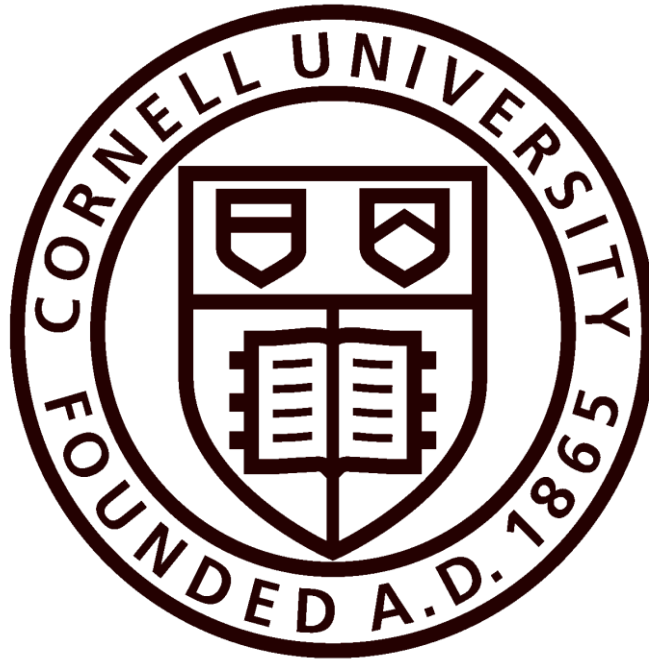


SCHRÖDINGER'S CATEGORIES:
THE INDETERMINACY OF
FOLK METAETHICS



A Dissertation

*Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy*

by

Lance Spencer Bush
May 2023

© 2023 Lance Spencer Bush

SCHRÖDINGER'S CATEGORIES: THE INDETERMINACY OF FOLK METAETHICS

Lance Spencer Bush, Ph.D.

Cornell University 2023

Metaethics is a field of philosophy that addresses fundamental questions about the nature of morality. One of the central disputes in metaethics is whether *moral realism* is true. Moral realism is the claim that there are *stance-independent moral facts*, moral facts that are true independent of the standards or values of individuals or groups, much like scientific facts (e.g., the shape of the earth) aren't made true by personal preference or cultural consensus. *Moral antirealism* is the claim that *there are no stance-independent moral facts*. Research on *folk metaethics* studies whether ordinary people (i.e., *nonphilosophers*) endorse realism or antirealism, or speak and think in ways that commit them to one of these views. Some researchers maintain that nearly everyone endorses either realism or antirealism, but not both. Yet most research suggests significant interpersonal and intrapersonal variation in *folk metaethics*: some people are more inclined towards realism, and others antirealism, while most people are *metaethical pluralists*: they are moral realists about some moral issues and antirealists about others. Regardless of the account in question, *all* existing research presumes that there is a *determinate* fact about whether people are realists or antirealists. I argue that existing evidence does not support this conclusion. Instead, the best account of folk metaethics may be *metaethical indeterminacy*: ordinary people are neither realists nor antirealists, and neither best explains the way people speak or think. The case for *metaethical indeterminacy* proceeds in two steps. First, I argue that *all published studies on folk metaethics rely on invalid measures*. Second, I present empirical evidence that challenges the validity of existing research on folk metaethics and supports metaethical indeterminacy. I evaluate the proportion of people who interpret questions about metaethics as intended, using open response questions, as well as multiple choice questions and Likert scale items. These studies show that most people do not interpret questions

about metaethics as researchers intend. I conclude with a study that demonstrates how forced choice paradigms can create the misleading appearance of a genuine pattern of determinate folk philosophical views, even where none plausibly exist.

BIOGRAPHICAL SKETCH

I have been told that I tend to focus on the negative, and I think that's mostly correct. I'd like to think that this gives me an edge. But it's just who I am. It's not in my nature to provide a positive biographical sketch; the kind that presents an array of highlights and memories like polished trophies. We all know these glossy histories are at best only half-truths. Sure, I've had bright moments. My mom tucking me into bed, folding half the blanket over me and declaring me a "taco." Foggy memories plodding through the swamps of New Jersey with my brother, casting a wary gaze over my shoulder for signs of the Jersey Devil. Winning first place in the pinewood derby. Tearing open a pack of cards from the *Weatherlight* expansion of "Magic: The Gathering" to find a Thundermare staring back at me, eyes ablaze.

Childhood was a time of wonder. But it was also a time of pain, confusion, and loneliness. I'd like to think the person that I became benefited from the hardships, but I fear I'm simply all that remains after so many enervating experiences left me pitted and gouged. I dropped out of high school in tenth grade—a rocky start, but I turned things around. Community college at sixteen. Finished two bachelor's degrees, then completed an MA at Tufts. Hobnobbed with students at Harvard and Oxford. Spent a few months as a research assistant in the Morality Lab at Boston College, and a brief stint teaching philosophy before I managed to get into Cornell. I've had a lot of opportunities, to be sure, and no small amount of luck.

But like I said, I tend to focus on the negative. In August 2019 I developed a chronic medical condition that nearly scuttled any hope I had of finishing my dissertation. A few months later, my mom died after a long battle with cancer—she beat the cancer, but she died anyway. I see pictures of myself from 2018 and I don't recognize the smile there anymore, or the absence of gray hair.

This dissertation is the culmination of over ten years of thinking about metaethics, moral

psychology, and moral philosophy. My aims are largely critical: critical of every study on the psychology of metaethics, of moral psychology, and of philosophy in general. If my central arguments are correct, then an entire literature is fundamentally misguided and thoroughly mistaken. Yet I think it'd be a mistake to see my ambitions as merely destructive, or to think that nothing good can come from an emphasis on criticizing the work of others. When we clear away the mistakes and confusions of the past, wherever possible, we make way for something new, something better. But the pursuit of something better is a task for someone else. Because, after all, I tend to focus on the negative.

To mom

ACKNOWLEDGEMENTS

No self-respecting person would begin their dissertation without an obligatory tipping of the hat to everyone who helped them along the way. As cynical as I am, this is one ritual I wholeheartedly endorse. Since I have many people to acknowledge, I'll begin by listing everyone I mention: Dr. David Pizarro, Dr. Laura Niemi, Dr. Shaun Nichols, Dr. Morten Christiansen, David Moss, Andres Montealegre, Tyler Millhouse, Don Loeb, Michael Gill, Sam Rosen, Brian Tebbitt, Bryce Gessell, Todd K. Shackelford, Sebastian Deri, Rebekah Davis, Thomas Pölzler, Dillon Bowen, Lieuwe Zijlstra, Geoffrey Goodwin, Jennifer Cole Wright, Ross Colebrook, Adam Gibbons, Tyler John, David Newman, Geoffrey Goodwin, and my wife Priscila Serpa.

First, I owe the deepest debt of gratitude to my advisor and the chair of my committee, Dr. David Pizarro. I am grateful for the chance to demonstrate what someone with a fundamentally different approach to psychology can bring to the field and can only hope the gamble pays off. I am also grateful for the approach David took as an advisor and a mentor. I'm an inveterate loner. Stubborn and independent, many of my greatest shortcomings stem from my unwillingness to conform with conventions even when they are there for good reason. David provided the breathing room for me to retain a sense of autonomy, Yet David also knew how and when to step in as it was necessary. There were many times I was afraid to step out of my comfort zone and address the weaknesses in my knowledge and skills. More than merely serving as a mentor, David excelled at creating an atmosphere that acknowledged the pressures of graduate study, but with an ever-present humor that made the whole process tolerable.

I also want to acknowledge the rest of their committee, Dr. Laura Niemi, Dr. Shaun Nichols, Dr. Morten Christiansen, for their insights, thoughts, and contributions. I haven't had the chance to

interact with the rest of the committee as much as I'd have liked, but I am nevertheless grateful for their insights, feedback, and criticisms.

I owe a special thanks to David Moss. David has contributed more than anyone to my research and my development as a thinker. David has been a constant source of insight, moderating my brasher impulses with a more measured approach to research and presenting countless insights and observations that have furthered my knowledge of psychological research and folk philosophy. In fact, although I've refined and implemented the coding scheme employed in chapter 4, it was David Moss who initially proposed this approach. David has also contributed a great deal to the studies reported here, assisting in the formulation of the hypotheses and the measures used, as well as contributing to the many criticisms I level against existing research on folk metaethics. David is in every way a research collaborator and deserves much of the credit for the final product.

I also want to thank my other colleagues and collaborators. Andres Montealegre has been a friend, research colleague, and supporter of the direction of my research for many years now. We have worked on numerous projects, and I have learned more from Andres about how to use R, quality research design, open science practices, the importance of a rigorous approach to survey design, preregistration, and much more than I ever could have imagined.

Thanks, as well, to Tyler Millhouse. Tyler was my roommate when I was a philosophy graduate student at Tufts. Our many discussions about philosophy both deepened my understanding of metaethics and broadened the scope of my knowledge with other areas of philosophy. Tyler also tempered many of my worst philosophical habits and served as a role model for the ability to approach philosophical disagreements with calm, composure, and candor. Tyler has provided feedback on several projects related to folk metaethics and has worked with me on many of my earliest projects, and he continues on occasion to work with me on current projects.

Don Loeb has also earned my gratitude. Don's work has had a profound influence on me, inspiring and motivating me to continue my pursuits with the confidence that a graduate student without the academic credentials of professional philosophers wasn't completely misguided in thinking there was something wrong with the methods and conclusions of contemporary metaethics. Don has also been an open book about his life, sharing many stories and anecdotes. His avuncular demeanor and his ability to befriend virtually everyone he meets is a reminder that philosophers need not be insular and aloof. I have yet to learn how to be as kind and compassionate towards others, and I often fear that it's beyond me. Thanks for being so awesome, Don!

I also want to thank Michael Gill. Dr. Gill's paper on indeterminacy and variability in metaethics could be seen as the key paper motivating my research. This is the only paper I know of that explicitly argued for metaethical indeterminacy. In many ways, this project is a very long footnote to Loeb's incoherentism and Gill's Indeterminacy-variability thesis.

I want to thank Sam Rosen for suggesting the title for the paper. I love bad jokes and terrible wordplay, and *Schrödinger's Categories* scratched that itch perfectly.

Thanks also to Brian Tebbitt, who provided valuable feedback on the content and direction of the dissertation and helped remind me to focus rather than go on a bunch of tangents. Thanks also for the many helpful discussions which helped shape the way I think by refining my views about language and the relationship between language, philosophy, and psychology.

There are far too many other people to thank as well. Thanks to the editors and reviewers at *New Ideas in Psychology* and *Diametros*. A special thanks to Bryce Gessell for helpful discussions and feedback. Thanks to Todd K. Shackelford for inspiring me to pursue psychology in addition to philosophy, for providing so many opportunities to publish articles and to interact with incredible researchers, and most of all for helping me to develop a richer understanding of evolution. If his classes hadn't been so engaging, I may have never moved into my current line of research. Thanks to

Sebastian for extensive help with R and feedback on how to write a dissertation. Thanks to everyone who attended the folk moral relativism workshop and to Thomas Pözlner for organizing it. Thank you to Rebekah Davis for providing critical feedback on the dissertation and for encouragement with my research. Thanks to many others for their helpful discussions and feedback, including Dillon Bowen, Lieuwe Zijlstra, Geoffrey Goodwin, Jennifer Cole Wright, Ross Colebrook, Andrew Maul, Hagop Sarkissian, Adam Gibbons, Tyler John, David Newman. I'd also like to thank Geoff Goodwin for generously sharing data from his research, despite my critical analysis of portions of it. Thank you to everyone else at Cornell who provided support or feedback or simply attended one of my talks. Thanks also to the many people I don't remember or never met who saw one of my posters or presentations at conferences. Thanks also to my parents, who supported me the whole way. Thanks to our wonderful dog Tchutchu, who is sadly no longer with us. She was my sole company for most of the writing process and throughout the solitude of the pandemic. I'd like to think she did the most to motivate me.

