceptual analysis as one of his main reasons for making Smith's moral internalism the subject of his pilot study.[154] According to Nichols, Smith's approach commits him to viewing his moral internalist thesis as partly determined by a "systematized set of platitudes that characterize the folk concept of morality."[155] Nichols writes,

> Although the project of systematizing the platitudes will presumably require serious analytic resources, the project also has substantive

[154] Nichols (2004b), p.74-5.
[155] Ibid., p.75

> *empirical* checks since the platitudes themselves are supposed to be claims that most people would accept.[156]

Let's consider the below argument to be a formal reconstruction of the kind of connection Nichols takes there to be between experimental philosophy and the plausibility of moral internalism as construed in light of the above account of conceptual analysis:

> P1: If inferential and judgmental dispositions serve an essential role in analyzing the concept of a moral judgment, then experiments designed to systematically investigate these inferential and judgmental dispositions can provide evidence as to whether moral internalism captures an essential feature of this concept. (Nichols' inference)
>
> P2: Inferential and judgmental dispositions serve an essential role in analyzing the concept of a moral judgment. (Lewisian conceptual-role semantics)
>
> C: Therefore, experiments designed to systematically investigate these inferential and judgmental dispositions can provide evidence as to whether moral internalism captures an essential feature of this concept.

Since this argument is valid, it seems one can dismiss it only by denying one of its premises. For the moment, we are stipulating that P2 is true. Are there reasons to doubt P1? Some philosophers have thought so.[157] Let's look at their concerns.

Kauppinen (2007) grants that there are philosophical debates that can (and, more strongly, must) be addressed by appeal to pre-theoretical intuitions. In fact, as we saw in the previous section, he uses the dispute over moral internalism as particularly paradigmatic![158] Yet despite his concession that some philosophical positions predict that the ordinary folk will have this or that intuition, Kauppinen argues that experimental philosophers cannot capture the folk intuition in these cases, at least not by using the

---

[156] Ibid.
[157] Kauppenian (2007); Joyce (2008).
[158] Kauppinen (2007), p.95-6.

standard "survey model."[159] The insurmountable problem for experimental philosophy, as Kauppinen sees it, is that the surveys can only record (what he calls) "surface" intuitions, whereas the intuitions relevant to philosophical debate are (what he calls) "robust."[160] Robust intuitions are derived from individuals who are (1) competent with the concept in question, (2) especially careful in their assessment of the given thought experiment, and (3) appropriately reflective such that they do not respond to "pragmatic"[161] aspects of the presented case.[162] If the subject lacks any one of these three characteristics, then Kauppinen holds that their responses are surely irrelevant to philosophical inquiry and debate.

If Kauppinen's criticism is right, then even if inferential and judgmental dispositions serve the essential role in conceptual analysis that we've been discussing, experiments designed to systematically investigate these inferential and judgmental dispositions would *not* provide the appropriate evidence. So Kauppinen offers a challenge to P1 of our reconstruction above.

In response, I want to begin by conceding a good deal to Kauppinen's criticism, generally speaking. First, I agree that what matters for at least some philosophical projects and debate are *robust* intuitions just as he defines them. Second, I agree that the

---

159 Ibid., p.97.
160 Ibid., p.97.
161 Nahmias and Nadelhoffer (2007) offer an example of what Kauppinen has in mind here, ultimately showing that experimental philosophy has the resources to address this concern. Results from a study by Joshua Knobe seemed to suggest that people do not think intentionally X-ing requires that one intended to X. However, some critics suggested that the results were due to the subjects being misled by a particular pragmatic consideration: perhaps the subjects viewed *moral responsibility* for X-ing as implying intentionally X-ing. However, when another study was setup to account for this pragmatic consideration, the results remained the same as in the Knobe study.
162 Kauppinen (2007), p.101.

average non-philosopher (and even to some extent the average philosopher) is bound to lack competence with a large number of concepts. So, for instance, I accept that there are a number of philosophical disputes (e.g., within philosophy of science, and maybe even epistemology) that need not consider folk conceptual commitments for this very reason: it is likely that the folk have a very naïve (or at least confused) grasp of scientific concepts and (some) epistemological concepts.[163] So, insofar as the individual lacks the relevant competence with the concept in question, and insofar as the survey is not capturing robust intuitions, I think that Kauppinen is right.

But I am concerned with disputes in metaethics, and thus the concepts I am interested in are *moral* concepts.[164] Do we have any good reason to think that the majority of folk *lack* competence with *moral* concepts? If we did, then it seems we would have arguably good reason to believe that the folk are not really moral agents (i.e., they never really behave morally, because they lack sufficient understanding of, for example, what distinguishes moral behavior from any other kind of behavior, just like very small children or non-human animals).[165] But we clearly have good reason (at least *prima facie*)

---

[163] The response I have in mind here is captured by Nadelhoffer and Nahmias (2007): "On this moderate view, the philosophical relevance of folk intuitions will vary from topic to topic. However, just because some intuitions may not be relevant to philosophy, it does not follow that we should so hastily banish *all* folk intuitions" (p.15 – emphasis in original)

[164] And metaethical concepts, but only indirectly. It's surely the case that most non-philosophers are not competent with the *philosopher's* metaethical concepts. However, it may still be possible to infer the metaethical inclinations of the folk by way of their employment of moral concepts.

[165] A possible objection might be the following argument from analogy: If the folk lacked *epistemic* concepts, it wouldn't follow that they weren't *epistemic* agents (i.e., knowers). Thus, why should it follow that the folk aren't moral agents if they lacked moral concepts? There seem to be two possible responses to this objection. First, it might actually follow that the folk wouldn't be epistemic agents either, precisely because they might be incapable of having justified beliefs, insofar as justification

to think that the vast majority of folks are indeed moral agents. Thus, these folks must be competent with *moral* concepts, at least to a sufficient degree for their agency to be described as moral.

One might grant that perhaps the folk are competent with moral concepts, but then argue that this might simply show the relevance of experimental philosophy to normative, first-order theorizing (so experimental *ethics*, but not necessarily experimental *meta*ethics). To get relevance to traditional metaethics – the objection continues – would require that the folk be competent with *metaethical* concepts. So concerning the dispute over moral internalism, we would need some reason to think that the folk are competent with the concept of a moral judgment, a presumably metaethical concept. Lacking such a reason, it seems Kauppinen's concern about the folk lacking robust intuitions remains.

First, it should be noted that what's being asked of participants in experimental *ethics* is presumably *no*t a conceptual question, but a *normative* one: the intuitions being probed our intuitions about what is right or wrong *to do* in the circumstances.[166] Thus, competence with moral concepts wouldn't necessarily be sufficient for one's intuitions to

---

involves the possession of certain epistemic concepts, and epistemic agency involves the capacity to have justified beliefs. But I admittedly don't find this response all that plausible. A better response may be that there is good reason to believe that moral agency is *special* in this sense – moral responsibility, moral culpability, and moral criticism in general seems to presuppose that agents possess such concepts (e.g., "he knew that was morally wrong but he did it anyway"), in a way that epistemic responsibility, epistemic culpability, and epistemic criticism in general may not presuppose an analogous possession of epistemic concepts. We don't think it's necessary that you know the difference between a justified belief and an unjustified belief in order to be said to have knowledge. Yet we arguably do think it's necessary that you know the difference between a right action and a wrong action in order to be said to be a moral agent. And presumably knowledge of the difference between right and wrong requires competence with moral concepts. Thanks to Philip Robbins for raising this objection.

[166] Phillips and Worsnip. (unpublished manuscript).

be evidence that one's moral view is *correct*, anymore than competence with scientific concepts would be sufficient for one's intuitions to be evidence that one's scientific view is *correct*.[167] Regarding the project at hand, what matters is not one's metaethical intuitions *per se* (e.g., whether realism or anti-realism is true), but rather one's conceptual/semantic intuitions regarding the concept of a moral judgment.

Having said this, is there reason to think that the folk are competent with the concept of a moral judgment? I think so. First, consider that we intuitively make judgments concerning moral responsibility or moral blameworthiness.[168] Part of what's involved in making judgments concerning blameworthiness are further judgments concerning whether or not the individual we are judging "knew the difference between right and wrong" or "knew that what he was doing was morally wrong." These latter judgments plausibly involve our folk psychological capacity to ascribe mental states to other people, something that we do reliably and without much effort. And in this case, the relevant mental states we are ascribing must essentially involve *moral judgment*: we are employing our mind-reading ability to determine whether an individual *knew* that what they were doing was morally wrong.

Consider the following example. I have twin toddlers, one boy and one girl. They occasionally get upset with each other, and resort to hitting or throwing toys. Now it seems safe to say that it's morally wrong to physically injure someone simply because they are playing with a toy that you want. However, it's unlikely that any of us would say

---

[167] And this seems right even if we grant (as I do) that moral truths are likely (partly) *a priori* truths.

[168] These notions may come apart. If so, we should consider the one that has the closest ties to our judgments about others' mental states (likely moral blameworthiness).

that my twins are morally blameworthy in such circumstances. This is (at least in part) because we doubt that my twins understand the moral salience of their actions (in this case, that they are being cruel to one another). But how could we have come to this conclusion without some competence with the concept of a moral judgment? How could we feel confident in our folk-psychological assessment if not for what appears to be our conception of a mental state described as a moral belief, moral knowledge, or some doxastic grasp of moral considerations? The fact that we have this capacity to ascribe the relevant mental states to individuals as part of our assessment of their moral blameworthiness seems to provide a reason to think that the folk are competent with some metaethical concepts, particularly the concept of a moral judgment. The folk must be employing such a concept in order to effortless assign moral blameworthiness to other individuals.

Supposing this addresses Kauppinen's first criteria for the presence of robust intuitions, there remain two more: (2) that subjects should be appropriately reflective, and (3) that subjects should attend only to semantic considerations. His second criteria can be construed in at least two ways: (a) that subjects should be as reflective as philosophers, or (b) that subjects should be as reflective as a reasonable, English-speaking person whose competent with the relevant concepts. While there may be projects where (a) is arguably necessary, there's good reason to deny its relevance for the current project concerning moral internalism. Experimental philosophical research on moral internalism was born out of the intuitive stalemate between internalist and externalist philosophers. Requiring subjects to reflect in this way would seem to defeat part of the purpose of presenting

amoralist scenarios to non-philosophers. Thus, (b) is likely all that's required for the current project.

With the above in mind, it seems that experimental philosophers can account (and are currently accounting) for Kauppinen's last two criteria by devising better studies. Experimenters are designing studies that place more focus on comprehension and reducing pragmatic effects. To be sure, even these changes will not *guarantee* that each subject is appropriately reflective and attentive. But as Nadelhoffer and Nahmias (2007) remind us, the statistical methods used by most experimental philosophers have a built-in way to detect the effects of subject mischief and inattentiveness: "Using sufficiently large sample sizes, you can show that the probability is extremely low that the relevant results obtained because of the irrelevant factors."[169] Moreover, study replication may also help reduce the likelihood that the results are due to some of the kinds of confounds that Kauppinen raises.

Considering the studies I've presented in this dissertation, the reflection and attention demands on participants were presumably quite low. For instance, it may be recalled that, in one study, participants read a short vignette that quoted an individual Jane saying she agrees with her friend Sally that getting an abortion merely to avoid parenthood is morally wrong, but that she just doesn't care if she does what's wrong. Participants were then asked (among other questions) whether Jane understands/believes that she shouldn't get an abortion merely to avoid parenthood. Given what appears to be a relatively simple vignette and test question, it's not obvious to me where reflection or

[169] Nadelhoffer and Nahmias (2007), p.26.

attention issues could arise.[170] But it should also be noted that, out of the four studies I ran in an attempt to explain the *Factitivty Effect*, three of those four were constructed on the assumption that there must be *some* failure of reflection or attention. Yet none of these studies' deflationary hypotheses received confirmation. The takeaway, I think, is that even if there is good reason to suspect reflection or attention failures, such suspicions ought to be translatable into testable hypotheses concerning the studies in question. Thus, experimental philosophers can presumably address Kauppinen's last two criteria by running more experiments in an effort to catch such failures or, as in my case, to provide more empirical reasoning for doubting that the effect in question is due to such failures.

But even supposing Kauppinen's three criteria are met, there may be an additional worry when it comes to experimental metaethics. Richard Joyce charges that some metaethical disputes—in particular, the dispute over moral internalism—may not be affected by folk intuitions *even if* the folk share competence with philosophers regarding certain concepts.[171] This is because this shared competence, according to Joyce, is exemplified only in the kind of knowledge related to one's use or abilities with the concept, sometimes referred to as knowledge-*how*.[172] But presumably, Joyce claims,

---

[170] One might cite the systematic failure of one of our comprehension questions as evidence of either a reflection or attention failure. But it should be noted that this failure was likely due to issues with the question itself, given that the vast majority of participants gave correct answers for the other comprehension questions. By analogy, if almost every student in a class misses a question on an exam, whereas the other exam questions have a more typical breakdown, it seems reasonable to assume that question construction may be at fault.

[171] Joyce (2008), p.380-83. It should be noted that, in this section, Joyce is offering, for argument's sake, a possible defense for Michael Smith against Shaun Nichols' experimentally-infused argument against Smith's form of moral internalism; Joyce clarifies that he does not necessarily endorse Smith's project.

[172] Lenman (2008) discusses the distinction between knowledge-*how* and knowledge-*that* in this context.

knowledge relevant to disputes over moral internalism will require something like knowledge concerning the proper application of the concept. But this kind of knowledge appears to be a kind of knowledge-*that*, given that it is knowledge *about* the concept itself.[173]

Is it reasonable to think the folk have knowledge-*that* concerning their own moral concepts? Joyce says no, concluding that experimental probing of folk intuitions will be of little help even if the folk are competent (knowledge-*how*) with such concepts. Joyce draws an analogy to asking a champion swimmer to describe his swimming technique.[174] Though the swimmer is certainly competent with the technique, he may nevertheless fail to "articulate" the nature of the technique itself (this is in line with the well-known phenomenon of great sports stars making peculiarly poor teachers of their own practice). Similarly, while the folk may be able to apply their concepts with ease, they may not be able to accurately assess when certain concepts of theirs are being applied appropriately. In this way, even conceptual truths can thus be "exceedingly unobvious to ordinary speakers."[175] Joyce concludes that, insofar as one is seeking information about whether speakers are internalists or externalists, then one would do better to find "an expert who has examined the patterns of use of moral language as it is employed in real life."[176]

Some experimental philosophers have responded by pointing out that appeals to expertise seem useless when there is considerable dispute among the experts

---

[173] Ibid.
[174] Joyce (2008), p.382.
[175] Ibid.
[176] Ibid. This seems similar to, if not the same as, a point Kauppinen (2007) presses against experimental philosophy, p.109-110.

themselves.[177] In these cases, so the reply goes, probing folk intuitions may provide some much needed evidence to help break the impasse.

I do not find this reply very appealing. While the existence of this dispute among the experts can explain why we might be interested in an experimental philosophy approach, it doesn't by itself *justify* this approach. If the folk lack knowledge-*that* concerning their own moral concepts, and if Joyce is right that only this kind of knowledge is relevant to the dispute at hand, then folk intuitions cannot provide any important evidence *regardless* of whether there is considerable dispute among the experts or not.

But I think there is a better response to Joyce's charge. Joyce claims that the folk must have knowledge-*that* regarding their own concepts if experimental results are to be relevant to disputes over conceptual issues, like that over moral internalism. But this isn't necessary. To see this, consider Joyce's champion swimmer analogy again. While it's likely true that the swimmer may not be able to accurately explain their own technique, this swimmer could probably still recognize if another swimmer was using their technique incorrectly. In other words, while the champion swimmer may not be able to *explain* what *exactly* has gone wrong, the swimmer could still determine that *something* has gone wrong, whatever it is. Similarly, the folk's ability to recognize that *something* is amiss—whatever it is—in the scenarios provided by experimenters may be sufficient for folk responses to be relevant to disputes over certain conceptual issues, particularly in metaethics.

[177] Nichols (2008), p.399; Nahmias et al. (2006), p.34.

4. Defending the Conceptual Analysis Approach

In my view, the conceptual analysis approach outlined in the previous section requires a metasemantic account of what fixes the content of our moral terminology. In particular, such an account will need to show how it can be the case that many non-philosophers are sufficiently competent with moral terminology in a way that is relevant to the metaethical question at hand. In this section, I will sketch a metasemantic account of our moral terminology that, if true, I think makes the conceptual analysis approach to certain metaethical questions particularly plausible.

4.1. Moore's Open Question *Revisited*

As may be recalled from chapter 2, moral internalism arose as an explanation for a rather curious feature of our normative concepts (particularly moral and value concepts) brought out by Moore's *Open Question* argument: of any proposed analysis of goodness (or rightness), we seem to be able to intelligently and coherently ask whether the *analysans* is *in fact* good (or right). As may also be recalled, since Moore first raised this challenge at the beginning of the 20th century, considerable work has been done in the philosophy of language and mind to raise doubts about the coherence or relevance of the challenge itself, not the least of which has been the possibility of *a posteriori* identities.[178] But despite the variety of concerns philosophers have raised against Moore's *Open Question* argument, different versions of the challenge continue to be raised.[179] Stephen Darwall thinks the "staying power" of the argument can be explained by the appeal of moral internalism itself:

[178] A nice review of this work can be found in LaPorte (2016).
[179] Horgan and Timmons (1992); Gampel (1996); Darwall (1992).

> Recall that intrinsic goodness for Moore (in *Principia*) has unqualified normativity built into it. Moore alternately formulated an option's being intrinsically good as its "ought[ing] to exist for its own sake," and this entails that anyone who can realize it ought, other things equal, to do so. We can therefore put the thought behind the open question argument in this way: for any complex of naturalistic or metaphysical properties, it seems we can intelligibly ask of an arbitrary agent in a position to realize this complex, whether she ought to, other things being equal. But what explains why this question should prove so difficult to close? Judgment and constitutive internalists believe the explanation has to do with the necessary connection between unqualified normativity (or normative judgment) and motive.[180]

More strongly, however, Eric Gampel suggests that the *Open Question* argument helps reveal why *a posteriori* identities should be viewed as the "exception, not the rule" regarding the philosophical import of semantic openness. Briefly put, Gampel argues that the referential intensions involved in the terms that underwrite standard *a posteriori* identities are lacking when it comes to moral terms. For example, part of the referential intensions of our term 'water' involves appeal to the causal features of the stuff that fills our lakes and rivers. It is partly this feature of 'water' that explains why we don't get semantically open questions regarding whether the H_20 that is nourishing the plants is really water. But Gampel argues that our terms 'moral,' 'good,' 'right,' and so on, lack this kind of referential intension, leaving questions like Moore's semantically open. Let's consider this argument in more depth.

The key, argues Gampel, is to note why we *don't* get semantic openness in a case like a theory that identifies water with H_20. According to Gampel, such natural kind theories are "guided by the further question: What property E causally explains K's everyday marks, and accounts for K's various causal roles?"[181] It is essentially because

[180] Darwall (1992), p.160.
[181] Gampel (1996), p.199-200.

the referential intension – in this case, concerning our term 'water' – is *in fact* to pick out *whatever* causally explanatory property that *makes* something *water*. But is there a similar referential intension concerning our moral terms, like for 'right,' 'wrong,' 'praiseworthy,' and 'blameworthy'? Gampel says no:

> Even if we discovered some causal/explanatory characterization N for the case of moral rightness, it would not be sufficient to justify reduction. For given the absence of a preexisting intention to pick out a causally explanatory property, why think having N was what we were trying to pick out by our ordinary moral predicate, with its ordinary criteria of employment?[182]

Thus, concludes Gampel, it is this absence of a referential intention to pick out whatever causally explanatory property that makes, say, an action morally wrong—or a person morally blameworthy—that explains why Moore's original open question seems to remain open.

I think it is likely this feature of our moral terms that arguably makes them best analyzed via conceptual role semantics, even if conceptual role semantics is not the best way to capture the content and reference of *non*-moral terms. Something like Gampel's metasemantic account of moral terms is what one would predict if conceptual role semantics is the best method of doing conceptual analysis on the concepts employed within our moral practice.

Before I go into more depth about how this metasemantic account of moral terms, if correct, can help secure the relevance of experimental philosophy for some metaethical questions (in particular, the question over moral internalism), it will be useful to look at how Gampel handles a particular objection. The objection goes like this: it seems one could concede that our moral terms are not associated with any referential intensions that

[182] Ibid.

aim to pick out some causally explanatory property in the world, but then suggest that ethical theorists can nevertheless go on identifying *rightness, wrongness, praiseworthiness,* and *blameworthiness* with whatever causally explanatory properties *happen* to correlate with our use of moral terms. After all (the objection continues) we seem to do something like this regarding those natural kind terms that involve some conflict between the folk, everyday sense and a more scientific sense associated with what we are actually picking out. For instance, many people may continue to say that the metal rod-part of a classroom desk is *cooler* than the wooden table-part of the desk, even though each part has the same average kinetic energy.[183] Surely, the objector presses, we wouldn't want to say that the everyday, folk sense of 'cooler' should serve as a counterexample to the scientific claim that both parts of the desk are equal in temperature. So by analogy, ethical theorists can insist that, say, what's *right* is whatever action is *optimific* regardless of whether our term 'right' has any referential intensions that aim to pick out actions that maximize overall well-being.[184]

Gampel answers this objection by pointing out a relevant disanalogy between scientific theorizing about temperature and ethical theorizing about moral rightness. In the case of the folk resisting the reduction of temperature to average kinetic energy, it seems one could either say that (a) the folk in this case are just confused about temperature or (b) the folk wish to "preserve the commonsense term, with its ordinary criteria of employment, and there's no harm in letting them do so for everyday contexts."[185] But, argues Gampel, (b) doesn't seem to be an option for the ethical theorist,

[183] I borrow this objection and example from Gampel (1996), p. 203.
[184] Ibid.
[185] Ibid.

given that "the primary purpose of ethical discourse is its use in everyday contexts."[186] If Gampel is right, then a ethical revisionist would be left with claiming that the folk are just confused about moral rightness – a very unattractive implication, assuming we want to continue holding the folk morally accountable for their actions.