Finally, I'd like to thank my wife, Priscila. She has had to put up with all of my flaws and shortcomings the most, yet she's done more than everyone else to keep my head in the game and my fingers typing. Priscila undoubtedly deserves the most credit for seeing this work to completion. Priscila helped with formatting my figures and tables and assisted with numerous aspects of the text, from feedback on the overall organizational structure, deciding which studies to include, providing an outsider's perspective on studies and how I frame my arguments, to virtually every element of writing, formatting, and thinking about the topics I discuss. Yet all the academic feedback and support in the world would be irrelevant if I didn't have the motivation and focus to finish. A lot of terrible things happened in the past few years that held me back and made finishing this project an immense struggle. Priscila has been with me every step of the way, providing encouragement and support, and pushing

me to get organized, to get focused, and to take care of my mental and physical health. I couldn't have done this without her.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH.....	14
ACKNOWLEDGEMENTS.....	17
CHAPTER 1: INTRODUCTION	24
CHAPTER 2: CRITIQUE OF THE DISAGREEMENT PARADIGM	37
2.0 INTRODUCTION	37
2.1 THE DISAGREEMENT PARADIGM	38
2.2 OVERVIEW OF CRITIQUE OF THE DISAGREEMENT PARADIGM.....	42
2.3 INADEQUATE RESPONSE OPTIONS.....	46
2.4 CONFLATIONS WITH UNINTENDED CONCEPTS.....	50
2.5 A SUMMARY OF REMAINING PROBLEMS	62
2.6 CONCLUSION.....	66
CHAPTER 3: CRITIQUE OF ALTERNATIVE PARADIGMS.....	69
3.0 INTRODUCTION	69
3.1 SCALES.....	70
3.2 PROBLEMS OF VALIDITY IN METAETHICS SCALES: AN EXAMPLE.....	71
3.3 TRAINING PARADIGMS	82
3.4 CONCLUSION.....	91
CHAPTER 4: OPEN RESPONSE PARADIGMS.....	92
4.0 INTRODUCTION	92
4.1 RATIONALE FOR USING OPEN RESPONSE DATA	93
4.2 GENERAL METHODS AND PROCEDURES.....	99
4.3 GENERAL PREDICTIONS	104
4.4 SUMMARY OF STUDIES	106
4.5 STUDIES 1A & 1B: SOURCE OF DISAGREEMENT IN GOODWIN & DARLEY (2008)	106
4.6 STUDY 1C.....	117
4.7 STUDY 2: INTERPRETING OTHER PEOPLE’S RESPONSES TO METAETHICS STIMULI.....	117
4.8 STUDY 3: CHARITY & NONCOGNITIVISM.....	127
4.9 STUDY 4: DIRECT REALISM AND RELATIVISM	132
4.10 STUDY 5: INTERPRETATION OF THE MORAL RELATIVISM SCALE (MRS)	143
4.11 STUDY 6: THE NEW METAETHICS QUESTIONNAIRE (NMQ)	149
4.12 STUDY 7: FOLK MORAL OBJECTIVISM SCALE (FMO).....	149
4.13 GENERAL DISCUSSION	150
4.14 CONCLUSION	154
CHAPTER 5: MULTIPLE CHOICE & INDETERMINANCY PARADIGMS.....	156
5.0 INTRODUCTION	156
5.1 MULTIPLE CHOICE PARADIGMS	156
5.2 STUDY 2: INDETERMINACY IN NON-METAETHICAL DOMAINS.....	182
5.3 CONCLUSION.....	214
CHAPTER 6: THIRD PERSON PARADIGM & NORMATIVE ENTANGLEMENT	216
6.0 INTRODUCTION	216
6.1 STUDY 1: THIRD PERSON PARADIGM (PILOT)	229
6.2 STUDY 2: THIRD PERSON PARADIGM WITH CONCRETE MORAL ISSUES.....	243
6.3 STUDY 3: THIRD PERSON PARADIGM (NONMORAL DOMAINS).....	243
6.4 STUDY 4: NORMATIVE ENTANGLEMENT.....	257
6.5 GENERAL DISCUSSION	279
6.6 CONCLUSION.....	289

CHAPTER 7: CONCLUSION	291
REFERENCES	307
APPENDIX A: GLOSSARY	353
APPENDIX B: BIBLIOGRAPHY OF FOLK METAETHICS STUDIES	363
APPENDIX C: METAETHICS SCALES.....	368
APPENDIX D: RESEARCH MATERIALS	375
APPENDIX E: METAETHICS & THE MORAL CONVENTIONAL DISTINCTION	426

CHAPTER 1:

Introduction

Alex and Sam are two typical Americans. They are planning to have a barbeque tomorrow, but according to a local weather forecast, it might rain. If it does rain, they will have to cancel their plans.

This has led to an argument about the weather:

Alex: “*Given the indeterministic nature of the universe, it is ultimately unknowable whether it will rain tomorrow.*”

Sam: “*Nonsense. It will definitely both rain and not rain tomorrow. The universe will diverge down two separate branches. It will rain in one universe, and not in the other.*”

Alex and Sam seem to be endorsing the Copenhagen and Many Worlds interpretations of quantum mechanics, respectively. This disagreement probably seems absurd. It is incredibly unlikely that interpretations of quantum mechanics would arise in everyday disputes about the weather, and even if they did, neither perspective would be helpful in resolving the practical question of whether to cancel a barbeque. The scenario is also absurd because *most people don't have a position on how to interpret quantum mechanics*. Such considerations are irrelevant to everyday decisions, and nothing about the way ordinary people think or speak requires them to take a side in disputes that occupy theoretical physicists.

I believe that a similar absurdity plagues empirical research on *folk philosophy*. *Folk philosophy* refers to the philosophical stances and commitments of ordinary people. *Philosophical stances* are the philosophical beliefs ordinary people hold. *Philosophical commitments* are the philosophical positions implicit in the way ordinary people speak, think, and act, independent of any particular psychological states.¹ *Ordinary people* are people who lack significant formal philosophical training and have not

¹ I will often refer to these as simply *stances* and *commitments*.

engaged in significant philosophical reflection. The goal of empirical research on folk philosophy is to *describe the philosophical stances and commitments of ordinary people*.

Researchers have devised a cunning array of paradigms for cataloging the stances and commitments of ordinary people, spanning every major branch of philosophy from *metaphysics* (Dink & Rips, 2017; Korman & Carmichael, 2017; Nichols & Bruno, 2010; Paul, 2010; Rose, Schaffer, & Tobia, 2018) to *epistemology* (Nagel, San Juan, & Mar, 2013; Starmans & Friedman, 2012; Swain, Alexander, & Weinberg, 2008; Weinberg, Nichols, & Stich, 2001) to *ethics* (Alicke, & Gordon, & Rose, 2013; Greene, 2008; May & Holton, 2012; Phillips, Nyholm, & Laio, 2014) and *aesthetics* (Cova et al., 2015; Cova & Pain, 2012; Rabb et al., 2020).²

Researchers have frequently interpreted these studies as evidence that ordinary people have stances or commitments that correspond to the categories, accounts, and distinctions recognized by philosophers, e.g. *deontology* and *consequentialism* (e.g. Greene, 2008; Kahane et al., 2018; Johansson-Stenman, 2012; May, 2014; cf. Mihailov, 2022), *compatibilism* and *incompatibilism* (e.g., Carstensen, 2022; Nadelhoffer et al., 2020; Nahmias & Murray, 2011; Nahmias et al., 2004; 2005; 2006; 2015; Nichols, 2012; Nichols & Knobe, 2007, cf. Nadelhoffer, Murray, & Murray, 2021), *psychological* and *non-psychological* conceptions of *personal identity* (e.g. Nichols & Bruno, 2010; Shoemaker & Tobia, 2022; Strohminger & Nichols, 2014; 2015; cf. Starmans & Bloom, 2018a; 2018b), adherence to or rejection of the ‘*ought*’ *implies* ‘*can*’ *principle* (Buckwalter & Turri, 2015;

² Some of these studies may only be concerned with measuring philosophical *intuitions* of a kind that do not correspond to what I mean by philosophical *stances* and *commitments*. “Intuition” is a term regularly employed by philosophers, but for which there is unfortunately no single definition. Researchers studying folk *intuitions* may or may not be studying stances or commitments, e.g., they might be studying *dispositions* to believe certain propositions, rather than studying what people pretheoretically believe are committed to (Earlenbaugh & Molyneux, 2009). Since I am offering a stipulative set of terminological distinctions other authors have not used, I would be unable to confirm whether they conceive of their findings as indicators of stances and commitments without consulting each author, explaining the distinction, and asking them for their position. Failing that, I could selectively opt for studies that appear to fit my distinctions, this task would be largely guesswork on my part. Instead, I chose broadly representative research characterized by scope and impact. I believe these studies provide a more useful picture of the kind of research I am referring to than a curated list of more obscure studies that would risk presenting a skewed notion of what folk philosophical research is about, even if some of these studies are not subject to my criticisms.

Chituc et al., 2016; Cohen, 2018; Henne et al., 2016; 2019; Kissinger-Knox, Aragon, & Mizrahi, 2018; Mizrahi, 2015; Semler & Henne, 2019; cf. Thompson, 2022) *causal-historical* and *descriptivist* theories of reference (e.g. Machery et al., 2004, Mallon et al., 2009; van Dongen et al., 2021), and so on.

Yet few researchers have seriously considered the possibility that ordinary thought and language is *indeterminate* with respect to at least some of these philosophical distinctions.³ In other words, there may be philosophical distinctions that are absent from the way ordinary people speak and think, and that, for these philosophical issues, *ordinary people don't have any particular philosophical stances or commitments at all*. If so, then there may be no way to resolve competing analyses of the philosophical content of folk philosophy, since these analyses could equally accommodate (or fail to accommodate) the data (Gill, 2009). If this is the case, the implications would be catastrophic for much ongoing research, since this research would be attempting to describe features of ordinary thought and language that don't exist.

I am not claiming that ordinary people have *no* philosophical stances or commitments. Even the most skeptical account of folk philosophy would acknowledge that ordinary people hold some minimally construed stances and commitments. To the extent that people endorse first-order moral judgments (e.g., “torture is morally wrong”) they could be said to have a philosophical stance. And insofar as ordinary people speak as though some beliefs are more justified than others, that there is an external world, or that we have a *prima facie* duty to keep our promises, ordinary people may be said to have epistemic, metaphysical, and moral commitments, respectively. I am not challenging the existence of determinate stances and commitments of this kind. Yet these are mundane and uncontroversial beliefs that are rarely the focus of research on folk philosophy.

³ Gill (2009) and Pravato (2020) are notable exceptions. Gill explicitly argues for the possibility of indeterminacy in folk metaethics (for replies see Johansson & Olson, 2015; Sinnott-Armstrong, 2009). Pravato (2020) argues that normative language (e.g., “good,” “ought”) is indeterminate, though it is not meaningless.

Research on folk philosophy instead tends to focus on how ordinary people think about ongoing disputes central to academic philosophy, often with the goal of illuminating or resolving these disputes (Alexander, Mallon, & Weinberg, 2010; Knobe, 2007; Sytsma & Livengood, 2012). As a result, most research on folk philosophy offers insight into ordinary thought and language that is nonobvious or at least subject to reasonable doubt. For instance, researchers studying folk philosophy could plausibly wonder whether ordinary people are committed to the notion that ‘*ought*’ implies ‘*can*’. Yet researchers studying folk philosophy would have little interest in studying whether people think it’s morally wrong to torture babies for fun. This is not to say that researchers wouldn’t ask this sort of question, but simply that responses to it would not be interpreted as discoveries about folk philosophy, and they wouldn’t play an important role in resolving philosophical disputes. Such measures would instead serve some conventional psychological purpose, such as measuring psychopathy.