Moreover, I think it's important to note that the kind of revisionism suggested by the objector just looks like a *change of subject*. While temperature scientists will be able to point to theory-independent reasons to retain their sense of 'temperature' in the face of folk resistance, what theory-independent reasons could ethicists point to in order to buttress such a departure from the everyday sense of moral terms? I suspect that ethicists will have trouble citing reasons that do not turn out to be, on closer investigation, question-begging (i.e., citing moral reasons, or reasons that have moral implications).

My purpose for considering Gampel's response to this objection at length is that I think something similar should be said regarding metaethical inquiry. If ethics is partly defined by what we mean by ethical terms, partly defined by what kinds of conceptual commitments we actually have when we engage in moral discourse, then metaethics—as philosophy *of ethics*—should be sensitive to what we mean, and what conceptual commitments we have, which partly define the object under investigation. In this way, metaethics is essentially a *descriptive* investigation into the very nature of morality *as we find it*. This includes moral discourse and moral judgment. If this were not the case, then most metaethical accounts would seem like non-sequiturs! What would be the point of defending an error theory if your interlocutors can simply *grant it* but then go on to offer

[186] Ibid. E. M. Adams (1960) also makes this point about the inability to distinguish between the everyday context and some other, more scientific, context to serve as a place for ethical theorizing.

a *revisionist* moral realist account of morality? In what sense would this realist account *be an account of morality*? A similar point can be made concerning the debate over moral internalism. What would be the point of defending an internalist theory of the nature of moral judgment if externalists can simply grant that theory but go on to offer a *revisionist* moral externalist account of moral judgment? In what sense would this externalist account *be an account of moral judgment*? I would argue then that, in some ways analogous to the collapse of the theory-level sense with the everyday sense that Gampel argued is the case with ethical theorizing, so too is there no room for revisionist metaethical theorizing given that the everyday phenomenon of morality—characterized in part by our moral talk and judgments—partly *determines* the very object of metaethical inquiry.

With this in mind, I want to now flesh out a bit more why I think that, if Gampel's metasemantic account is correct, we get a relatively straightforward way of demonstrating how experimental philosophy can contribute to metaethics (particularly, the question over moral internalism). And here I want to borrow an insight from Don Loeb concerning another metaethical question: are moral judgments truth-apt descriptive beliefs about the moral realm (so more analogous to our synthetic judgments about the physical realm), or are they instead expressions of our wants/desires/feelings (so more analogous to our gustatory judgments about what tastes good or bad)? Philosophers known as *cognitivists* take the former view, whereas *non-cognitivists* take the latter view. Regarding this metaethical debate, Loeb presses a seemingly meta-metaethical question: what determines whether the cognitivist or non-cognitivist is right? Loeb suggests that, at the end of the day, presumably *empirical* facts about what we are *actually* talking about

when we use moral vocabulary determine whether the cognitivist or the non-cognitivist (or, on Loeb's view – *both*) is correct. This is because, as Loeb explains, while metaphysical truths are not semantic truths, there does seem to be a dependence relation of the former on the latter, at least when it comes to the nature of moral properties:

> It is true that whatever properties exist do so whether we talk about them or not. But those properties are indeed the *moral* properties if and only if they are the properties we are talking about when we talk about morality.[187]

Yet, Loeb continues, any investigation into what "we are talking about when we talk about morality" must presumably involve some empirical investigation into what Loeb calls our "linguistic dispositions": the "intuitions, patterns of thinking and speaking, semantic commitments, and other internal states (conscious or not) of those who employ [moral vocabulary]."[188]

I would argue that part of the conceptual analysis approach is to view experimental philosophy as investigating these "linguistic dispositions." To be sure, one might grant this while nevertheless highlighting the various difficulties involved in empirical inquiry concerning moral and metaethical linguistic dispositions.[189] But, as Loeb puts it:

> [L]ike the drunk searching for his wallet under a streetlight because the light is better there than it was where he dropped it, we cannot expect to succeed if we don't look in the right place.[190]

So, given that (a) metaethics is traditionally offering descriptive accounts of the nature of moral judgment, (b) moral terms are best analyzed through conceptual role semantics,

---

[187] Loeb (2008a), p. 355.
[188] Ibid., p.355-6.
[189] Such as those raised by Kauppinen (2007)
[190] Loeb (2008b), p.381.

and (c) most adults are indeed competent with moral vocabulary, we thus arguably get a better picture of how the conceptual analysis approach grounds the possibility of experimental philosophy contributing to the metaethical question over moral internalism.

Now one may concede the metaethical relevance of information about linguistic dispositions, but then deny that most adult speakers have the appropriate linguistic dispositions.[191] This is the objection to which I will now turn.

4.2. Objection: Linguistic Dispositions are Relevant, but Only When Coming From a Select Population of Speakers

Geoffrey Sayre-McCord offers an argument by analogy to show that certain metaethical disputes seem to retain their intelligibility even if folk concepts differ drastically from those of the metaethicists. This is intended to show that not just any speaker's linguistic dispositions can be relevant to these metaethical questions. Here's the argument: just as the ontological dispute over God's existence would remain coherent even if it was discovered that the folk conception of God is merely "love and mystery," as opposed to an omnipotent creator of the universe, so too would the metaethical dispute over cognitivism and non-cognitivism remain coherent even if it was discovered that the folk conception of moral judgment isn't determinately cognitive nor non-cognitive.[192] "[I]f those concerns and questions were intelligible in the first place," Sayre-McCord concludes, "discovering that the terms one was disposed to use to express them are not

---

[191] Dowell (forthcoming) challenges the very relevance of linguistic dispositions (specifically, semantic intuitions) to metaethics. While I cannot go into this here, it is worth noting that Dowell's argument, if successful, would undermine some traditional methods as well as experimental methods of investigating metaethical questions.

[192] Sayer-McCord (2008), p.406-407.

suitable does nothing to address the concerns and questions."[193] So while he grants that the theories of cognitivists and non-cognitivists should be tracking the linguistic dispositions of those people whom the *metaethicists* view as engaged "in the sort of thought and talk at issue," he denies that this domain of people will encapsulate all "ordinary speakers."[194] And in the event that a significant number of the metaethicists' select people are shown to have drastically different concepts than that held by philosophers, Sayre-McCord points out that such an empirical finding would still be compatible with the notion that "fewer people than we assumed are actually engaging in moral thought and talk."[195] At its most extreme, presumably a single metaethicist might only count *herself* as engaging in moral thought and talk!

Although Sayre-McCord doesn't mention other metaethical disputes, it seems that his argument can reasonably be applied to the question over moral internalism and amoralist skepticism.[196] If this is correct, and if Sayre-McCord's argument is sound, then this would seem to threaten the plausibility of the conceptual analysis approach.

I should first note that I am sympathetic to the notion that certain philosophical disputes need not be affected by folk concepts. For example, I agree that there is surely an intelligible and interesting question about the ontological nature of electrons that doesn't turn on how the folk conceive of electrons (if they even do). Another example: I agree that there is surely an intelligible and interesting question about the nature of knowledge—understood in a post-Gettier sense—despite the fact (if it is a fact) that the

---

193 Ibid., p.407.
194 Sayre-McCord (2008), p.409.
195 Ibid., p.410.
196 Gill (2008), p.388 makes this same point in response to Loeb's argument that the cognitivism/non-cognitivism debate turns largely on empirical facts about what the folk are doing when they make moral claims.

folk happen to be competent only with the pre-Gettier sense of knowledge. This concession is perhaps best expressed in Frank Jackson's (2011) suggestion, in his article "On Gettier Holdouts," that epistemologists could justifiably choose to ignore what he calls "Gettier holdouts":

> It would be a species of chauvinism for analytical philosophers to insist a priori that their concept of knowledge is everyone's. But it isn't a species of chauvinism for analytical philosophers to insist that their concept is better, or better for some purposes, than true justified belief, or that it raises interesting questions that had escaped our attention pre-Gettier. Physicists aren't being chauvinists when they tell us which ways of categorizing systems of particles are important. They are doing their job.[197]

As philosophers, we take ourselves to be in a special position to critically evaluate the different ways in which one could carve up the world and our experiences. Pre-Gettier, it seemed perfectly sensible to view cases of justified true belief as cases of knowledge, and thus all of the normative significance of having knowledge would, we believed, be attained when one had a justified true belief. Post-Gettier, as Jackson points out, it no longer seems *appropriate* to consider agents with justified true beliefs as having knowledge. This sense of appropriateness is, I think, deriving from the philosopher's authority as an expert in the field. Upon considerable reflection over a line of inquiry going back to ancient Greece, we tentatively, but still reasonably, hold that some ways of carving up the world and our experience are better supported than others.

But when it comes to questions in metaethics—questions about the nature of morality—I think the philosopher cannot afford to be quite so cavalier. First, it's important to notice that folk concepts appear to play crucial roles in some metaethical projects. For instance, J. L. Mackie seems to require that folk concepts regarding moral

[197] Jackson (2011), p.479-80.

facts essentially involve both a claim to objectivity and a claim to prescriptivity in order for him to demonstrate that moral claims are systematically rife with ontological error.[198] In another instance we've already discussed, Michael Smith takes his version of moral internalism to correspond to a folk platitude, setting the foundation for his response to what he calls the *moral problem*. The problem is one of reconciling three individually plausible, but together seemingly inconsistent, positions: (1) we take moral facts to be objective facts, (2) we take moral judgments to involve motivation to comply, and (3) the Humean theory of motivation seems right.[199] Notice that (1) and (2) seem to make reference to a shared conception of moral facts and moral judgments.

Second, as Loeb points out, part of what's presently at issue is how *principled* the metaethicists' distinction between "ordinary speakers" and those "engaged in the sort of thought and talk at issue" can be.[200] The existence of pervasive disagreement among metaethicists over the explananda itself (in both the cognitivism/non-cognitivism dispute and our own dispute over moral internalism) seems to tell against the metaethicist's ability to provide a non-question-begging sorting procedure. Furthermore, as Loeb explains, this seems to illustrate how Sayre-McCord's analogy to talk of God "seems misleading." Insofar as 'God' actually does pick out "love and mystery," then questions about the existence of an omnipotent creator of the universe will clearly be *changing the subject*, despite the coherence of such questions.[201] Likewise, insofar as "moral rightness" actually does (aim to) pick out a property that is both objective and prescriptive, then

198 Mackie (1990), p.35.
199 Smith (1994).
200 Loeb (2008b), p.420.
201 Ibid., p.419-420.

questions about the existence of subjective, non-prescriptive properties are questions about a different subject matter, coherent though they may be.[202]

In sum: there is a reasonable doubt that metaethicists can principally (i.e., non-question-beggingly) distinguish between speakers that have the relevant linguistic dispositions and those that do not. And if they decide to ignore this altogether and form an entirely novel question to investigate—free from the pre-theoretical conceptions that underlie our actual moral thought and talk—then it's unclear why this wouldn't just be a change of subject.