At the other extreme, it seems unlikely that ordinary people hold philosophical stances about more esoteric philosophical issues, or speak and think in ways that exclusively fit one or another side of obscure philosophical disputes. For example, Gill (2009) finds it implausible that ordinary people speak or think in ways that commit them to mathematical Platonism or anti-Platonism when they use mathematical language, since he believes “[t]he way people use numbers in everyday math simply does not contain answers to the questions that animate philosophy of mathematics” (p. 218). Likewise, it is unlikely that the way people speak or think commits them to a particular stance about the legitimacy of the analytic-synthetic distinction, the optimal decision-theoretic solution to Newcomb’s problem, or other debates that concern academic philosophers, and it is even less plausible that ordinary people have explicit stances on these issues.⁴

Thus, there is a continuum between folk stances and commitments that uncontroversially exist and those that are likely absent from folk philosophy altogether. Somewhere between these extremes

⁴ For other examples of popular philosophical disputes, see Bourget and Chalmers (2014; ms).

lies a nebulous middle ground of philosophical accounts, concepts, and distinctions that possess a less-than-certain status as features of folk philosophy. This middle ground is the uncertain territory where I intend to build my case for folk indeterminacy. In making my case, I am not claiming that there are *no* determinate features of folk philosophy between the extremes. There are many features of the way ordinary people think and speak that have been or could be discovered, and at least some of these findings may correspond to traditional philosophical distinctions (e.g., perhaps most ordinary people really are *compatibilists* or *incompatibilists* about “free will”). Yet each hypothesis must be evaluated on an individual basis. We should not presume that a given distinction that is important to philosophers is a part of folk philosophy. In the end, empirical evidence will be the final arbiter of the content of folk philosophy.

While my concerns apply to folk philosophy in general, I will focus exclusively on *folk metaethics*. Metaethics is a branch of philosophy that deals with abstract questions about the nature of morality, such as whether there are moral facts, what makes moral facts true, and how we might acquire knowledge of moral facts (Sayre-McCord, 2014). *Folk metaethics* is simply the subset of *folk philosophy* dedicated to studying the metaethical stances and commitments of ordinary people. While there are several questions that fall under the purview of metaethics, most research focuses on whether ordinary people endorse or are committed to some form of *realism* or *antirealism* about moral facts (Pölzler &

Wright, 2020a; 2020b). Roughly speaking, this distinction concerns whether there are *stance-independent*⁵ facts about what is morally right or wrong (Shafer-Landau, 2003, p. 15).⁶

Realism: *There are stance-independent facts about what is morally right or wrong*

Antirealism: *There are no stance-independent facts about what is morally right or wrong*⁷

Most studies have found that the majority of ordinary people are *metaethical pluralists* who endorse realism about some moral issues and antirealism about others (Beebe, 2014; Beebe et al., 2015; Beebe & Sackris, 2016; Goodwin & Darley, 2008; 2012; Pölzler & Wright, 2019; 2020a; 2020b; Wright, 2015; 2018; Wright, Grandjean, & McWhite, 2013, Zijlstra, 2019; cf. Beebe, 2020). For instance, a person may judge that there is a stance-independent moral fact about whether murder is morally wrong, but that there is no stance-independent moral fact about whether abortion is morally wrong. They might instead think that claims about abortion can only be true or false relative to the moral standards of different individuals or cultures, or that such claims merely expression nonpropositional content, e.g., a negative emotion or an imperative to not get an abortion.

⁵ I follow Shafer-Landau's lead by using the term stance-independent rather than the more common term mind-independent. This is because, as Shafer-Landau notes, the latter is more disposed to prompt confusion since there is one respect in which realism entails that moral facts are mind-independent: they are not made true by the beliefs or values of people. Yet there is another respect in which some moral facts may crucially depend on people's attitudes or values. People sometimes mistakenly think that mind-independence means that whether an action is right or wrong depends on the psychological impact that action would have on an individual. For instance, facts about whether it would be okay to hit someone would depend on whether that person would suffer. And since suffering is a mental state, we might think of claims like "hitting someone is wrong" as mind-dependent, since the reason hitting someone may be wrong is because it causes suffering. Yet this is simply a different sense in which a moral fact could depend on mental states than is intended by the notion of mind-independence or stance-independence.

⁶ Realism is sometimes defined more minimally, and instead consists of two claims: that moral sentences are truth-apt and at least some of them are true (Sayre-McCord, 2015). There seems to have been a shift in recent decades towards reserving the term realism for more robust forms of realism that include stance-independence or other considerations.

⁷ Technical terms like "stance-independent" are likely to prove unhelpful. Roughly, the distinction concerns whether there are moral facts that are not made true by our goals, standards, or values, or whether there are no such facts. Comparison with more familiar concepts might help. Many of us, on reflection, may agree that scientific facts don't depend on our goals, standards, or values. Believing, or really wanting the earth to be flat couldn't make it true that it was flat. Yet many of us would, if we reflected on the matter, deny that there are any stance-independent facts about which food or music is best. We might instead insist that it's a matter of personal preference, and that facts about which food tastes good or bad can only be understood relative to different people's stance towards the food (e.g., whether it tastes good or bad to them), if there are any facts at all. If so, we might say that we're gastronomic antirealists: we deny that there are facts about which foods taste good or bad that are stance-independently true. See Loeb (2003) for an amusing discussion of gastronomic realism and its relation to moral realism.

Some researchers have accepted these findings at face value and argued that *folk metaethical pluralism* is the best interpretation of the data (Davis, 2021; Feltz & Cokely, 2013; Hopster, 2019; Pölzler, 2017; Wright, Grandjean, & McWhite, 2013). Others have interpreted findings in folk metaethics to support the claim that most people are realists (Goodwin & Darley, 2008; 2012) or antirealists (Beebe, 2020; Sarkissian et al., 2011; Pölzler & Wright, 2020b) or at least that we can rule out specific folk metaethical views such as folk noncognitivism (Pölzler & Wright, 2020a). All of these interpretations share the presumption that folk metaethics is *determinate* (Gill, 2009).⁸ In other words, they presume that we can decisively demonstrate that ordinary people adopt a realist or antirealist stance towards moral claims, or speak in ways that best fit some form of realism or antirealism.

I contend that all of this research is fundamentally flawed, and that these interpretations are all mistaken. I don't mean that the studies are poorly designed, and that, with a few tweaks, we'll be able to properly assess what the folk think about realism and antirealism. Nor do I mean that a proper interpretation of these studies would allow us to determine which stances or commitments ordinary people have regarding realism and antirealism. I mean that such efforts cannot succeed in principle because, with few exceptions, *ordinary people have no determinate stances or commitments about the truth status of moral claims*. With respect to realism and antirealism, most ordinary people exist in a state of philosophical superposition that collapses only by engaging in philosophy. I will refer to this as the *metaethical indeterminacy thesis*, which is the claim that *ordinary people have no determinate metaethical stances or commitments with respect to moral realism and antirealism*.

The primary purpose of this dissertation is to explain why I think this is the case, and to convince you that *metaethical indeterminacy* is a plausible account of folk metaethics. Since all existing research on folk metaethics has been interpreted as evidence that ordinary people have determinate

⁸ One exception to this is Pölzler and Wright (2019; 2020a). They don't presume determinacy so much as explicitly argue that their findings cast doubt on indeterminacy. I am indebted to them for clearly laying out their reservations and doubts, and for offering such high quality and insightful studies for our disagreements to center around.

metaethical stances and commitments, this leaves me with three central tasks: (1) to demonstrate that the measures used in existing research on folk metaethics are invalid, (2) to demonstrate this can be readily explained by metaethical indeterminacy and (3) to provide empirical evidence that supports metaethical indeterminacy.

My third objective is intentionally modest. I do not intend to provide anything even approaching decisive evidence of metaethical indeterminacy. This would be an incredibly difficult task, since there are few straightforward ways to convincingly demonstrate that ordinary people have no determinate metaethical stances or commitments. Researchers could always insist that they do, but that existing methods have simply failed to reveal them. Such a task, if it is to succeed, will call for multiple, converging lines of evidence and a broader theoretical foundation than is feasible for the scope of this project. My efforts here should be seen, instead, as laying the groundwork for such a project, providing both a theoretical rationale for the plausibility of metaethical indeterminacy and preliminary empirical support.

It should also quickly become apparent *why* demonstrating metaethical indeterminacy will prove difficult. Although the folk metaethics literature is new and of modest size, no researchers have interpreted their findings as evidence of metaethical indeterminacy. Quite the contrary, all existing research supports a determinate account of folk metaethics, with the only question being what the distribution of different folk metaethical positions is (i.e., whether most people are some form of realist, antirealist, or a combination of the two). Overturning an entire literature is no easy feat. Nevertheless, I *do* believe I can decisively succeed at this task. The methodological critiques that I level against existing studies on folk metaethics reveal far more than the mere difficulty of measuring folk metaethics, they raise serious doubts about whether there is any viable way to measure folk metaethics *at all*. Even if the case for metaethical indeterminacy cannot be decisively established, but only gestured at, I nevertheless hope to show that existing studies fail to provide good evidence that people have

determinate metaethical stances and commitments. In other words, before demonstrating that I am right, I first set out to demonstrate that everyone else is wrong.

If I am correct that ordinary people have no determinate metaethical stances or commitments, the implications would be considerable. Whole research lines in metaethics will prove little more than misguided attempts to measure metaethical stances that most people simply don't have, or to measure commitments that are almost entirely absent from folk philosophy, and, where present, do not vindicate traditional metaethical accounts or existing interpretations of the data. This troubling possibility has been obscured by flawed research design, poor measurement, and a host of methodological problems (Beebe, 2015; Bush & Moss, 2020; Pölzler, 2018b; 2018c). Taken together, these issues give the superficial appearance of nascent fields of research that simply need to refine their tools before progress can be made.

In the chapters that follow, I will address these methodological shortcomings, and argue that researchers have not defended the most plausible interpretation of available data. Yet in doing so, I do not wish to give the impression that the solution is to devise better measures. Rather, I hope to show that the observed patterns of results, as well as the underspecified questions, ubiquitous confounds, pervasive ambiguity in scale items, and other methodological shortcomings are not merely the result of flawed research design, but an unavoidable byproduct of the mismatch between philosophical theories and ordinary thought and language (Bush & Moss, 2020). As I will show, most participants do not understand questions about metaethics in the ways researchers intend, and I will argue that a plausible explanation for this is not that researchers have yet to ask questions in a way participants interpret as intended, but that the types of responses researchers are trying to elicit simply aren't features of folk thought and language to begin with.

I don't believe the problem of folk indeterminacy is a problem exclusive to metaethics. There may be broad ramifications of the kind of analyses I marshal to support this conclusion. If an entire

field of research could present the superficial appearance of legitimate psychological phenomena where none exist, or at least exist in a form radically different from what researchers propose, could the same be true of other lines of research as well? Much to my alarm, I fear an answer in the affirmative.

Given these potentially broader implications, it would be reasonable to ask why I focus on metaethics in particular. Why choose an obscure field with a comparatively small body of literature? The main reason is simply that I'm interested in the topic. But focusing on folk metaethics is also strategic. I am more familiar with research on folk metaethics than any other area of study, and metaethics is my primary area of interest in philosophy. Most importantly, it is an area where the case for indeterminacy is especially strong. Finally, I also suspect that I can build a more persuasive case by focusing on a single topic than by spreading myself too thin by addressing many different areas of research.

If I am successful, this will open the door to future efforts to build similar cases for indeterminacy in other areas of folk philosophy. I am currently developing a similar case for indeterminacy about folk notions of free will and will turn my attention to other folk philosophical distinctions in the future.⁹ But for now, folk metaethics is a rich, active, and evolving field of research that provides fertile ground for developing a template for similar arguments in favor of indeterminacy in other areas of folk philosophy. My critique of research on folk metaethics is thus intended also as a proof of concept for the broader possibility of folk indeterminacy with respect to other philosophical distinctions, and that other areas (or potential areas) of research on folk philosophy may be dead ends.

⁹ I could have presented a similar case for indeterminacy with respect to other areas of folk philosophical research, such as *folk epistemology* (Beebe, 2012; Kim & Yuan, 2015; Machery et al., 2004; Nichols, Stich, & Weinberg, 2003; Weinberg, Nichols, & Stich, 2001) or *personal identity* (De Freitas et al., 2017; Molouki & Bartels, 2017; Shoemaker & Tobia, 2019; Starmans & Bloom, 2018a; 2018b; Strohminger & Nichols, 2014; 2015), among other possibilities. Hopefully researchers who are more familiar with these bodies of literature will investigate the possibility of indeterminacy in these areas of research as well. Even if there are determinate folk views with respect to these philosophical literatures, my approach may aid in refining the measurement tools used to assess other topics.

As bleak as this may seem, my hope is that coming to terms with the intractability of some research programs will redirect efforts towards new lines of research. After all, beliefs and attitudes related to many aspects of morality—including agency, responsibility, and normativity—are deeply embedded in our languages and cultures, and are thus plausibly part of our evolved psychology (Cosmides & Tooby, 2016; Devaine, Hollard, & Daunizeau, 2014; Mallon and Machery, 2010, Zahn-Waxler & Robinson, 1995). There must be patterns in the way people speak and think that could shed light on these and other topics that traditionally fall within the ambit of moral philosophy. The same is true of folk epistemology and metaphysics. Yet our understanding of how ordinary people think about these issues has been hindered by an overreliance on the traditional concepts, distinctions, and framing of questions that dominates analytic philosophy and Western philosophical traditions generally. We should draw insights from academic philosophy whenever possible, but I hope to show that attempts to force ordinary thought into the artificial strictures that animate historical philosophical disputes can distort and obscure the factors that *do* characterize folk philosophy.

On a personal note, these criticisms are not motivated by hostility towards research on folk philosophy or folk metaethics in particular. After years of frustrating attempts to study folk philosophy, I have grudgingly accepted that a great deal of projects are dead ends. Rather, I am motivated by the desire for philosophers and psychologists working at the intersection of our fields to conduct the highest quality research possible. I suspect that others share my feeling of never being completely at home in a philosophy or psychology department. But if we intend to craft an interdisciplinary home for ourselves, I want to ensure it is one we are proud to live in. Much as criticism of the rigor and methods of social psychology can be motivated by a love for the discipline and optimism about its future, I see my efforts not as a call to abandon the study of folk philosophy, but as an effort to rectify significant methodological shortcomings in existing research and, more importantly, to call on philosophers to show more appreciation for the potentially rich and surprising

ways in which folk philosophical thought might differ from academic philosophy. I have no doubt that I will err many times here and elsewhere, and that if my work receives any attention at all, much of it will be to correct such mistakes. I also have no illusions that the case presented here will convince everyone or stem the tide of misguided research on metaethics or folk philosophy in general, but at the very least, I hope my criticisms will contribute to the improvement of future research.

The chapter structure is as follows. In chapter two, I critique *the disagreement paradigm*, the most common method used to measure folk realism and antirealism. In chapter three, I critique the most prominent alternative measures to the disagreement paradigm and a few attempts to experimentally manipulate folk realism and antirealism. In both chapters, my goal is to show that existing methods suffer such substantial methodological shortcomings that they are unlikely to provide valid measures of folk metaethics. In chapter four, I present a series of studies that center on the collection and evaluation of open response data. My goal in this chapter is to demonstrate that ordinary people struggle to interpret questions about metaethics as intended. Chapter five is split into two sections. In the first section, I supplement the studies conducted in the previous chapter with a series of multiple choice studies and a study using Likert scale items that also assess interpretation rates for metaethics stimuli. Once again, I show that ordinary people struggle to interpret metaethics stimuli as intended. Finally, I present a study which asks participants to assess a highly sophisticated theoretical issue: different interpretations of quantum mechanics. I argue that it is highly implausible that ordinary people would have sufficient prior knowledge of quantum mechanics to have a meaningful philosophical stance, and that it is even less plausible that ordinary language is implicitly committed to distinct philosophical positions on quantum mechanics. Nevertheless, nonrandom patterns of data emerge when using a forced choice paradigm. This illustrates that even when we possess strong theoretical grounds for presuming folk *indeterminacy*, studies can create the superficial appearance of

folk *determinacy*. Finally, in the conclusion, I offer a brief commentary on the implications of my arguments and findings, and an optimistic prognosis for the future of research into folk philosophy.

I am also aware of the interdisciplinary nature of this project. I employ jargon in both philosophy and psychology, and introduce new terminology. I have put together an extended explanation of some of the major terms, themes, concepts, and arguments introduced here and throughout the dissertation. I also include caveats and qualifications to the central argument which were too lengthy to include in the main text. These are available in **Supplement 1**. For a brief summary of the relevant terms used in this chapter and subsequent chapters, please consult the glossary in **Appendix A**.

CHAPTER 2: Critique of the Disagreement Paradigm

2.0 Introduction

Most research on whether ordinary people are moral realists or antirealists employs the *disagreement paradigm* (Bush & Moss, 2020). Researchers have also employed a variety of other measures, including scales (e.g., Collier-Spruel et al., 2019; Zijlstra, 2019), implicit measures (Wagner, Pölzler, & Wright, 2021; Zijlstra, 2021), fMRI (Theriault et al., 2017; 2020) and training paradigms (Pölzler & Wright, 2020a; 2020b; Wright, 2018). This chapter provides a summary of the many methodological reasons to doubt the validity of the disagreement paradigm.¹⁰ I explore problems with other measures in the following chapter. Both chapters provide a cursory overview of the many methodological shortcomings with existing folk metaethics paradigms. Although their methods differ, most of these studies provide *prima facie* evidence of moral pluralism, though some researchers have argued that their findings, or folk metaethical findings, support folk uniformity (e.g., Zijlstra, 2021).¹¹

Unfortunately, *none of these paradigms provide valid measures of folk metaethical stances or commitments*. Each study exhibits a host of methodological shortcomings that render its findings at best inconclusive. Of course, I cannot review every study. Instead, I have organized studies by the methods used, and chose examples that are most representative of those paradigms. There are no established conventions or standardized paradigms for folk metaethical research, so my organizational scheme will not perfectly distinguish all study designs, and some studies will exhibit characteristics that overlap with multiple categories or that will not comfortably fit into any particular category. In some cases,

¹⁰ Although my critiques are more comprehensive than others, I am not the first to raise these or similar objections. Pölzler (2018a; 2018b) has previously raised some of the same criticisms of research on folk metaethics, and has raised some concerns that I don't focus on here. Although I've consolidated many critiques here, Pölzler's criticisms are excellent and worth considering as well.

¹¹ This includes uniform folk relativism (Beebe, 2020), a nearly uniform tendency towards antirealism, but pluralism about the type of realism (i.e., most people endorse cultural relativism, others individual subjectivism, and others noncognitivism; Pölzler & Wright, 2020a; 2020b; Wright 2018, Pölzler and Wright 2019).

specific studies exhibit virtues or shortcomings that are not present in the examples I use to represent that paradigm, or employ additional measures that I did not deem sufficiently important to consider. Where relevant, I will make note of the relevant idiosyncrasies of individual studies. Yet there are so many methodological problems with this research, both in general and with respect to individual studies, that there is no practical way to adequately present them in the main text. I provide a more comprehensive critique of the disagreement paradigm in **Supplement 2** and a more comprehensive critique of alternative measures in **Supplement 3**.

Philosophers may also question the value of research on folk metaethics. After all, aren't we more concerned with whether moral realism or antirealism are true than what people think or how they speak? Aside from the fact that the answer for some of us will simply be "no," there are still important questions about the relationship between academic metaethics and folk metaethics. I address this in **Supplement 1** where I argue that a great deal of academic metaethics *is* concerned with folk metaethics and that many disputes in metaethics turn in part on empirical questions about how ordinary people speak and think. Some researchers may also wonder whether research on the Moral/Conventional Distinction (the MCD) provides evidence of folk realism. I argue that it does not in **Appendix E**.¹²

2.1 The disagreement paradigm

There is no single, canonical form of the disagreement paradigm, but they all share a similar structure:

- (1) Typically, participants are first asked to rate how much they agree or disagree with a moral claim, e.g., *"Robbing a bank in order to pay for an expensive vacation is morally bad."*

¹² I must apologize for the inclusion of so many supplements and appendices. I have done so to spare the reader an interminable slog through the methodological musings that motivated the empirical portions of this project, which appear in chapters four and five. However, it would feel like a betrayal of my principles to simply omit such discussions from my work. One of the central objectives of this project is to criticize the tendency for researchers to dive straight into experimental research without adequately situating their research on solid theoretical foundations reinforced by substantive descriptive findings. It would therefore make little sense to conceal the theoretical contributions I can offer when they can be made available to those interested in advancing the study of folk metaethics.

(2) Next, they are presented with a disagreement about the moral claim. This disagreement can be between themselves and another person, or between two other people.

(3) Finally, they must select a response that indicates their metaethical position. Usually, this involves a set of multiple choice response options that reflect different realist and antirealist positions.¹³

Precise wording may vary, but all versions of the disagreement paradigm attempt to distinguish realism from antirealism by asking whether (i) *both people can be correct* (which indicates *antirealism*), or if (ii) *at least one person must be incorrect* (which indicates *realism*). For example, Beebe and Sackris (2016) present the disagreement paradigm as follows:

If someone disagrees with you about whether [claim], is it possible for both of you to be correct or must one of you be mistaken?

☐ *At least one of you must be mistaken*

☐ *It is possible for both of you to be correct*^{14,15}

The judgment that *at least one person must be mistaken* is interpreted as *realism* because it seems to reflect the view that there is a single stance-independent standard of truth. Conversely, the judgment that both people could be correct is interpreted as an expression of *antirealism* because, if a person that judges that a moral claim can be *correct* according to one person's moral standards but *incorrect* according to another person's standards, they seem to believe the truth of moral claims can only be judged relative to each person's moral standards and can vary depending on what that person believes. This judgment reflects a form of *relativism*, which for our purposes is a form of *antirealism*.¹⁶

¹³ Studies sometimes differ. Sarkissian et al. (2011) present a realist statement then ask participants how much they agree or disagree with it on a 7-point Likert scale.

¹⁴ I changed the order of the response options because it seems more natural to describe realism first, then antirealism.

¹⁵ I chose Beebe and Sackris's version of the disagreement paradigm for its clarity and simplicity.

¹⁶ To illustrate the rationale behind the paradigm, consider how we judge disagreements about nonmoral issues. Some disagreements involve mutually incompatible truth claims. If Alex claims that "Water is H₂O," and Sam claims that "Water is not H₂O," we know they cannot both be correct, because these claims represent jointly exhaustive and mutually incompatible descriptions of reality. But if Alex claims that "Cheese is delicious," and Sam claims that "Cheese is not delicious" it is not obvious that one of them must be incorrect. Perhaps these statements are best understood not as competing truth claims, but as indexicalized expressions of the subjective preferences of the individual making the claim. That is, Alex could mean "I find cheese delicious," while Sam could mean "I don't find cheese delicious." If so, Alex and Sam would each be making an assertion that is true *relative to their own preferences*. These statements may *appear* to conflict with one another, but this is simply because the indexical element in each statement is implicit. Once it is made explicit,

We could ask participants whether people who disagree about morality (in general) can both be correct or if at least one must be incorrect, but most studies present participants with a series of concrete moral claims (e.g., murder, abortion, donating to charity, etc.) and repeat the same set of questions for each moral issue. This allows researchers to assess variations in metaethical judgments towards distinct moral issues. For instance, a person may provide a realist response for the claim “murder is wrong,” but an antirealist response for the claim that “abortion is wrong.” Most researchers also include questions about nonmoral issues, such as claims about science or history, social conventions, and matters of taste or aesthetic judgment. This allows for cross-domain comparisons. For instance, researchers may hypothesize that most people will consistently choose the *realist* response for disagreements about science, but *antirealist* responses for social conventions and matters of taste. Researchers then assess the total proportion of *realist* and *antirealist* responses participants provided for each moral and nonmoral claim. This allows them to assess the proportion of realist and antirealist responses for each moral and nonmoral issue individually and to assess the overall proportion of realist and antirealist responses within each domain.

Most studies find that the majority of participants endorse realism for some moral issues and antirealism for others. In fact, *all studies that employ the disagreement paradigm find at least some evidence of intrapersonal variation*. While some participants do provide uniformly realist or antirealist responses, they are typically in the minority (but see Pölzler & Wright, 2020a; 2020b for notable exceptions). In other words, most studies suggest not only that pluralism is a common response, but that *most* people appear to be metaethical pluralists. These studies also reveal consistent evidence of *interpersonal variation*, since

any apparent conflict dissolves. We are already familiar with indexical claims, and have no trouble understanding them in other contexts or when they are made explicit. Obviously, there would be no conflict if Alex were to say, “I am Alex,” and Sam were to say, “I am not Alex.” Both statements would be true and obviously don’t conflict with one another.

some participants exhibit a consistent inclination for realism, while others are more disposed towards antirealism.¹⁷

Many researchers have accepted these findings at face value, i.e., people appear to be metaethical pluralists because they *are* metaethical pluralists (e.g., Colebrook, 2020; Pölzler & Wright, 2018; 2020a; 2020b; Wright, Grandjean, & McWhite, 2013; Wright, 2018). However, others have characterized the overall evidence as favoring a uniform account, or at least a substantially uniform account even if they acknowledge a bit of pluralism (e.g., Beebe, 2020; Goodwin & Darley, 2008; Sarkissian et al., 2011; Zijlstra 2021), while still others have expressed skepticism about the plausibility of pluralism on theoretical grounds (Johansson & Olson, 2015; Sinnott-Armstrong, 2009). Finally, one could favor metaethical indeterminacy by arguing that existing studies are invalid and contend that they show neither pluralism nor uniformism because ordinary people don't have determinate metaethical stances or commitments. This leaves us with three competing hypotheses about folk metaethics: *uniformism*, *pluralism*, and *indeterminism*.