5. General Objections to Experimental Philosophy and Replies.

Despite the development of the two approaches sketched above, there's likely to remain some general objections about the relevance of experimental philosophy to traditional philosophical questions that, if correct, would seem to imply that the above approaches must fail for some reason or another. In this final section, I intend to defend the relevance of experimental philosophy to certain metaethical questions against some wholesale objections to experimental philosophy.

[202] Prinz (2007) uses a similar line to argue that morality is *not* objective: "My thesis in this chapter is that morality is not objective. By that I mean that moral concepts, as they currently exist, do not refer to objective properties. In this respect, my project is descriptive [. . . .] if consequentialism could identify an objective source for values, it wouldn't follow that our values are objective—it wouldn't follow that our moral concepts refer to the features that consequentialists define as the foundations of right action. Put more pointedly, if moral vocabulary is fixed by how we use moral terms, then consequentialism is best construed, not as an account of morality, but as an alternative to it" (p.159).

5.1. Objection #1: Philosophical Theories Are Not Constrained By Folk Conceptual Commitments

It has recently been argued that experimental philosophy lacks philosophical significance because the plausibility of philosophical theories does not turn on the conceptual commitments of the folk.[203] But this is simply too quick, at least in respect to metaethics, for there are a number of metaethicists whose theories place emphasis on *our* intuitive responses or on the content of the ordinary platitudes, either implicitly (like Dreier) or explicitly (like Smith). And as Thomas Nadelhoffer and Eddy Nahmias (2007) point out in their recent defense of experimental philosophy:

> Any areas of philosophy that rely on (a) intuition-pumps and thought experiments, (b) appeals to commonsense and pre-philosophical intuition or (c) conceptual analysis based in part on ordinary usage or "platitudes" are ripe for investigation by experimental philosophers who are, above all, interested in examining these things in a controlled and systematic way.[204]

It is presumably uncontroversial that part of the arguments for moral internalism or amoralist skepticism relies on thought experiments, appeals to commonsense intuition, and/or conceptual analysis involving reference to platitudes. Thus, even if this objection gains traction with other philosophical questions, it's not clear that it does so for the question over moral internalism.

---

[203] Timothy Williamson is interpreted, by Joshua Alexander (2010), as defending this position.

Herman Cappelen (2012) argues that philosophical theorizing and argumentation doesn't rely on intuitions at all, much less that of the folk. However, reviewers have criticized Cappelen's characterization of intuitions as too narrow. They have also raised questions about Cappelen's reliance on merely replying to possibly unrepresentative arguments for the relevance of intuitions to philosophy, as opposed to offering positive arguments demonstrating the irrelevance of intuitions. See Ichikawa (2012) and Chalmers (Forthcoming) for these and other criticisms of Cappelen (2012).

[204] Nadelhoffer and Nahmias (2007), p.5

5.2. Objection #2: Philosophical Theories Should Not Be Constrained By Folk Conceptual Commitments

In response to the first objection, I noted that the relevance of folk conceptual commitments to metaethics is assured given that the positions or theorists implicitly or explicitly rely on folk intuitions or folk platitudes. But maybe it seems that I have evaded the *real* objection: namely, that the conceptual commitments of the folk *ought to be* irrelevant to the plausibility of philosophical theories. Insofar as philosophers claim that their theories are held hostage to folk intuitions or platitudes, this objection implies that these philosophers are simply mistaken.

In response, I think it is important to first notice, as Nadelhoffer and Nahmias (2007) point out, that even outspoken *critics* of experimental philosophy still think that philosophers need to take into account folk concepts and intuitions:

> Kauppinen, for instance, asks "why should anybody care about what philosophers do if they just argued about their own inventions?" On his view, because folk intuitions are used at least in part to adjudicate between competing analyses, philosophers ought to be interested in the intuitions of non-partisans—i.e., individuals who are not invested in the philosophical debate that is under investigation. This shows that both Kauppinen and at least some of the experimental philosophers he criticizes agree that conceptual analysis often does and should focus on folk concepts.[205]

Frank Jackson (1998), in his book *From Metaphysics to Ethics*, argues for the connection between metaphysical theorists and folk concepts as being analogous to a bounty hunter and the relevant picture/description of the wanted individual:

> Although metaphysics is about what the world is like, the questions we ask when we do metaphysics are framed in a language, and thus we need to attend to what the users of the language mean by the words they employ

[205] Ibid., p.16.

> to ask their questions. When bounty hunters go searching, they are searching for a person and not a handbill. But they will not get very far if they fail to attend to the representational properties of the handbill on the wanted person. These properties give them their target, or, if you like, define the subject of their search.[206]

The idea, I think, is that most (or all?) metaphysical theories implicitly rely on some kind of conceptual analysis. And, on the plausible assumption that philosophers are attempting to answer generally *shared* questions, it is necessary that philosophers first capture the shared concepts that (partly) [207] make up the meaning of the questions. So, if the question is one that is presumably shared by philosophers and folk alike, then any attempt by philosophers to answer it is (or should be) constrained by the relevant concept applications that (partly) serve to make up the question itself.

5.3. Objection #3: There Aren't Any Folk Conceptual Commitments; Just Folk Beliefs

The present objection is this: instead of their being a folk metaethics made up of conceptual commitments regarding the nature of moral facts and moral judgments—as Smith and Mackie claim—there is really just a lot of explicit metaethical *beliefs*. As such, these folk metaethical beliefs are not essentially tied to our existing moral or metaethical theories and practice any more than folk *physics* beliefs are essentially tied to our existing physics theories and practice.

In response, I will first concede that the folk likely do have some explicit metaethical beliefs (e.g., the relativism espoused by college freshmen). But even so, this

---

[206] Jackson (1998), p.30.

[207] This qualification is necessary in order to leave open the possibility that much (but not all) of semantic meaning is captured by Kripkean-like externalist considerations. The position being defended here is not committed (or need not be committed) to some strong form of descriptivism about meaning. See Jackson (1998) p.37-41.

fact need not imply that Mackie and others are wrong to hold that there are generally-held conceptual commitments to moral facts and moral judgments being of a certain nature. And, at the end of the day, it seems that experimental philosophy is really our only scientifically respectable way of determining whether Mackie and company are right.

Moreover, if the objector is right, then when experimental philosophers turn their (properly tuned) instruments towards folk metaethics, they should end up getting responses that merely reflect the surrounding cultures' positions on these issues. It would be highly unlikely to find cross-culturally shared responses to these studies, at least on the assumption that they lack any conceptual commitments related to metaethics. However, if there is a shared conceptual commitment, then experimental philosophers should find *uniformity* among folk metaethical responses (e.g., compare the uniformity within folk *psychology*). And, as it turns out, this is in fact what some researchers have found to be the case.[208]

6. Metaphilosophy

The debate over what experimental philosophy can tell us about traditional philosophical problems seems to have clear affinities with a much more foundational debate within meta-philosophy: where should philosophers land on the empiricist/rationalist spectrum? Should we be radical empiricists, embracing whatever seems to follow from experience? It seems that experimental philosophy will turn out fine on such a view. Or should we be stubborn rationalists, insisting that our concepts are what *ought* to drive the interpretation

[208] Beebe et al. (2015).

of experience? Perhaps experimental philosophy would have a tougher time on this line of thought.

As one might expect at this point, my view is that philosophers should try to navigate a middle-ground between the two extremes. There's something imprudent, in my view, about defending either full-fledged radical empiricism or the buck-stopping rationalist position. It seems like a philosophical virtue to be open to empirical pressure on some traditional ways of conceiving reality. I think the debate over moral internalism fits this rather well. As Kennett and Fine argued against Smith, there should be some line drawn in the empirical sand where internalism ought to look implausible. To instead hold that no one has ever really made a moral judgment seems to reflect the vice of residing in the buck-stopping rationalist part of the spectrum. But I think Smith is certainly right that not just *any* empirical evidence can give us insight into the nature of moral judgment precisely because internalism is a view about what is to *count* as a moral judgment in the first place. To think that empirical evidence is sufficient to determine the nature of moral judgment seems to reflect the vice of residing in the radical empiricist part of the spectrum.

Of course, there's perhaps a more insidious (in my view) approach to doing philosophy that may suggest a wholesale rejection of experimental philosophy. Some philosophers may feel that they already *know* the answer to some philosophical question, and that their task now is merely to provide the most convincing arguments for their position to their interlocutors. Let's call this the *apologetics* way of doing philosophy (after the movement of the same name within Christianity to provide rational arguments for theistic beliefs). Here, philosophy is being done not in service of finding out, as

Sellars famously said, "how things, in the broadest possible sense of the term, hang together, in the broadest possible sense of the term."[209] Instead, philosophical tools and discussion are used for dialectic purposes *only*.

Some philosophers have charged that this apologetics method may be endemic to entire fields of philosophy, such as the philosophy of religion.[210] One of the main problems with apologetics is its tendency to stifle new avenues of research and new ways of conceiving of old problems. And it's not difficult to see why this would be: these philosophers aren't actually *investigating* anything. No shared philosophical inquiry is being had between the interlocutors. Instead, what we have is something very much like a lawyer defending her client *come what may*.

If this is what some philosophers are doing concerning debates in metaethics, then it may come as no surprise that experimental philosophy would get short shrift. The dialectic approach notwithstanding, how could *any* possible empirical evidence of *any* sort be relevant if you already *know* the *truth*? But I hope I've said enough to show why apologetics is not in the spirit of the Sellarian image of doing philosophy with which I opened this chapter.

[209] Sellars (1962); p.1.
[210] Draper and Nichols (2013).

Chapter 6: Conclusion

This dissertation is the culmination of over four years of research. I have examined how moral internalism and the challenge raised to it by amoralist skepticism have each arisen and developed within traditional metaethical inquiry into the contemporary internalism/externalism debate over the relationship between moral judgment and motivation. I have critically reviewed the experimental challenges raised against moral internalism and in favor of amoralist skepticism, as well as some recent experimental work that appears more sympathetic to moral internalism. I have developed an experimental argument for reworking the debate over moral internalism so that it takes into account the surprising *Factivity Effect*: people's intuitions lean externalist when evaluating factive amoralist scenarios, whereas people's intuitions lean relatively internalist when evaluating non-factive amoralist scenarios. And finally, in a Sellarian effort to keep my "eye on the whole," I have offered some reflections on the place of experimental methodology within traditional philosophical inquiry, culminating in an argument to the effect that experimental results can indeed impact some metaethical questions (even if such results are not relevant to all philosophical questions, or even all metaethical questions).

I would like to conclude this dissertation with some remarks regarding the importance of metaethical inquiry into the relationship between moral judgment and motivation. In chapter 2, I briefly addressed a concern some readers might have over the traditional dispute itself: "why should philosophers debate whether a very strange kind of individual such as the amoralist is conceivable?" While I gave a brief response in that

chapter, I would like to continue my response here. I think part of the answer can be found in how internalism and externalism appear to have drastically different implications for how we view ourselves and our relationship to morality. In his critique of Svavarsdóttir's externalist account of moral motivation, R. Jay Wallace seems to capture the gravity of embracing externalism over internalism:

> The worry I have about this can be put by saying that she [Svavarsdóttir] makes the connection between moral judgment and moral motivation seem altogether optional and arbitrary. Some agents happen to be moved to do what they judge to be morally right or good, while other agents are not so moved; but there is little in the view she advocates that would under gird the claim that those who are morally motivated are somehow responding appropriately or correctly to the moral distinctions they grasp. Perhaps we can explain, in social terms, why it is desirable that people should be brought up to have reliable dispositions to do what they believe to be morally right or good. But from the point of view of practical reason there is nothing to require that an agent who endorses moral claims should be motivated to comply with them. In this respect, the desire to be moral seems to be a mere optional extra, something some of us just happen to have—in this respect rather like a taste for clams or the color azure.[211]

If I understand Wallace's worry correctly, the problem is that externalist accounts of moral motivation (or, at least Svavarsdóttir's particular account) must ultimately end up treating our interest in behaving in accordance with perceived moral duties and obligations as merely contingent features about us. You happen to care about morality while I happen to find it boring and uninteresting—perhaps no different in kind from you happening to care about football while I happen to find football boring and uninteresting. But such an assimilation of moral concern with comparatively trivial concerns seems to flout, as Wallace puts it, the "grain of truth in internalism": "the common thought that the connection between moral judgment and moral motivation is noncontingent."[212]

---

[211] Wallace (2001).
[212] Ibid.

Although Svavarsdóttir responds by saying that such a complaint seems only to show that her account is incomplete,[213] I think the worry is that it is in the very nature of an externalist account that it cannot do justice to the above "grain of truth." On this view, there seems to be no amount of bells and whistles that could be added to an externalist account that would allow for a principled distinction between kinds of contingently-held motivational states. To be sure, perhaps internalists like Wallace may need to re-evaluate what has led them in the first place to find moral internalism to be irreplaceable.

Perhaps the strongest language concerning the importance of this question over moral internalism and amoralist skepticism comes from W. K. Frankena. Given his call to interdisciplinary investigations into this question, I think it's doubly fitting to end this dissertation with his words:

> Here the true character of the opposition appears. Each theory has strengths and weaknesses, and deciding between them involves determining their relative total values as accounts of morality. But such a determination calls for a very broad inquiry. It cannot be based on individual preference. We must achieve "clarity and decision" about the nature and function of morality, of moral discourse, and of moral theory, and this requires not only small-scale analytical inquiries but also studies in the history of ethics and morality, in the relation of morality to society and of society to the individual, as well as in epistemology and in the psychology of human motivation. The battle, if war there be, cannot be contained; its field is the whole human world, and a grand strategy with a total commitment of forces is demanded of each of its participants. What else could a philosopher expect?[214]

[213] Svarvasdóttir (2001).
[214] Frankena (1976), p.73.

Appendix

Description: The following is the complete materials used, and the complete data, from the studies discussed in the paper.

Exploratory Study

Complete materials:

Condition 1:

Please read the following story and answer the questions that follow.

John is an adult of normal intelligence. He doesn't care whether hurting or killing people is morally wrong. Indeed, he hurts and even kills people when he wants their money. John has a long history of violent behavior, dating back to childhood and continuing to the present day. Most recently, he shot a woman when she refused to hand over her wallet during a robbery. Still, John says, "I understand that hurting and killing people is morally wrong."

Does John really understand that hurting and killing people is morally wrong? (Definite No = 1; Definite Yes = 7)

Is it stated in the paragraph above that John says he understands that hurting and killing people is morally wrong? Y/N

Is it stated in the paragraph above that John understands that hurting and killing people is morally wrong? Y/N

Is it stated in the paragraph above that John has never hurt or killed anyone? Y/N

Is it stated in the paragraph above that John doesn't care whether hurting or killing people is morally wrong? Y/N

Please add any comments that you might have in the space below.

Condition 2:

Please read the following story and answer the questions that follow.

John is an adult of normal intelligence. He doesn't care whether hurting or killing people is morally wrong. Indeed, he hurts and even kills people when he wants their money. John has a long history of violent behavior, dating back to childhood and continuing to the present day. Most recently, he shot a woman when she refused to hand over her wallet during a robbery. Still, John says, "I believe that hurting and killing people is morally wrong."

Does John really believe that hurting and killing people is morally wrong? (Definite No = 1; Definite Yes = 7)

Is it stated in the paragraph above that John says he believes that hurting and killing people is morally wrong? Y/N

Is it stated in the paragraph above that John believes that hurting and killing people is morally wrong? Y/N

Is it stated in the paragraph above that John has never hurt or killed anyone? Y/N

Is it stated in the paragraph above that John doesn't care whether hurting or killing people is morally wrong? Y/N

Please add any comments that you might have in the space below.

Condition 3:

Please read the following story and answer the questions that follow.

Dave is an adult of normal intelligence. He doesn't care whether harming the environment is morally wrong. Indeed, he commutes a long distance to work in a car that gets very poor gas mileage. He realizes that this practice is wasteful and harmful to the environment, and he can easily afford to buy a much more fuel-efficient car. Still, Dave says, "I believe that harming the environment is morally wrong."

Please answer the following questions:

Does Dave really believe that harming the environment is morally wrong? (Definite No = 1; Definite Yes = 7)

Is it stated in the paragraph above that Dave says he believes that harming the environment is morally wrong? Y/N

Is it stated in the paragraph above that Dave believes that harming the environment is morally wrong? Y/N

Is it stated in the paragraph above that Dave prefers to drive a high-mileage car? Y/N

Is it stated in the paragraph above that Dave doesn't care whether harming the environment is morally wrong? Y/N

Please add any comments that you might have in the space below.

Condition 4:

Please read the following story and answer the questions that follow.

Chris is an adult of normal intelligence. He doesn't care whether giving his extra money to charity is morally right. Indeed, Chris spends his extra money on video games and fancy clothes, rather than giving it to charity. Still, Chris says, "I believe that giving my extra money to charity is morally right."

Does Chris really believe that giving his extra money to charity is morally right? (Definite No = 1; Definite Yes = 7)

Is it stated in the paragraph above that Chris says he believes that giving his extra money to charity is morally right? Y/N

Is it stated in the paragraph above that Chris believes that giving his extra money to charity is morally right? Y/N

Is it stated in the paragraph above that Chris gives his extra money to charity? Y/N

Is it stated in the paragraph above that Chris doesn't care whether giving his extra money to charity is morally right? Y/N

Please add any comments that you might have in the space below.

2x2 Factorial Study

Complete Materials:

Group 1: Understands/right

Chris is a 21-year-old undergraduate student. He and his friend Bill are in line to buy a video game. At the counter is a charity donation jar with only $30 in it. Chris has $20 for the game, but Bill argues that giving it to charity is the morally right thing to do. After listening to Bill's argument, Chris says, "I agree, Bill - I understand that giving my money to charity is the morally right thing to do; however, I don't care at all about doing the morally right thing." After saying this, Chris buys the video game with his $20.

Chris understands that giving his money to charity is the morally right thing to do. (Definite No = 1; Definite Yes = 7)

Chris understands that giving his money to charity is the ethical thing to do.

Chris understands that he should give his money to charity.

Is it stated in the paragraph above that Chris says the following: "I understand that giving my money to charity is the morally right thing to do"? Y/N

Is it stated in the paragraph above that Chris understands that giving his money to charity is the morally right thing to do? Y/N

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that telling the clerk about the mistake is the morally right thing to do. After listening to Sandy's argument, Jane says, "I agree, Sandy - I understand that telling the clerk about the mistake is the morally right thing to do; however, I don't care at all about doing the morally right thing." After saying this, Jane pays $15 and leaves with the headphones.

Jane understands that telling the clerk about the mistake is the morally right thing to do. (Definite No = 1; Definite Yes = 7)

Jane understands that telling the clerk about the mistake is the ethical thing to do.

Jane understands that she should tell the clerk about the mistake.

Is it stated in the paragraph above that Jane says the following: "I understand that telling the clerk about the mistake is the morally right thing to do"? Y/N

Is it stated in the paragraph above that Jane understands that telling the clerk about the mistake is the morally right thing to do? Y/N

Group 2: Believes/right

Chris is a 21-year-old undergraduate student. He and his friend Bill are in line to buy a video game. At the counter is a charity donation jar with only $30 in it. Chris has $20 for the game, but Bill argues that giving it to charity is the morally right thing to do. After listening to Bill's argument, Chris says, "I agree, Bill - I believe that giving my money to charity is the morally right thing to do; however, I don't care at all about doing the morally right thing." After saying this, Chris buys the video game with his $20.

Chris believes that giving his money to charity is the morally right thing to do. (Definite No = 1; Definite Yes = 7)

Chris believes that giving his money to charity is the ethical thing to do.

Chris believes that he should give his money to charity.

Is it stated in the paragraph above that Chris says the following: "I believe that giving my money to charity is the morally right thing to do"? Y/N

Is it stated in the paragraph above that Chris believes that giving his money to charity is the morally right thing to do? Y/N

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that telling the clerk about the mistake is the morally right thing to do. After listening to Sandy's argument, Jane says, "I agree, Sandy - I believe that telling the clerk about the mistake is the morally right thing to do; however, I don't care at all about doing the morally right thing." After saying this, Jane pays $15 and leaves with the headphones.

Jane believes that telling the clerk about the mistake is the morally right thing to do. (Definite No = 1; Definite Yes = 7)

Jane believes that telling the clerk about the mistake is the ethical thing to do.

Jane believes that he should tell the clerk about the mistake.

Is it stated in the paragraph above that Jane says the following: "I believe that telling the clerk about the mistake is the morally right thing to do"? Y/N

Is it stated in the paragraph above that Jane believes that telling the clerk about the mistake is the morally right thing to do? Y/N

Group 3: Understands/wrong

Chris is a 21-year-old undergraduate student. He and his friend Bill are in line to buy a video game. At the counter is a charity donation jar with only $30 in it. Chris needs $20 for the game, but Bill argues that taking it from the jar is morally wrong. After listening to Bill's argument, Chris says, "I agree, Bill - I understand that taking money from the jar is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Chris buys the video game with $20 from the jar.

Chris understands that taking money from the jar is morally wrong. (Definite No = 1; Definite Yes = 7)

Chris understands that taking money from the jar is unethical.

Chris understands that he shouldn't take money from the jar.

Is it stated in the paragraph above that Chris says the following: "I understand that taking money from the jar is morally wrong"? Y/N

Is it stated in the paragraph above that Chris understands that taking money from the jar is morally wrong? Y/N

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - I understand that buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones.