(i) **Uniformism** holds that there is a single determinate account of folk metaethics with respect to a particular metaethical distinction.

(ii) **Pluralism** holds that there is ineliminable but determinate variability with respect to one or more traditionally competing accounts of folk metaethics.

(iii) **Indeterminism** holds that there is no determinate position that characterizes some or all folk metaethics with respect to one or more metaethical distinctions.¹⁸

¹⁷ Since every study finds evidence of metaethical variability, this might appear to be overwhelming evidence of *pluralism*. If *uniformism* were correct, we would expect little interpersonal or intrapersonal variation. This doesn't mean we should expect *perfect* uniformity. People can be subject to performance errors, or misunderstand questions, or lack competence with regard to the relevant terms or concepts, or even deliberately select responses that do not reflect their beliefs out of spite or amusement. The uniformist is not, therefore, committed to the position that *all* respondents should *always* choose the same response in practice. They need only capture a consistent and strong majority. Yet even this low bar cannot be met. Instead, we find most major metaethical positions well-represented among participants. If these findings accurately reflect the metaethical stances or commitments that participants genuinely hold, this would be straightforward evidence of metaethical pluralism, at least among the populations these samples represent (i.e., mostly WEIRD populations; see **Supplement 2**).

¹⁸ Each of the hypotheses may also apply to one dispute but not another, and are thus not incompatible in that respect, either. It may be that folk metaethics is determinately cognitivist, and that prototypical moral claims are thus best understood as propositional claims (i.e. *uniformism* towards cognitivism), while there could also be no determinate answer

At present, all published interpretations of folk metaethical data support either uniformism or pluralism. Though Gill (2008; 2009) was the first to defend metaethical indeterminacy, this proposal was coupled with pluralism and preceded the bulk of empirical research on folk metaethics, though it nevertheless served as an impetus for this project. Thus, to my knowledge, no one has argued that indeterminism (with a qualified splash of pluralism) may offer a plausible explanation of current empirical findings and the position most likely to be vindicated by future research.^{19,20} In the remainder of this chapter, I review the state of the empirical research on folk metaethics and argue that existing data is too methodologically flawed to support uniformism or pluralism.²¹

2.2 Overview of critique of the disagreement paradigm

There are many reasons to doubt the disagreement paradigm's ability to reliably distinguish realists from antirealists. Documenting all these shortcomings may seem like overkill. But I do so to make a point: measuring folk metaethics is *difficult*. By reviewing the many pitfalls that researchers face when studying folk metaethics, I intend not merely to illustrate that the disagreement paradigm should be discarded, but to gesture towards the more pessimistic conclusion that surveys are not up to the task

as to whether these assertions are relative or not (i.e., *indeterminism* towards relativism vs. nonrelativism), or there could be instances of relativism and nonrelativism that cannot be explained away as conceptually confused or nonstandard (i.e., *pluralism* about relativism vs. nonrelativism, Gill, 2008, p. 218).

¹⁹ However, Gill (2009) does mention early empirical evidence that supports metaethical pluralism.

²⁰ According to Gill (2009), it is:

“...quite likely that meta-ethical indeterminacy characterizes much of ordinary discourse. But it may not characterize all of it. The best descriptive analysis of some other uses of moral terms might involve robust meta-ethical commitments. Those commitments, however, might not all be uniformly consistent with one side of the traditional meta-ethical debate over the other.” (p. 218)

I agree. But where Gill and I differ is that I emphasize indeterminacy to a much greater extent.

²¹ Pluralism and indeterminism are not mutually incompatible but may instead characterize subsets of folk thought and language. I discuss this in **Supplement 1**. Note that the compatibility of pluralism and indeterminacy could take multiple forms. It could be that some people's moral thought and language is consistently determinate but other people's moral thought and language is consistently indeterminate, or it could be that moral thought and language tend to be determinate in particular contexts but not others, or towards particular moral issues but not others, or among the members of some communities but not others, or among adults but not children, etc. There are too many possibilities for me to entertain all of them.

of evaluating folk metaethics.²² Pölzler (2018b; 2018c) has done much of the work in documenting these shortcomings, and several of the concerns I raise summarize or are similar to points he has made. However, I raise a variety of novel concerns and add my perspective to criticisms originally raised by Pölzler.

All criticisms of the disagreement paradigm are summarized in **Table 2.1**. Including each subcategory of a particular criticism, there are at least 28 methodological criticisms. It would be impossible to discuss each in detail. I will discuss a handful in detail and provide a summary of the rest. If any of these criticisms seem obscure, confusing, implausible, or underdescribed, see **Supplement 2** for a more complete discussion.

Table 2.1

Summary of methodological problems with the disagree paradigm

Problem		Description
1	<i>Inadequate response options</i>	Missing options: Typical response options are not mutually exhaustive and often fail to include standard forms of realism and antirealism
1.1	<i>Missing noncognitivism</i>	Studies typically exclude a response option for noncognitivism
1.2	<i>Missing error theory</i>	Studies typically exclude a response option for error theory
1.3	<i>Missing subcategories of relativism</i>	Studies typically fail to distinguish different forms of relativism (e.g., subjectivism, cultural relativism, appraiser relativism, agent relativism)
1.4	<i>Missing other subcategories</i>	Studies typically fail to distinguish different forms of realism, including naturalism, non-naturalism, constructivism, relation-designating accounts, incoherentism, hybrid theories, indeterminacy, pluralism, and quietism
2	<i>Poor specificity</i>	Typical response options have poor specificity: they cannot distinguish different metaethical positions from one another
3	<i>Conflations</i>	Stimuli is often interpreted in unintended ways due to conflating realism and antirealism with unintended concepts and distinctions
3.1	<i>Conflating metaethics with normative ethics</i>	Participants often conflate metaethics stimuli with <i>normative</i> moral concerns. This undermines the validity of the disagreement paradigm since many

²² Pölzler (2018) was the first to document many of the methodological shortcomings discussed in this section. I am grateful for his clear and careful documentation of many of these shortcomings.

		responses express the participant's normative moral views, <i>not</i> their views on realism or antirealism
3.2	<i>Conflating metaethics with epistemic concerns</i>	Participants often conflate metaethics stimuli with <i>epistemic</i> concerns. This undermines the validity of the disagreement paradigm since many responses express the epistemic judgments, <i>not</i> their views on realism or antirealism
3.3	<i>Conflating realism with universalism</i>	Participants often conflate realism with <i>universalism</i> . This undermines the validity of the disagreement paradigm since many responses express judgments about universalism, <i>not</i> their views on realism or antirealism
3.4	<i>Conflating realism with absolutism</i>	Participants often conflate realism with <i>absolutism</i> . This undermines the validity of the disagreement paradigm since many responses express judgments about absolutism, <i>not</i> their views on realism or antirealism
3.5	<i>Conflating relativism with contextualism</i>	Participants often conflate relativism with <i>contextualism</i> . This undermines the validity of the disagreement paradigm since many responses express judgments about contextualism, <i>not</i> their views on realism or antirealism
3.6	<i>Conflating relativism with descriptive claims</i>	Participants often conflate relativism with <i>descriptive claims</i> . This undermines the validity of the disagreement paradigm since many responses express judgments about descriptive claims, <i>not</i> their views on realism or antirealism
4	<i>Evaluative standard ambiguity</i>	Studies often fail to provide context that would resolve ambiguities about which moral standards people who disagree are indexing. Providing context may fail to resolve the problem since it may not be salient and may not provide a valid measure for all forms of realism (e.g., appraiser relativism)
5	<i>Abstract norm ambiguity</i>	Moral disagreements can be readily attributed to different ways of conforming to the same abstract moral principles. Participants often judge two people could both be correct because both actions conform to the same abstract moral principle, <i>not</i> because the participant endorses relativism.
6	<i>Misattributing source of disagreement</i>	People may misattribute the source of disagreement, e.g., to differences in nonmoral beliefs or misunderstanding the question, rather than a difference in fundamental moral values.
7	<i>Domain classification inconsistency</i>	Studies rely on <i>a priori</i> categorization of issues as "moral." However, research demonstrates that ordinary people (a) exhibit high levels of variability in which issues they consider moral <i>within</i> samples, (b) high levels of variability <i>between</i> populations and (c) systematically differ from researchers in what they consider moral issues. This can introduce considerable noise and interpretative difficulties and may reduce measurement invariance
8	<i>Presumption of correspondence theory of truth</i>	As Pölzler and Wright observe, studies on realism and antirealism presume participants are committed to a <i>correspondence theory</i> . However, there is little empirical evidence suggesting most participants endorse or are committed to correspondence theories of truth in relation to moral claims. If they do not, it is not clear if their responses are valid.
9	<i>Signaling and reputational concerns</i>	Participants may be motivated to select response options in order to signal desirable character traits rather than because those responses accurately reflect their views on realism and antirealism.

10	<i>Lack of realism</i>	Stimuli suffer a variety of problems that undermine their capacity to induce consistent, intended psychological states
10.1	<i>Lack of experimental realism</i>	Stimuli are often implausible or unrealistic, which can reduce engagement
10.2	<i>Lack of mundane realism</i>	Stimuli are often underspecified. This may force participants to “fill in the blanks” with their own inferences, decreasing consistency in interpretations
10.3	<i>Lack of psychological realism</i>	Stimuli sometimes include humorous or unrealistic stimuli (e.g., aliens that want to turn everything into pentagons), which can prompt unintended psychological states
11	<i>Lack of external validity</i>	Studies exhibit poor external validity due to overreliance on WEIRD populations, unrepresentative response options, and lack of ecological validity
11.1	<i>Overreliance on WEIRD populations</i>	Most studies are conducted on WEIRD populations and college students, limiting generalizability
11.2	<i>Stimulus as fixed effect problems</i>	Studies employing concrete moral issues are subject to the stimulus-as-fixed-effect fallacy
12	<i>Forced choice obscures indeterminacy</i>	Studies force participants to choose from available response options. This can create the misleading appearance of determinacy
13	<i>Inaccurate, biased, or misleading stimuli</i>	A problem originally described by Pölzler (2018b; 2018c) stimuli may mislead or bias participants, prompting unintended interpretations or motivating people to choose responses that inappropriately favor particular response options, though stimuli may also simply be inaccurate
14	<i>Questionable a priori categorization</i>	Hypotheses are derived primarily from concepts and distinctions drawn from contemporary analytic metaethics, a small, elite academic field almost exclusively composed of people from WEIRD populations, whose ways of thinking may not reflect how ordinary people think
15	<i>Intrinsic complexity</i>	<p>The metaethical concepts studies are designed to assess are:</p> <ul style="list-style-type: none"> (i) complex and difficult to grasp (ii) unfamiliar to ordinary people (iii) abstract and distant from everyday contexts of moral judgment (iv) there are typically unintended interpretations of stimuli that are more plausible than the intended (metaethical) interpretation <p>The complexity of metaethical concepts makes efforts to mitigate the conflation and ambiguities outlined in this list difficult, if not infeasible, using conventional social scientific methods.</p>
16	<i>Cumulative invalidity</i>	In isolation, the preceding problems may not decisively threaten the validity of the disagreement paradigm. But the cumulative effect of so many issues does. Adequately addressing all such issues may be too difficult without training participants, which could result in spontaneous theorizing or cause them to differ so much from the population they were sampled from that studies lose external validity

2.3 Inadequate response options

One problem with standard versions of the disagreement paradigm is lack of response options that reflect the full range of possible metaethical stances and commitments. There are two reasons why the response options used in standard versions of the disagreement paradigm are inadequate: (1) response options are *not jointly exhaustive*, and thus fail to provide response options that reflect the full range of possible metaethical positions²³ and (2) response options are typically *low resolution*: they *may* tell us whether the participant is a realist or antirealist, but even still, they don't tell us *what kind*.