Jane understands that buying the headphones at the mistaken price is morally wrong. (Definite No = 1; Definite Yes = 7)

Jane understands that buying the headphones at the mistaken price is unethical.

Jane understands that she shouldn't buy the headphones at the mistaken price.

Is it stated in the paragraph above that Jane says the following: "I understand that buying the headphones at the mistaken price is morally wrong"? Y/N

Is it stated in the paragraph above that Jane understands that buying the headphones at the mistaken price is morally wrong? Y/N

Group 4: Believes/wrong

Chris is a 21-year-old undergraduate student. He and his friend Bill are in line to buy a video game. At the counter is a charity donation jar with only $30 in it. Chris needs $20 for the game, but Bill argues that taking it from the jar is morally wrong. After listening to Bill's argument, Chris says, "I agree, Bill - I believe that taking money from the jar is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Chris buys the video game with $20 from the jar.

Chris believes that taking money from the jar is morally wrong. (Definite No = 1; Definite Yes = 7)

Chris believes that taking money from the jar is unethical.

Chris believes that he shouldn't take money from the jar.

Is it stated in the paragraph above that Chris says the following: "I believe that taking money from the jar is morally wrong"? Y/N

Is it stated in the paragraph above that Chris believes that taking money from the jar is morally wrong? Y/N

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - I believe that buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones.

Jane believes that buying the headphones at the mistaken price is morally wrong. (Definite No = 1; Definite Yes = 7)

Jane believes that buying the headphones at the mistaken price is unethical.

Jane believes that she shouldn't buy the headphones at the mistaken price.

Is it stated in the paragraph above that Jane says the following: "I believe that buying the headphones at the mistaken price is morally wrong"? Y/N

Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N

Complete data - results for the Items Averaged, for both between-subjects and within-subjects results.
Legend for Significance Indication: < .05 = *; < .01 = **; < .001 = ***

Between-subjects results

Figure 1 – Video Game Items Averaged.

Items Averaged	*F(1, 288)*	*p*	η^2
Attitude	61.41	<.001***	.178
Valence	1.42	.23	.005
Attitude x Valence	.16	.70	.001

Figure 2 – Video Game Items Averaged.

Attitude	**Vignette**	*Mean*	*Std. Deviation*	*N*
Understand	Right	6.00	.97	76
	Wrong	6.12	1.41	67
	Total	6.05	1.19	143
Believe	Right	4.72	1.42	75
	Wrong	4.97	1.44	70
	Total	4.84	1.43	145
Total	Right	5.36	1.37	151
	Wrong	5.53	1.53	137
	Total	5.44	1.45	288

Figure 3 – Headphones Items Averaged.

Items Averaged	*F(1, 288)*	*p*	η^2
Attitude	96.19	<.001***	.253
Valence	7.48	.007**	.026
Attitude x Valence	1.30	.26	.005

Figure 4 – Headphones Items Averaged.

Attitude	**Vignette**	*Mean*	*Std. Deviation*	*N*
Understand	Right	6.37	.81	76
	Wrong	5.82	1.25	67
	Total	6.11	1.07	143
Believe	Right	4.81	1.21	75
	Wrong	4.59	1.48	70
	Total	4.70	1.35	145
Total	Right	5.60	1.29	151
	Wrong	5.19	1.51	137
	Total	5.40	1.41	288

Within-subjects results

Figure 5 – Understands Right Condition

Item	*F(1,288)*	*p*	η^2
Vignette: Items Averaged	12.49	.001**	.129

Figure 6 – Understands Right Condition

Item	**Vignette**	*Mean*	*Std. Deviation*
Items Averaged	Video Game	6.00	.103
	Headphones	6.35	.092

Figure 7 – Understands Wrong Condition

Item	*F(1, 288)*	*p*	η^2
Vignette: Items Averaged	9.66	.003**	.105

Figure 8 – Understands Wrong Condition

Item	**Vignette**	*Mean*	*Std. Deviation*
Items Averaged	Video Game	6.19	.144
	Headphones	5.93	.131

Figure 9 – Believes Right Condition

Item	*F(1, 288)*	*p*	η^2
Vignette: Items Averaged	.84	.36	.010

Figure 10 – Believes Right Condition

Item	**Vignette**	*Mean*	*Std. Deviation*
Items Averaged	Video Game	4.82	.155
	Headphones	4.93	.137

Figure 11 – Believes Wrong Condition

Item	*F(1, 288)*	*p*	η^2
Vignette: Items Averaged	6.66	.01*	.076

Figure 12 – Believes Wrong Condition

Item	**Vignette**	*Mean*	*Std. Deviation*
Items Averaged	Video Game	5.13	.159
	Headphones	4.77	.170

Deflationary Explanation #1: The Inverted-Commas Response (ICR) Hypothesis

Complete materials:

Marijuana/Understands

Please read the following story:

Chris is a 21-year-old college student. He and his friend Bill are in their dorm room. From a box hidden under his bed, Chris pulls out some marijuana. Chris prepares to smoke it, but Bill argues that using marijuana recreationally is morally wrong. Chris replies that it's controversial whether smoking marijuana for enjoyment is immoral, and Bill concedes that there's currently no consensus on the matter, but Bill continues to make his case. After listening to Bill's argument, Chris says, "I agree, Bill – recreational marijuana use is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Chris proceeds to smoke his hidden stash of marijuana.

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

Chris understands that recreational marijuana use is morally wrong.

Chris understands that recreational marijuana use is unethical.

Chris understands that he shouldn't use marijuana recreationally.

Is it stated in the paragraph above that Chris says the following: "I agree, Bill – recreational marijuana use is morally wrong"? Y/N

Is it stated in the paragraph above that Chris understands that recreational marijuana use is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

In my personal view, it's morally wrong for anyone to use marijuana recreationally.

Abortion/Understands

Please read the following:

Jane is a 24-year-old graduate student. She and her friend Sandy are outside a clinic that performs abortions. Jane intends to end her pregnancy because she doesn't want to be a parent, but Sandy argues that an abortion for that reason is morally wrong. Jane replies that it's controversial whether abortion to avoid parenthood is immoral, and Sandy concedes that there's currently no consensus on the matter, but Sandy continues to make her case. After listening to Sandy's argument, Jane says, "I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane enters the clinic and a doctor ends her pregnancy as she requests.

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

Jane understands that ending her pregnancy to avoid parenthood is morally wrong.

Jane understands that ending her pregnancy to avoid parenthood is unethical.

Jane understands that she shouldn't end her pregnancy to avoid parenthood.

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong"? Y/N

Is it stated in the paragraph above that Jane understands that ending her pregnancy to avoid parenthood is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

In my personal view, it's morally wrong for anyone to end their pregnancy to avoid parenthood.

Marijuana/Believes

Please read the following story:

Chris is a 21-year-old college student. He and his friend Bill are in their dorm room. From a box hidden under his bed, Chris pulls out some marijuana. Chris prepares to smoke it, but Bill argues that using marijuana recreationally is morally wrong. Chris replies that it's controversial whether smoking marijuana for enjoyment is immoral, and Bill concedes that there's currently no consensus on the matter, but Bill continues to make his case. After listening to Bill's argument, Chris says, "I agree, Bill – recreational marijuana use is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Chris proceeds to smoke his hidden stash of marijuana.

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

Chris believes that recreational marijuana use is morally wrong.

Chris believes that recreational marijuana use is unethical.

Chris believes that he shouldn't use marijuana recreationally.

Is it stated in the paragraph above that Chris says the following: "I agree, Bill – recreational marijuana use is morally wrong"? Y/N

Is it stated in the paragraph above that Chris believes that recreational marijuana use is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

In my personal view, it's morally wrong for anyone to use marijuana recreationally.

Abortion/Believes

Please read the following:

Jane is a 24-year-old graduate student. She and her friend Sandy are outside a clinic that performs abortions. Jane intends to end her pregnancy because she doesn't want to be a parent, but Sandy argues that an abortion for that reason is morally wrong. Jane replies that it's controversial whether abortion to avoid parenthood is immoral, and Sandy concedes that there's currently no consensus on the matter, but Sandy continues to make her case. After listening to Sandy's argument, Jane says, "I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane enters the clinic and a doctor ends her pregnancy as she requests.
Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

Jane believes that ending her pregnancy to avoid parenthood is morally wrong.

Jane believes that ending her pregnancy to avoid parenthood is unethical.

Jane believes that she shouldn't end her pregnancy to avoid parenthood.

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong"? Y/N

Is it stated in the paragraph above that Jane believes that ending her pregnancy to avoid parenthood is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

In my personal view, it's morally wrong for anyone to end their pregnancy to avoid parenthood.

Complete data - results for the Items Averaged.

Legend for Significance Indication: < .05 = *; < .01 = **; < .001 = ***

Figure 1 – Items Averaged.

Items Averaged	*F(1,77)*	*p*	η^2
Attitude	*16.32*	*<.001***	*.183*
Vignette	.17	.68	.002
Attitude x Vignette	1.78	.185	.024

Figure 2 – Items Averaged.

Attitude	**Vignette**	*Mean*	*Std. Deviation*	*N*
Understand	Marijuana	5.50	1.12	14
	Abortion	5.79	1.14	26
	Total	5.69	1.13	40
Believe	Marijuana	4.64	1.87	11
	Abortion	4.08	1.28	26
	Total	4.24	1.48	37
Total	Marijuana	5.12	1.53	25
	Abortion	4.94	1.48	52
	Total	4.99	1.49	77

Deflationary Explanation #2: the Dispositional Belief Hypothesis

Complete materials:

Rightness/Understands

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that telling the clerk about the mistake is the morally right thing to do. After listening to Sandy's argument, Jane says, "I agree, Sandy - telling the clerk about the mistake is the morally right thing to do; however, I don't care at all about doing the morally right thing." After saying this, Jane pays $15 and leaves with the headphones. When they get back to their apartment, Sandy takes a nap while Jane listens to music with her new headphones.

Please answer the following reading comprehension questions:

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - telling the clerk about the mistake is the morally right thing to do"? Y/N

Is it stated in the paragraph above that Jane understands that telling the clerk about the mistake is the morally right thing to do? Y/N

Is it stated in the paragraph above that Sandy argues that telling the clerk about the mistake is the morally right thing to do? Y/N

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

Sandy (despite being asleep) understands that telling the clerk about the mistake was the morally right thing to do.

Sandy (despite being asleep) understands that telling the clerk about the mistake was the ethical thing to do.

Sandy (despite being asleep) understands that Jane should have told the clerk about the mistake.

Jane understands that telling the clerk about the mistake was the morally right thing to do.

Jane understands that telling the clerk about the mistake was the ethical thing to do.

Jane understands that she should have told the clerk about the mistake.

Rightness/Believes

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that telling the clerk about the mistake is the morally right thing to do. After listening to Sandy's argument, Jane says, "I agree, Sandy - telling the clerk about the mistake is the morally right thing to do; however, I don't care at all about doing the morally right thing." After saying this, Jane pays $15 and leaves with the headphones. When they get back to their apartment, Sandy takes a nap while Jane listens to music with her new headphones.

Please answer the following reading comprehension questions:

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - telling the clerk about the mistake is the morally right thing to do"? Y/N

Is it stated in the paragraph above that Jane believes that telling the clerk about the mistake is the morally right thing to do? Y/N

Is it stated in the paragraph above that Sandy argues that telling the clerk about the mistake is the morally right thing to do? Y/N

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

Sandy (despite being asleep) believes that telling the clerk about the mistake was the morally right thing to do.

Sandy (despite being asleep) believes that telling the clerk about the mistake was the ethical thing to do.

Sandy (despite being asleep) believes that Jane should have told the clerk about the mistake.

Jane believes that telling the clerk about the mistake was the morally right thing to do.

Jane believes that telling the clerk about the mistake was the ethical thing to do.

Jane believes that she should have told the clerk about the mistake.

Wrongness/Understands

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones. When they get back to their apartment, Sandy takes a nap while Jane listens to music with her new headphones.

Please answer the following reading comprehension questions:

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - buying the headphones at the mistaken price is morally wrong"? Y/N

Is it stated in the paragraph above that Jane understands that buying the headphones at the mistaken price is morally wrong? Y/N

Is it stated in the paragraph above that Sandy argues that buying the headphones at the mistaken price is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Strongly Disagree = 1; Strongly Agree = 7):

Sandy (despite being asleep) understands that buying the headphones at the mistaken price was morally wrong.

Sandy (despite being asleep) understands that buying the headphones at the mistaken price was unethical.

Sandy (despite being asleep) understands that Jane shouldn't have bought the headphones at the mistaken price.

Jane understands that buying the headphones at the mistaken price was morally wrong.

Jane understands that buying the headphones at the mistaken price was unethical.

Jane understands that she shouldn't have bought the headphones at the mistaken price.

Wrongness/Believes

Jane is a 24-year-old graduate student. She and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones. When they get back to their apartment, Sandy takes a nap while Jane listens to music with her new headphones.

Please answer the following reading comprehension questions (Strongly Disagree = 1; Strongly Agree = 7):

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - buying the headphones at the mistaken price is morally wrong"? Y/N

Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N

Is it stated in the paragraph above that Sandy argues that buying the headphones at the mistaken price is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements:

Sandy (despite being asleep) believes that buying the headphones at the mistaken price was morally wrong.

Sandy (despite being asleep) believes that buying the headphones at the mistaken price was unethical.

Sandy (despite being asleep) believes that Jane shouldn't have bought the headphones at the mistaken price.

Jane believes that buying the headphones at the mistaken price was morally wrong.

Jane believes that buying the headphones at the mistaken price was unethical.

Jane believes that she shouldn't have bought the headphones at the mistaken price.

Complete data - results for the Items Averaged

Legend for Significance Indication: < .05 = *; < .01 = **; < .001 = ***

Figure 1 – Items Averaged.

Items Averaged	*F(1,307)*	*p*	η^2
Sandy Attitude	.05	.83	<.001
Jane Attitude	94.16	<.001***	.237
Sandy Valence	.15	.70	.001
Jane Valence	.72	.40	.002
Sandy Attitude x Valence	4.35	.04*	.014
Jane Attitude x Valence	.6	.44	.002

Figure 2 – Sleeping Sandy Items Averaged.

Sandy Average Attitude	**Valence**	*Mean*	*Std. Deviation*	*N*
Understand	Right	6.16	1.31	76
	Wrong	6.37	.89	79
	Total	6.27	1.12	155
Believe	Right	6.40	.99	78
	Wrong	6.08	1.19	74
	Total	6.24	1.10	152
Total	Right	6.28	1.16	154
	Wrong	6.23	1.05	153
	Total	6.26	1.11	307

Figure 3 – Jane Items Averaged.

Jane Average Attitude	**Valence**	*Mean*	*Std. Deviation*	*N*
Understand	Right	5.85	1.41	76
	Wrong	5.84	1.16	79
	Total	5.84	1.28	155
Believe	Right	4.55	1.31	78
	Wrong	4.31	1.20	74
	Total	4.43	1.26	152
Total	Right	5.19	1.50	154
	Wrong	5.10	1.40	153
	Total	5.14	1.45	307

Deflationary Explanation #3: Alternative Factives/Non-factives

Complete materials:

Group 1: Factive – 'knows'
Jane and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones.

Please rate your agreement/disagreement with the following statement (Strongly Disagree = 1; Strongly Agree = 7):

Jane knows that she shouldn't buy the headphones at the mistaken price.

Please answer the following reading comprehension questions:

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - buying the headphones at the mistaken price is morally wrong"? Y/N

Is it stated in the paragraph above that Jane knows that buying the headphones at the mistaken price is morally wrong? Y/N

Group 2: Non-factive – 'thinks'
Jane and her friend Sandy are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy - buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones.

Please rate your agreement/disagreement with the following statement (Strongly Disagree = 1; Strongly Agree = 7):

Jane thinks that she shouldn't buy the headphones at the mistaken price.

Please answer the following reading comprehension questions:

Is it stated in the paragraph above that Jane says the following: "I agree, Sandy - buying the headphones at the mistaken price is morally wrong"? Y/N

Is it stated in the paragraph above that Jane thinks that buying the headphones at the mistaken price is morally wrong? Y/N

A Substantive Explanation: the Moral Emotions Hypothesis

Complete materials:

Group 1: Understands

A customer in front of Chris unknowingly drops a $20 bill, and Chris needs $20 to buy a video game. Chris understands that taking the $20 is morally wrong. Nevertheless, Chris proceeds to buy the video game with the dropped $20 bill.

Please answer the following reading comprehension question:

Is it stated in the paragraph above that Chris understands that taking the $20 is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Completely Disagree = 1; Completely Agree = 7):

Chris felt bad about taking the $20.

Chris felt regret about taking the $20.

Chris felt guilty about taking the $20.

Unbeknownst to the clerk, the headphones Jane wants to buy mistakenly ring up $15 instead of $150. Jane understands that buying the headphones at the mistaken price is morally wrong. Nevertheless, Jane proceeds to pay $15 and leaves with the headphones.

Please answer the following reading comprehension question:

Is it stated in the paragraph above that Jane understands that buying the headphones at the mistaken price is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Completely Disagree = 1; Completely Agree = 7):

Jane felt bad about buying the headphones.

Jane felt regret about buying the headphones.

Jane felt guilty about buying the headphones.

Group 2: Believes

A customer in front of Chris unknowingly drops a $20 bill, and Chris needs $20 to buy a video game. Chris believes that taking the $20 is morally wrong. Nevertheless, Chris proceeds to buy the video game with the dropped $20 bill.

Please answer the following reading comprehension question:
Is it stated in the paragraph above that Chris believes that taking the $20 is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Completely Disagree = 1; Completely Agree = 7):

Chris felt bad about taking the $20.

Chris felt regret about taking the $20.

Chris felt guilty about taking the $20.

Unbeknownst to the clerk, the headphones Jane wants to buy mistakenly ring up $15 instead of $150. Jane believes that buying the headphones at the mistaken price is morally wrong. Nevertheless, Jane proceeds to pay $15 and leaves with the headphones.

Please answer the following reading comprehension question:

Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N

Please rate your agreement/disagreement with the following statements (Completely Disagree = 1; Completely Agree = 7):

Jane felt bad about buying the headphones.

Jane felt regret about buying the headphones.

Jane felt guilty about buying the headphones.

Complete data - results for the Items Averaged

Legend for Significance Indication: < .05 = *; < .01 = **; < .001 = ***

Figure 21 – Items Averaged.

Items Averaged	*F(1,139)*	*p*	η^2
Attitude	6.54	.01*	.046
Vignette	4.64	.03*	.033
Attitude x Vignette	2.61	.31	.008

Figure 22 – Items Averaged.

Attitude	**Vignette**	*Mean*	*Std. Deviation*	*N*
Understand	Video game	3.53	1.46	25
	Headphones	2.65	1.50	36
	Total	3.01	1.54	61
Believe	Video game	3.97	1.64	29
	Headphones	3.65	1.70	49
	Total	3.77	1.67	78
Total	Video Game	3.77	1.56	54
	Headphones	3.22	1.68	85
	Total	3.43	1.65	139

Bibliography

Adams, E. M. (1960). *Ethical Naturalism and the Modern World-view*. Chapel Hill, University of North Carolina Press.

Alexander, J. (2010). Is experimental philosophy philosophically significant? *Philosophical Psychology*, 23, 377–389.

Alfano, M. & Loeb, D. (2014) Experimental moral philosophy. *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta (ed.), URL =<http://plato.stanford.edu/archives/sum2014/entries/experimental-moral/>.

Appiah, K. A. (2008). *Experiments in Ethics*. Cambridge, Harvard University Press.

Baldwin, T., "George Edward Moore", *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2010/entries/moore/>.

Baumeister R. F., Masicampo E. J., & DeWall C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin* 35, 260–62.

Bedke, M. S. (2009). Moral judgment purposivism: saving internalism from amoralism. *Philosophical Studies*, 144, 189-209.

Beebe, J. & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind and Language*. 25, 474–498.

Beebe, J., Qiaoan, R., Wysocki, T., & Endara, M. A. (2015). Moral objectivism in cross-cultural perspective. *Journal of Cognition and Culture*, 15, 386-401.

Björklund, F., Björnsson, G., Eriksson, J., Olinder, R.F., & Strandberg, C. (2012). Recent work: motivational internalism. *Analysis*, 72, 124-137.

Björnsson, G. (2002). How emotivism survives immoralists, irrationality and depression. *The Southern Journal of Philosophy*, 60, 327-344.

Björnsson, G., Eriksson, J., Strandberg, C., Olinder, R.F., & Björklund, F. (2015). Motivational internalism and folk intuitions. *Philosophical Psychology*, 28, 715-734.

Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford, UK: Clarendon Press.

Boyd, R. (1988). How to be a moral realist, in *Essays on Moral Realism*, Sayre-McCord, G. Ed., 187-228, Ithaca, NY: Cornell UP 1988.

Brink, D. (1989). *Moral Realism and the Foundations of Ethics*. New York, NY: Cambridge UP.

Buckwalter, W., Rose, D., & Turri, J. (2013). Belief through thick and thin. *Noûs*. DOI: 10.1111/nous.12048.

Cappellen, H. (2012). *Philosophy Without Intuitions*. New York, NY: Oxford UP.

Chalmers, D. (2014). Intuitions: a minimal defense. *Philosophical Studies*, 171, 535-544.

Cholbi, M. (2006). Moral belief attribution: a reply to Roskies. *Philosophical Psychology*, 19, 629–638.