2.3.1 Response options are not exhaustive

Although there is an overarching division between metaethical *realism* and *antirealism*, both positions may be further subdivided into a number of subcategories. Realists may be subdivided into naturalists and non-naturalists, while antirealists are generally divided into relativists, noncognitivists, and error theorists, though antirealism also includes constructivist and response-designating accounts, as well as less common positions, such *incoherentism* (Loeb, 2008) and my own view, *metaethical quietism*.²⁴ Thus, both realism and antirealism are umbrella terms that sit at the top of a hierarchy of lower-level categories and subcategories.

In principle, this does not pose any challenge to the validity of the disagreement paradigm.

The goal of standard versions of the disagreement paradigm is to simply determine whether the

²³ Pölzler (2018a; 2018b) discusses this shortcoming as well.

²⁴ Metaethical quietism is the view that the dispute between realism and antirealism is a pseudoproblem resulting from linguistic and conceptual confusions. In my own case, this is paired with the position that nontrivial accounts of realism are unintelligible. I cannot elaborate on these views here. Roughly, I maintain that notions such as irreducible normativity, or standard accounts of “reasons,” as they figure in contemporary analytic metaethics, are unintelligible, and thus by extension any metaethical account that appeals to such concepts is likewise unintelligible. Other accounts are merely *trivial* in a technical sense, which in most instances results from the account in question bottoming out in inert descriptive claims that lack the kind of authoritativeness or clout to compel action (e.g., typical naturalist accounts of realism). Although I reject the realism/antirealism dispute as a pseudoproblem, I still reject realism, making me an antirealist. However, since I reject other antirealist positions for buying into many of the conceptual errors that realism is subject to (e.g., confusions about “reasons”), I reject all conventional forms of antirealism as well. Given that my position rests on metaphilosophical foundations that reject the very framework in which metaethical positions are ordinarily classified, my position is difficult to place in conventional taxonomies.

participant believes (or is committed to) the claim that there are stance-independent moral facts (*realism*) or not (*antirealism*).²⁵ Yet in practice this is *not* what standard versions of the disagreement paradigm measure. Instead, participants are only able to judge whether two people who disagree about a moral issue (a) *are both correct* or if (b) *at least one must be incorrect*. This does not represent a jointly exhaustive pair of options. Participants could, in principle, think that *neither* side of a moral dispute could be correct. Yet they have no way to express this. Such a view is consistent with *noncognitivism*. Noncognitivism is the view that moral claims don't express propositions, and are thus incapable of being true or false. For instance, the moral claim "murder is morally wrong" could be understood as an imperative ("don't murder") or an emotional expression ("murder?! Boo!").

The failure to include adequate response options can also systematically *miscategorize* participants. Error theorists would have to choose option (b), and be classified as realists, even though error theory is an antirealist realist position. Error theorists believe that all claims about what is morally right or wrong are false, because they are implicitly committed to a false presupposition. For instance, if "stealing is morally permissible" presumes that stealing is *stance-independently* permissible, but there are no stance-independent moral facts, then this and all other moral claims like it are false, including the contrary moral claim that "stealing is morally impermissible." Since error theorists would agree that "at least one person must be incorrect," they would be included among realists. Changing the response option doesn't help either, e.g., changing it to "at most one could be correct" or indicating that just one of the two positions is correct isn't available to the antirealist, because they don't think anyone is correct.²⁶

²⁵ There is nothing wrong in principle with a simple, high-level measure that isn't designed to assess subcategories. If you want to study whether people like pizza, you don't *have to* exhaustively canvass everyone's favorite pizza toppings or risk having useless data. Just the same, it would be valuable in principle to know whether people are realists or antirealists.

²⁶ Technically, an error theorist could think it is logically possible that one of the positions is correct, but that it isn't *metaphysically* possible. I suspect researchers would have a difficult time designing measures that could detect this distinction.

Both the noncognitivist and the error theorist would think that *neither* side of a moral dispute could be correct, either because nobody could be correct *or* incorrect (*noncognitivism*) or because everyone is incorrect (*error theory*). Participants who hold such views have no way to adequately express their positions, leaving them without the ability to provide an accurate response to standard versions of the disagreement paradigm. Newer versions of the disagreement paradigm have corrected for this, by including noncognitivism or both noncognitivism and error theory (Beebe, 2015; Davis, 2021; Pölzler & Wright, 2020a; 2020b). However, this may solve one problem only to introduce another. Noncognitivism and error theory are subtle, sophisticated philosophical positions. By including additional response options, researchers increase the cognitive load on participants and risk introducing response options that are difficult to interpret as intended (Bush & Moss, 2020). Whatever these costs may be, the exclusion of legitimate response options is worse, since previous studies show that people do choose such options when they're provided. Forcing participants to choose from a narrow set of responses that they wouldn't otherwise choose risks undermining the validity of one's measures by mixing participants who don't endorse the response options they selected with those who do, and requiring participants to choose response options that don't reflect what they think. Such added complexity would only scratch the surface. I haven't even addressed the typical exclusion of different forms of realism (e.g., naturalism and non-naturalism) or antirealism (e.g., constructivism, relation-designating accounts, and so on), with the exception of Davis (2021), who does address different forms of realism, no other studies address any of these response options, and if they did, the complexity and number of response options would quickly become unmanageable.

2.3.2 Poor specificity

Standard versions of the disagreement paradigm also suffer from poor specificity. Even if the disagreement paradigm could tell us whether people are realists or antirealists, typical versions can't tell us *what kind* of realist or antirealist people are. Without additional follow-up questions, we cannot

further determine whether a realist endorses naturalism or non-naturalism, or whether an antirealist endorses relativism, noncognitivism, error theory, or some other position. The low resolution of the disagreement paradigm does not render it invalid. It simply means that it could at best only provide limited information.

One solution to this limitation is to ask a variety of auxiliary questions that allow one to provide a more fine-grained assessment of each participant's metaethical position. This is precisely what Davis (2021) did. Participants were asked whether they believed in God before follow-up questions, which influenced the subsequent set of questions they were asked. Once they were presented with the disagreement paradigm, participants were able to choose realism, relativism, error theory, or noncognitivism. Those who chose realism were then presented with a set of follow-up questions to determine whether they endorsed a supernatural, natural, or non-natural conception of realism. This is a step in the right direction, but it introduces a variety of potential concerns. Briefly, (1) the task becomes longer and more cognitively demanding, (2) introducing additional stimuli results in new ways to increase unintended interpretations and interpretation variation, (3) follow-up and auxiliary questions may still employ forced choices that limit response options and generate the artificial appearances of determinacy, (4) introducing additional questions can introduce novel methodological problems, (5) such approaches rely on potentially controversial judgments by researchers on how to classify different metaethical positions.

The disagreement paradigm also does not typically include different response options for different forms of relativism (Pölzler, 2018a; 2018b). In particular, there is no distinction between *subjectivism*, the view that moral claims are true or false relative to the standards of *individuals*, and *cultural relativism*, the claim that moral claims are true or false relative to the standards of different *groups*. Once Pölzler & Wright (2020b) included distinct options for each, they found a substantial number of participants who endorsed one or the other. This is yet another limitation that can be circumvented

by including additional response options, but once again at the cost of increased complexity and risk of unintended interpretation.

A less discussed limitation is the absence of any attempt to distinguish *appraiser* and *agent* relativism (Quintelier, De Smet, & Fessler, 2014). Relativism allows the truth of a moral claim to be relativized to the standards of the agent performing an action (agent relativism), the standards of those judging the agent judging the action (appraiser relativism), or both. Only one study tested for both, and found evidence of *both*, leading the authors to conclude that “there is inter-individual as well as intra-individual variation in whether individuals relativize moral standards to agents or appraisers” (p. 227).²⁷

In short, the disagreement paradigm does not typically allow us to measure people’s metaethical stances or commitments with any level of specificity, and any attempt to do so faces a tradeoff: the higher the resolution, the harder the task. With more response options, follow-up questions, and auxiliary questions used to pin down the precise metaethical stances or commitments of the participant, the more onerous the task and the greater the risk that something *else* goes wrong. This would not be a problem if the risk of something else going wrong were low, but as I show in the remaining sections and chapters, the risk is anything but low.

2.4 Conflations with unintended concepts

This section presents the most substantive and serious set of methodological shortcomings with the disagreement paradigm: *conflations*. *Conflations* occur whenever participants interpret stimuli in a particular unintended way, such that their response is influenced by (or fully reflects) their stance or commitment towards *something other* than what researchers were intending to measure.

²⁷ Notably, they acknowledge the increased complexity of drawing so subtle a distinction, adding that “Notwithstanding the fact that we excluded all participants who did not fill out all comprehension questions correctly, given the complexity of our scenarios and questions, future investigations would benefit from simpler materials” (p. 227).

There are two types of conflation: *formal conflations* and *informal conflations*. Formal conflations occur whenever stimuli *itself* conflates the intended interpretation with one or more unintended interpretations. Such errors can cause participants to systematically interpret stimuli in unintended ways due to researcher error, rather than participant misunderstanding. *Informal conflations* occur whenever participants conflate stimuli that have been presented accurately (i.e., without conflations) but participants nevertheless conflate the intended interpretation with some other interpretation(s), resulting in a response that does not reflect the measure of interest. Formal conflations could, in principle, be eliminated, though it may be infeasible to eliminate informal conflations. As such, the latter present the greater potential threat to the validity of the disagreement paradigm. The findings discussed in **Chapter 4** and **Chapter 5** suggest that unintended interpretations are the primary threat to folk metaethics research, and these unintended interpretations are, in turn, largely due to participants conflating stimuli intended to elicit judgments about realism and antirealism with other distinctions. Here, I document the most common conflations.

2.4.1 Conflating metaethics with normative ethics

Metaethical distinctions concern second-order questions about the nature of morality. Metaethics is distinct from *normative ethics*, which is concerned with first-order questions about what is in fact morally right or wrong, permissible or impermissible, etc. When a person judges that e.g., “murder is wrong,” this is a *first-order* (normative) position. When they judge that there is a stance-independent fact about whether murder is wrong, this is a *second-order* (metaethical) position. In other words, metaethical positions are “philosophical views *about* such first-order moral judgment” (Pölzler, 2018b, p. 657; emphasis original; see also Huemer, 2005, pp. 1-2).

Participants responding to the disagreement paradigm often appear to conflate metaethical considerations with normative considerations. For instance, suppose Alex expresses the view that abortion is morally permissible. Then Alex is told a previous participant judged that abortion is *not*

morally permissible. Alex is then asked whether both she and the previous participant can be correct, or if at least one of them must be incorrect. Alex might reason as follows: “Well, abortion isn’t morally wrong, so the other person is mistaken.” Alex may then select the response option “at least one of us must be incorrect.” This would be interpreted as a *realist* response, yet Alex’s thought process had nothing to do with realism. Alex was simply expressing a first-order moral judgment: abortion isn’t morally wrong, so someone who thinks that it is wrong is mistaken. When asked to explain their answers, many participants expressed first-order judgments of this kind, or, alternatively, simply stated that the person who disagreed with them was mistaken or a bad person (see **Chapter 4**, Study 1). Unfortunately, such responses cannot tell us whether the participant is a realist or not, since we cannot be sure whether the response option they selected reflects their metaethical position or simply reiterates their normative stance.

Unfortunately, participants are also far more familiar with expressing first-order moral judgments, and may have trouble “turning off” their tendency to think in terms of what’s right or wrong, rather than in terms of abstract considerations about what it means for something to be morally right or wrong. This possibility prompted Pölzler (2018b) to suggest that “Avoiding first-order moral intuitions in studies on folk moral realism altogether may be methodologically infeasible” (p. 658). If so, normative moral standards may represent an ineliminable impediment to measuring metaethical stances and commitments. At best, we may be able to mitigate the distorting influence of normative judgments, but doing so may come at the cost of further complicating studies and may limit the kinds of moral issues we can assess (e.g., we may not be able to assess more emotionally charged moral issues).