DeLapp, K. (2016). "Metaethics", *The Internet Encyclopedia of Philosophy*, ISSN 2161-002, http://www.iep.utm.edu/

Dowell, J. (forthcoming). The metaethical insignificance of moral twin earth" In *Oxford Studies in Metaethics, Vol. 11*, Shafer-Landau, R., ed., Oxford: OUP.

Draper, P. & Nichols, R. (2013). Diagnosing bias in philosophy of religion. *The Monist*, 96, 420-444.

Dreier, J. (1990). Internalism and speaker relativism. *Ethics*, 101, 6-26.

Frankena, W. K. (1976). *Perspective on Morality: Essays by William K. Frankena*. Ed. Goodpaster K. E., Notre Dame, IN: U of Notre Dame P.

Feltz A. & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Conscious Cognition*. 18:342–50

Gampel, E. H., (1996). A defense of the autonomy of ethics: why value is not like water. *Canadian Journal of Philosophy*, 26, 191-210.

Gill, M. B. (2008). Metaethical invariability, incoherence, and error. In *Moral Psychology, Vol. 2 - The Cognitive Science of Morality: Intuition and Diversity*, ed. Sinnott-Armstrong, W., 387-402, Cambridge, MA (2008): The MIT Press.

Goodwin, G. P. & Darley, J. M. (2008). The psychology of metaethics: exploring objectivism. *Cognition*, 106, 1339-1366.

--- (2010). The perceived objectivity of ethical beliefs: psychological findings and implications for public policy. *Review of Philosophy and Psychology*, 1, 1–28.

--- (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, 48, 250-256.

Greenberg, M. & Harman, G. (2006). Conceptual role semantics. In *Oxford Handbook of Philosophy of Language*, Lepore, E. and Smith, B. eds., (2006): Oxford University Press.

Hare, R. M. (1969). *The Language of Morals*. Oxford, UK: OUP.

---. (1981). *Moral Thinking: Its Levels, Method, and Point*. Oxford, UK: Clarendon Press.

Horgan, T. & Timmons, M. (1992). Troubles on moral twin earth: moral queerness revived. *Synthese*, 92, 221-260.

Hurka, T., "Moore's Moral Philosophy", *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2015/entries/moore-moral/>.

Ichikawa, J. J. (2012). Review of *Philosophy Without Intuitions*. Unpublished manuscript.

Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford: OUP.

---. (2011). On Gettier holdouts. *Mind & Language*, 26, 468-481.

Johnson, R. (1999). Internal reasons and the conditional fallacy. *The Philosophical Quarterly*, 49, 53-71.

Joyce, R. (2008). What neuroscience can (and cannot) contribute to metaethics. In Sinnott-Armstrong (2008b), 371-394.

Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations*, 10, 95–118.

Kennett, J. & Fine, C. (2008a). Internalism and the evidence from psychopaths and "acquired sociopaths." In Sinnott-Armstrong (2008b), 173-190.

---. (2008b). Could there be an empirical test for internalism? In Sinnott-Armstrong (2008b), 217-225.

Knobe, J., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H., & Sommers, T. (2012). Experimental philosophy. *Annual Review of Psychology*, 63, 81-99.

Knobe, J. & Burra, A. (2006). Intention and intentional action: a cross-cultural study. *Journal of Cognition and Culture*, 6, 113–132.

Korsgaard, C. (1986). Skepticism about practical reason. *The Journal of Philosophy*, 83, 5-25.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard UP.

LaPorte, J., "Rigid Designators", *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2016/entries/rigid-designators/>

Leben, D. & Wilckens, K. (2015). Pushing the intuitions behind moral internalism. *Philosophical Psychology*, 28, 510-528.

Lenman, J. (1999). The externalist and the amoralist. *Philosophia*, 27, 441-457.

---. (2008). Moral naturalism, *The Stanford Encyclopedia of Philosophy (Winter 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2008/entries/naturalism-moral/>.

Loeb, D. (2008a). Moral incoherentism: how to pull a metaphysical rabbit out of a semantic hat. In Sinnott-Armstrong (2008a), 355-386.

---. (2008b). Reply to Gill and Sayer-McCord. In Sinnott-Armstrong (2008a), 413-422.

Mackie, J. L. (1990). *Ethics: inventing right and wrong*. New York, NY: Penguin.

McDowell, J. (1978). Are moral requirements hypothetical imperatives? *Proceedings of the Aristotelian Society, Supplementary Volumes*, 52, 13-29.

Machery, E. (2008). The folk concept of intentional action: philosophical and experimental issues. *Mind & Language*, 23, 165–189.

Mallon, R. (2008). Knobe versus Machery: testing the trade-off hypothesis. *Mind and Language*, 23, 247–255

Miller, A. (2013). *Contemporary Metaethics: An Introduction 2nd Edition*. Malden, MA: Polity Press.

Milo, R. (1981). Moral indifference. *The Monist*, 64, 373-393.

Murray, D., Sytsma, J., & Livengood, J. (2013) God knows (but does God believe?). *Philosophical Studies*, 166, 83-107.

Myers-Schutz, B. & Schwitzgebel, E. (2013). Knowing that p without believing that p. *Noûs*, 47, 371-384.

Nadelhoffer, T. & Nahmias, E. (2007). The past and future of experimental philosophy. *Philosophical Explorations*, 10, 123-149.

Nahmias, E., Coates, D. & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: experiments on folk intuitions. *Midwest Studies in Philosophy*, 31, 214–242.

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J., (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73, 28-53.

Nahmias, E. & Murray, D. (2010). Experimental philosophy on free will: an error theory or incompatibilist intuitions. In *New Waves in Philosophy of Action*, Aguilar, J., Buckareff, A., and Frankish, K., eds., Hampshire, UK: Palgrave-Macmillan.

Nichols, S. (2002). How psychopaths threaten moral rationalism: is it irrational to be amoral? In *Moral Psychology: Historical and Contemporary Readings*, Nadelhoffer, T., Nahmias, E., and Nichols, S., eds., 73-83, West Sussex, United Kingdom (2010): Blackwell Publishing Ltd.

---. (2004a). After objectivity: an empirical study of moral judgment. *Philosophical Psychology*, 17, 3–26.

---. (2004b). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York, NY: Oxford University Press.

---. (2008). Moral rationalism and empirical immunity: comments on Joyce. In Sinnott-Armstrong (2008b), 395-407.

Philips, J. and Worsnip, A. Motivation internalism. Unpublished manuscript.

Prinz, J. (2007). *The emotional construction of morals*. New York, NY: Oxford UP.

Railton, P. (1986). Moral realism. *The Philosophical Review*, 95, 163-207.

Rakoczy, H., Behne, T., Clüver, A., Dallmann, S., Weidner, S. & Waldmann, M. (2015). The side-effect effect in children is robust and not specific to the moral status of action effects. *PLoS ONE* ,10, 1-10.

Robbins, P. & Jack, A. (2006). The phenomenal stance. *Philosophical Studies*, 127, 59–85.

Rosati, C. (2008). Moral motivation. *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), ed. Zalta, E., URL = <http://plato.stanford.edu/archives/fall2008/entries/moral-motivation/>.

Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy." *Philosophical Psychology*, 16, 51-66.

---. (2006). Patients with ventromedial frontal damage have moral beliefs. *Philosophical Psychology*, 19, 617–627.

---. (2008). Internalism and the evidence from pathology. In Sinnott-Armstrong (2008b), 191-206.

Roskies, A. & Nichols, S. (2008). Bringing moral responsibility down to earth. *Journal of Philosophy* 105, 371-388.

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind & Language*, 26, 482–505.

Sayre-McCord, G. (2008). Moral semantics and empirical enquiry. In Sinnott-Armstrong (2008a), 403-412.

Schwitzgebel, E. & Cushman, F. (2012). "Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers." *Mind and Language*, 27:2, 135–153.

Sellars, W. (1963). *Science, Perception and Reality*. New York: Humanities Press.

Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford, UK: Clarendon Press.

Sinnott-Armstrong, W. (ed.), (2008a), *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity*. (Volume 2), Cambridge: MIT Press.

---. (ed.), (2008b), *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (Volume 3), Cambridge: MIT Press.

Smith. M. (1994). *The Moral Problem*. Cambridge, MA: Blackwell.

---. (2008). The truth about internalism. In Sinnott-Armstrong (2008b), 207-216.

Stocker, M. (1979). Desiring the bad: an essay in moral psychology. *The Journal of Philosophy*, 76, 738-753.

Strandberg C. & Björklund, F. (2012). Is moral internalism supported by folk intuitions? *Philosophical Psychology*, 26, 319-335.

Svavarsdóttir, S. (1999). Moral cognitivism and motivation. *The Philosophical Review*, 108, 161-219.

---. (2001). Reply to commentators. *Brown Electronic Article Review Service*, Dreier, J. and Estlund, D. eds, World Wide Web,

(http://www.brown.edu/Departments/Philosophy/bears/homepage.html), Posted September 01, 2001.

Tresan, J. (2009). The challenge of communal internalism. *The Journal of Value Inquiry*, 43, 179-199.

Vaidya, A. (2011). The epistemology of modality, *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Zalta, E. (ed.), URL = <http://plato.stanford.edu/archives/win2011/entries/modality-epistemology/>.

Wallace, R. J. (2001). Wallace reviews Svavarsdóttir, *Brown Electronic Article Review Service*, Dreier, J. and Estlund, D. eds., World Wide Web, (http://www.brown.edu/Departments/Philosophy/bears/homepage.html), Posted August 10, 2001.

Williamson, T. (2011). Philosophical expertise and the burden of proof. *Metaphilosophy* 42, 215-229.

Wright, J. C., Grandjean, P. T., & McWhite, C. B. (2013). The meta-ethical grounding of our moral beliefs: evidence for meta-ethical pluralism. *Philosophical Psychology*, 26, 336-361.

Vita

Kenneth Wesley Shields was born in Dallas, Texas in 1982, to his parents Gary and Veronica Shields. He grew up in a suburb outside of Dallas, graduating secondary school from Forney High School in Forney, Texas in 2001. He studied music in high school, specifically the clarinet, saxophone and percussion.

Kenny continued to pursue his musical studies during his undergraduate career at Henderson State University—a small, liberal arts university in Arkadelphia, Arkansas—where he trained to be a secondary school music director. It wasn't until he had taken a summer introductory course in philosophy that Kenny first learned about doing philosophy professionally. While he continued his music education degree, he began sitting in on philosophy classes and regularly attending a weekly philosophy group. After a particularly unsatisfying internship as a secondary school music director, Kenny's mentor encouraged him to pursue philosophy.

Kenny began his graduate career in the liberal arts program at Henderson State University, where he studied the history of philosophy. His MLA thesis examined how philosophers from Plato to the present day explain weakness of will. He graduated with a master's of liberal arts in 2009.

A year after graduating with his master's of liberal arts, Kenny was accepted into the PhD program at the University of Missouri. He graduated with a master's degree in philosophy in 2012. During his six years in this program, he focused primarily on metaethics, moral psychology and experimental philosophy.

Kenny is married to Nicole Shields. They have two children together, Mary Claire and Wesley.