2.4.2 Conflating metaethics with epistemic concerns

Metaethical realism and antirealism are typically construed positions on the truth status of moral claims with *metaphysical* but not necessarily *epistemic* implications.²⁸ Metaphysical considerations are distinct, but related, to *epistemological* questions about how we can acquire moral *knowledge*, whether (and how) our moral beliefs can be *justified*, and whether we can be *certain* of our moral views (Pölzler, 2018). There is no easy way to disentangle metaphysical and epistemological considerations, since metaphysical stances often have epistemological implications, and vice versa. For example, if error theory or noncognitivism are true, then it is impossible to have moral knowledge because there are no moral facts. On occasion, philosophers also incorporate epistemic stances in their characterizations of moral realism, e.g., some claim that moral realism requires that we have (or can have) knowledge of at least some moral facts (Miller, 2009; Sinnott-Armstrong, 2009).²⁹

Despite their close relationship, questions about folk metaethical realism and antirealism are not intended to directly assess epistemic considerations about the means or possibility of moral knowledge or justification. Rather, they are concerned exclusively with whether there are moral facts, and if so, whether those facts are stance-independent (which are metaphysical or conceptual questions, *not* epistemic questions). Skepticism about moral knowledge is entirely consistent with the belief that

²⁸ I say “typically” because Parfit and Scanlon apparently maintain non-metaphysical notions of realism (see Veluwenkamp, 2017). This is likely a very uncommon view, and it’s unclear how plausibly it can be maintained. Nevertheless, Parfit is quite explicit on this point. As Veluwenkamp notes:

Parfit maintains that in the normative domain these truths have “no positive ontological implications” and are not “about metaphysical reality” [...]. And for Scanlon, normative truths “need no natural or special metaphysical reality in order to have the significance that we commonly grant them” [...]. (p. 751, see Parfit, 2011, vol. 2, p. 479, p. 747, and Scanlon, 2014, p. 52)

I have no idea what they are talking about. Of course moral realism has metaphysical implications. At the risk of sounding impertinent towards eminent scholars, I think these remarks are implausible and desperate attempts to insulate realism from objections through little more than stipulative fiat.

²⁹ This is a reasonable criterion to include, since it allows the realist to exclude undesirable forms of skeptical moral realism: that there are moral facts, but we can’t know any of them. People who believe that there are stance-independent moral facts are typically animated not just by the belief that they exist, but by the confidence (or at least hope) that we already know, or could eventually come to know, at least some of those moral facts.

there are stance-independent moral facts. Conversely, the belief that we can have moral knowledge may not directly entail whether that knowledge is of stance-independent facts or knowledge or relativized or response-dependent moral standards. In other words, skepticism about moral knowledge is consistent with realism, while a belief that moral knowledge is possible is consistent with antirealism. As a result, questions that do not neatly distinguish epistemological and metaphysical considerations from one another risk being unable to identify whether a response reflects an epistemic stance, a metaphysical stance, or both, and thus cannot serve as valid measures of realism/antirealism.

There are many ways in which participants could conflate metaethical and epistemic considerations. Here, I will focus on just one: *modal operator scope ambiguity*.³⁰ Consider the following claim:

- (i) Brazil and Germany could both win the World Cup this year.

This sentence could be making one of two distinct claims:

- (a) *Either* Brazil could win the World Cup this year, *or* Germany could win the World Cup this year, *but they cannot both win*.
- (b) Brazil and Germany could both win the World Cup this year *at the same time*.

To formalize this ambiguity, take $\Diamond(x)$ to mean that x is possible, take W to refer to the predicate “win the World Cup”, and take b to refer to Brazil and g to refer to Germany. This leaves us with two formalizations of the possible meaning of (i):

- (a) $\Diamond(Wb) \mathrel{\mathcal{C}} \Diamond(Wg)$ (exclusive)
- (b) $\Diamond(Wb \mathrel{\mathcal{C}} Wg)$ (inclusive)³¹

(i) is ambiguous between the first, *exclusive* reading, and the second, *inclusive* reading. In practice, context and pragmatics will tend to resolve this ambiguity. Given our background knowledge of the World Cup and the rules of tournaments in general, most of us know that only one team can win at a time,

³⁰ This problem was originally proposed by my colleague, Tyler Millhouse.

³¹ Tyler Millhouse formalized the inclusive and exclusive readings of the disagreement paradigm (Millhouse & Bush, 2016).

and this will lead us to conclude that when someone asserts (i), the exclusive interpretation is more plausible than the inclusive one. In other cases, context will tend to favor the opposite interpretation.

For instance, consider the statement:

(ii) Alex and Sam could both have cereal for breakfast this morning.

Despite its similarity to (i), there is little reason to assume that two people would be unable to eat the same thing for breakfast, so most people would lean towards an inclusive reading of this statement.

It is unclear how people interpret moral disagreements that exhibit the same ambiguity. However, versions of the disagreement paradigm closely mirror the wording used in these examples.

For instance, Beebe (2015) uses the following response options:

- (1) It is possible for both of you to be correct.
- (2) At least one of you must be mistaken.

The first option, “it is possible for both of you to be correct,” invites the same ambiguity between an inclusive and exclusive reading. This may lead participants to be uncertain how to resolve this ambiguity. They may interpret in the intended exclusive way, but they may also interpret in the inclusive way. One way the inclusive interpretation makes sense is if the participant is *uncertain* about whether the moral claim is true or false, and recognizes that future information could reveal that either they or the person they disagree with could ultimately turn out to be correct. If so, judging that they could both be correct would not represent a relativist stance because they will have interpreted the question as an (incomplete) *epistemic* question about moral knowledge, and not a metaethical question about whether conflicting moral claims are mutually exclusive.

This may be the best explanation for some otherwise puzzling results. Beebe et al. (2015) and Beebe and Sackris (2016) asked participants to judge disagreements about empirical claims, such as the age of the earth or the size of planets. Most people probably believe that claims like these have a single established answer that they could verify for themselves. But Beebe and colleagues also included

claims that may have a single, objectively correct answer in principle, but lack any practical means of verification; or, as they put it, are “not only unknown but practically unknowable” (Beebe & Sackris, 2016).

The most intriguing example involved conflicting claims about whether Julius Caesar ate soup on his 21st birthday.^{32, 33} If participants interpreted this question in the intended *inclusive* reading, they would have to consider whether two logically incompatible historical claims could both be true *at the same time*. It is possible that some participants might endorse a radical form of relativism, and believe that truth claims even in the domain of historical truth are relative. Nichols (2004) found some evidence of this in the explanations participants gave for their responses. However, these individuals represented only a handful of participants. Beebe & Sackris (2016) found that 45% of US participants judged that both people could be correct in 45% of the disagreements about Julius Caesar, while a similar pattern held for participants in China (53%), Poland (48%), and Ecuador (53%, Beebe et al., 2015). One plausible explanation for these findings is that participants interpreted what they were being asked as an epistemological question, and gave an exclusive reading of the question about whether both could be correct.

Scope ambiguity represents just one of the ways participants could conflate epistemic considerations with metaethical considerations. Even if scope ambiguity were minimized or eliminated, ordinary people may still be prone to conflating epistemic and metaphysical considerations more generally. Disentangling epistemic and metaphysical considerations is sufficiently difficult that training in philosophy may be necessary to develop the competence to recognize the distinction and tease it apart, and even philosophers, on occasion, slip in doing so. Like normative confluations, it may

³² Beebe et al. (2015) adapted the materials used by Beebe and Sackris (2016) to populations in Poland, China, and Ecuador, and modified the name of the historical figure to one that was more culturally relevant, e.g., Confucius.

³³ Beebe & Sackris (2016) asked about Caesar drinking wine rather than eating soup as in Beebe et al. (2015), but this difference is irrelevant to any points that I make here.

be difficult or impossible to fully eliminate the unintended influence of epistemic considerations on ordinary people's response to metaethical stimuli, and in practice, it may require so much training that it would be infeasible to provide participants with the requisite training; it may even be impossible, since the amount of training required would result in participants no longer being "ordinary people." We wouldn't expect, for instance, ordinary people to possess the requisite expertise to draw difficult, counterintuitive, and unfamiliar distinctions in mathematics or science, so it's puzzling that researchers would assume that they can, or simply ignore whether people could readily do so for highly abstract philosophical distinctions.

2.4.3 Conflating realism with universalism

Moral realism is often confused with *moral universalism*. Moral universalism is the view that a given moral principle or standard applies to all moral agents, regardless of their location in time or space.³⁴ For instance, if it is a universal moral fact that it is wrong to own slaves, then it is not only wrong to own slaves in the United States, it is also wrong to own them in any nation on earth, or anywhere else in the universe, so it would also be wrong even for aliens to enslave one another.³⁵ Moral universalism is sometimes contrasted with moral relativism, in that the former holds that moral standards apply to all people, while the relativist may hold that moral standards can vary depending on an individual's subjective values, or the standards of their culture or group. For instance, a relativist may claim that it

³⁴ A *moral agent* is any entity that is appropriately subject to moral appraisal. A typical adult human is a moral agent, while babies, nonhuman animals, and inanimate objects are not. This restriction is intended to limit the scope of universality to appropriate targets. A position may still count as *universal* even if it fails to hold lightning morally accountable for striking people.

³⁵ Assuming they possess the relevant characteristics to be appropriate subjects of moral consideration, e.g., relevant forms of agency such that they can be held to the same moral standards as humans. For the record, moral philosophers do not reference aliens as often as they should. For instance, most forms of group relativism seem to implicitly refer only to differences between human cultures. Yet in principle one could advocate species-relativism, and defend the view that moral standards can be correct or incorrect relative to the standards of an entire species, rather than to particular cultures within that species. This position is rarely explored, presumably because there are no known alien civilizations to compare ourselves to, but if there were, this might very well be a popular position. It is interesting to note, then, that the conceptual space of metaethical positions that people happen to defend seems to some degree circumscribed by contingent features of our circumstances. If multiple advanced species had evolved on earth (e.g., advanced elephantine or cephalopod civilizations) and existed today, species-relativism might be a common position.

is morally wrong for Catholics to have abortions, but it is not morally wrong for non-Catholics to have abortions.³⁶

Universalism represents one end of a continuum that represents the *scope* of moral facts. At one extreme, moral facts apply only on an individual basis: the individual subjectivist may believe that each of us ought to do that which is consistent with our personal moral standards. On such a view, the scope of a given person's moral standard, e.g., "do not murder," applies only to themselves. Someone else may endorse the same moral rule, but to the extent that moral rule applies to them, it does so in virtue of it being *their* standard, not someone else's. At the other extreme, the moral standard "do not murder" applies to everyone, everywhere, in all times and places. Other positions may fall somewhere between these two extremes e.g., cultural relativists may believe a moral rule applies to all members of a particular culture. It is also possible to believe some moral rules apply universally but others do not.

Regardless of where one falls on the continuum between moral facts being more or less universal in scope, universalism/localism is orthogonal to the distinction between realism and antirealism. Some researchers have drawn explicit attention to this and have sought to carefully avoid conflating universalism with realism (e.g., Goodwin and Darley, 2008).³⁷ Unfortunately, researchers have on occasion conflated realism and universalism (Yilmaz & Bahçekapili, 2015a; 2018; 2020). Even

³⁶ I provide this example rather than an example based on different cultures as a revolt against the common tendency to speak of relativism only in terms of cultural standards, rather than other potential group-based standards. Relativists need not hang their hat on one, and only one way in which one's standardized can be relativized.

³⁷ Goodwin and Darley are very clear not to mix the two up:

"[T]he question of whether ethical standards should apply to all cultures is a question about the *scope* of ethical standards, and is independent of the question of whether such standards and beliefs are objectively or subjectively true. Our interest centers on this second question, which concerns the *source* of such beliefs or standards - whether they derive their truth (or warrant) independent of human minds (i.e., objectively) or whether instead, their truth is entirely mind-dependent or subjective." (p. 1341)

so, participants often conflate universalism with relativism even when researchers don't formally conflate them (see **Chapter 4**).

2.4.4 Conflating realism with absolutism

People also conflate realism with *absolutism*. Absolutism is the view that certain moral rules or principles are *exceptionless* (Hawley, 2008). In other words, there are certain types of actions that an agent may not perform under *any* circumstances. For example, an absolute prohibition against killing other people would entail that it is wrong to intentionally kill people, even in self-defense or to prevent that person from killing a greater number of people.

Absolutism is a property of first-order moral norms, is orthogonal to the dispute between realism and antirealism, and *it does not entail realism*. A person who endorses absolute moral rules could have a stance or commitment towards cultural relativism or individual subjectivism. If so, then their endorsement of an absolute moral rule would best be understood as e.g., “there are no exceptions to this rule *according to my culture’s standards*,” or “there are no exceptions to this rule *according to my personal moral code*.” Likewise, realism does not entail absolutism. A realist may believe that there are stance-independent moral facts, but believe that those facts represent a body of rules and principles that are flexible and result in moral principles and rules that permit exceptions in particular circumstances. For instance, a realist may believe that it is morally wrong to kill people, *except* in self-defense or to stop someone from causing extraordinary harm to others. As I show in **Chapter 4**, the conflation between realism and absolutism is one of the most common causes of unintended interpretations.

2.4.5 Conflating relativism with contextualism

Another common source of unintended interpretations is the conflation between contextualism and relativism (Chappell, 2007).³⁸ Contextualism is the claim that whether a given action or type of action is permissible under some circumstances but not others. For instance, one might hold that it is morally permissible to kill someone in self-defense, but it is not morally permissible to kill someone for personal gain. Or it may be morally permissible to lie to an enraged murderer about the whereabouts of their intended victim, but morally wrong to lie to avoid responsibility for your actions.

Contextualism is *not* relativism. Contextualism is a feature of *normative* moral principles and standards. It concerns the sensitivity of a given category of action, e.g., “killing,” “lying,” “stealing,” to situational factors that influence the moral status of the action. Both realists and antirealists can favor moral principles that are sensitive to contextual considerations. Consider a contextualist account of the conditions in which killing is justified:

It is morally wrong to intentionally kill others except in self-defense, justified wars, or to prevent great harm.

A moral realist could believe this statement is stance-independently true. But a moral subjectivist could also think it is true, just not stance-independently true. In other words, *contextualism is completely orthogonal to moral realism and antirealism.*

Nevertheless, people routinely conflate relativism with contextualism. This confusion likely stems from the fact that both relativism and contextualism allow the moral status of an action to vary. Relativism has to do with moral claims being true or false on the basis of which standards they are indexed to. As such, relativism allows *the exact same action* to be right relative to some moral frameworks, and wrong relative to others. Contextualism doesn’t allow for this. Contextualism has to do with moral actions being right or wrong on the basis of specific details relevant to a given situation, and that can

³⁸ I use the term “contextualism,” here, but there is no standard usage in the literature, and it is also used to refer to other concepts or distinctions which I do not address here (e.g., Brogaard, 2008; Evers, 2014).

vary from one situation to another. Thus, it allows for an *action type* to be right in some cases and wrong in others, but the rightness or wrongness always turns on *differences* between these circumstances, e.g., lying could be morally wrong in situations where it is used for selfish gain, but morally right in situations where it would save lives.

Both relativism and contextualism allow the truth of some general moral action, like “stealing,” to be morally good sometimes and morally bad other times, but for completely different reasons. Thus, in a certain sense, relativism *does* hold that a moral claim is true or false “depending on the context,” but it refers to a very distinct type of context: the context of utterance. *Who* makes the claim and *what* moral framework is indexed by their claim determines the truth conditions for the statement in question. Yet this sense of “context” has nothing to do with the contextual considerations relevant to contextualism. Contextualism treats a given moral action as morally right or wrong on the basis of distinctive features of a given moral situation *other than* standards of the people engaging in moral actions or judging the moral actions of others. This is a subtle distinction that is difficult to convey. It is no surprise that it is one of the most common confluences people express (see **Chapter 4**).

2.4.6 Conflating relativism with descriptive claims

Some people believe abortion is morally permissible. Others believe abortion is morally wrong. The members of some societies are more likely on average to consider a particular action wrong than members of another society. Such observations involve a *descriptive* claim to the effect that different individuals and societies have different moral standards. And these claims are completely independent of whether the respective standards are *correct* or *incorrect*, and they are likewise independent of whatever metaethical position one might endorse. Yet, for reasons I cannot fathom, people seem to conflate the banal observation that different people have different moral standards with relativism. Not just participants, but researchers, too. As I discuss in Bush and Moss (2020), several of the items Yilmaz and Bahçekapili use conflate descriptive claims with metaethics, e.g., “*What is moral varies on the basis of*

context and society.”³⁹ Yet even when researchers avoid such conflations, they are rampant among participants across a variety of studies. For instance, participants were asked what one of the items on Zijlstra’s (2019) *Folk Moral Objectivism* (FMO) scale means:

“What is morally right and wrong is relative to the moral beliefs of an individual, culture, or society”

They routinely responded with statements like these:

Different groups of people have different morals.

Different cultures and societies have different morals to beliefs as to what is right and wrong.

It is true that different individuals and cultures have different morality belief systems

Responses like these are very common, and illustrate that even when people are presented with a claim that *explicitly* describes morality as “relative,” they frequently interpret such claims as *descriptive*.

2.5 A summary of remaining problems

There are many more methodological shortcomings, but they cannot be adequately addressed in detail. I will quickly summarize the remainder, before concluding with the implications not simply of each methodological worry taken in isolation, but the *cumulative* effect of so many.

2.5.1 Evaluative standard ambiguity

If two people who disagree about a moral issue are both referring to the same moral standards (such as the standards of a shared culture), relativists would agree that at least one of them must be incorrect. Studies that don’t specify which moral standards people’s claims are referring to risk encouraging some antirealists to choose the “realist” response option. Attempts at clarifying would complicate studies and may be ineffective.

³⁹ Note that they also conflate relativism with contextualism in the very same item.

2.5.2 Abstract norm ambiguity

There are different ways to comply with the same abstract moral rule. For instance, the proper way to respect the dead is burial in some cultures, and cremation in others. Participants may judge that people from these cultures may disagree, and both be correct, not because moral truth is relative, but because they believe both cultures have identified appropriate ways to comply with the same moral rule. This may result in participants choosing the antirealist option even though this position has nothing to do with realism or antirealism (but rather with the application conditions of normative standards).

2.5.3 Misattributing source of disagreement

Participants may attribute the source of the disagreement to something other than a fundamental difference in moral values, e.g., to a disagreement over the nonmoral facts or to misunderstanding the scenario. If so, their responses cannot serve as valid measures of realism or antirealism because the validity of the disagreement paradigm requires participants to interpret the cause of the disagreement as a difference in moral values.

2.5.4 Domain classification inconsistency

Most studies present participants with items that researchers have classified themselves as “moral.” Yet ordinary people do not agree with researchers or with one another about which issues are “moral” or not (Levine et al., 2021; Machery, 2018; Wright, Grandjean, & McWhite, 2013). Researchers who ignore this problem are unable to assess the extent to which this undermines the validity of their results or at least renders them much noisier. Researchers who account for this issue by asking participants to classify issues as “moral” or not (a) face yet another challenge, in that participants won’t agree with researchers or one another about what it means for an issue to be “moral,” (b) face validity issues with this task as well (e.g., interpretative difficulties, performance errors, etc.), and (c) if such measures

result in excluding responses participants don't classify as moral, cross-participant comparisons will be compromised, perhaps severely.

2.5.5 Presumption of correspondence theory of truth⁴⁰

The disagreement paradigm presumes participants endorse a correspondence theory of truth. There is no evidence that they do so with respect to moral claims (Pölzler & Wright, 2020a). The only pair of studies to specifically address this question suggest ordinary people *don't* exhibit a clear and uniform commitment to a correspondence conception of truth (Barnard & Ulatowski, 2013; Reuter & Brun, 2021). Instead, both studies find that ordinary people frequently reject features of correspondence theories and exhibit a variety of different positions, and may have multiple or (in Reuter and Brun's case) ambiguous meanings in ordinary language. For instance, Reuter and Brun claim that their findings “reveal that the use of ‘true’ shows substantial variance within the empirical domain, indicating that ‘true’ is ambiguous between a correspondence and a coherence reading” (p. 1).

2.5.6 Signaling & reputational concerns

Given the centrality of our moral thoughts, values, and behavior, ordinary people may be motivated to select response options on the basis of what they perceive would signal socially desirable characteristics. This may be exacerbated by a tendency to ascribe normative or ideological implications endorsing metaethical positions, or at least the kind of phrasing used to describe metaethical positions.

⁴⁰ This is a problem that was initially proposed by Pölzler (2018a; 2018b) and discussed in Pölzler & Wright (2020a; 2020b). It had not occurred to me to consider the dependence of folk metaethics on other philosophical presumptions that are often taken for granted. It leads me to wonder whether there are others that may threaten the validity of folk metaethics. Yet it also hints at a far greater worry: that many, perhaps *all* folk philosophical thought is so deeply nested in the rest of each individual's philosophical beliefs and commitments that the only way to assess any particular person's philosophical views would be holistically, taking all of their views (or at least the relevantly connected ones) into account. That is, any particular position a person takes on any given issue may only make sense by taking into consideration the rest of their beliefs and commitments, in which case studying individual philosophical views in isolation from the rest of each person's views may make little sense.

This could motivate some participants to select response options for reasons unrelated to a sincere expression of a metaethical position.

2.5.7 Lack of realism

Lack of *realism* can also threaten the external validity of the disagreement paradigm (Calder, Phillips, & Tybout, 1982; Findley, Kikuta, & Denly, 2021). There are at least three ways that such stimuli can lack external validity and thus exhibit poor generalizability:

(i) Lack of experimental realism: Participants may be less attentive or motivated to engage stimuli they find unserious or unrealistic (e.g., people may take a scenario involving space pirates less seriously than one involving ISIS)

(ii) Lack of mundane realism: *Unfamiliar* scenarios with *impoverished stimuli* may prompt responses that don't reflect more prototypical, familiar, and contextually rich instances of the relevant phenomena (e.g., people's moral evaluation of scenarios involving space pirates may tell us very little about moral judgment and behavior in real situations)

(iii) Lack of psychological realism: Stimuli that are amusing, strange, unrealistic, or difficult to imagine may prompt different psychological processes than stimuli that lack these features (e.g., if people find material funny, elevated mood may have an undesirable influence on how they respond)⁴¹

2.5.8 Lack of external validity

Almost all research has been conducted on WEIRD populations, with much of it conducted specifically among college students. What research has been conducted cross-culturally is both limited and limited both in terms of how many studies have been conducted and how representative the relevant samples are. This limits the generalizability of most folk metaethics research (Henrich, Heine, & Norenzayan, 2010). Furthermore, most studies represent the moral domain via a motley array of concrete moral issues that are inappropriately treated as representative of the moral domain as a whole,

⁴¹ These distinctions are my interpretation and integration of the categories proposed by Pölzler(2018b) and Bauman et al. (2014, p. 537). See Bauman et al. (2014) for a deeper discussion of these problems as they relate to popular stimuli used in moral psychology. The threats proposed here are more of a problem for that research than they are for research on folk metaethics.

exposing all research relying on concrete moral items to the stimulus-as-fixed-effect fallacy, further limiting the generalizability of existing studies (Baguley, 2012; Judd, Westfall, & Kenny, 2012).

2.5.9 Forced choice obscures indeterminacy

Almost all studies employ forced choice paradigms that require participants to endorse some form of realism or antirealism. This obscures the possibility of metaethical indeterminacy.

2.5.10 Inaccurate, biased, or misleading stimuli

Stimuli, including response options and instructions, can be inaccurate or misleading, or bias participants towards particular response options. While this is a mundane problem common to all research, metaethical concepts are subtle and hard to describe, and researchers have routinely misoperationalized realism, antirealism, and subcategories of each, suggesting it is especially difficult to operationalize metaethical concepts.

2.5.11 Questionable a priori theorizing

Studies rely on concepts and distinctions derived from contemporary analytic metaethics. Many of the relevant distinctions emerged in the 20th century and have been primarily described and refined among a highly educated academic elite, almost exclusively in the Anglophone world. Their distinctive way of thinking about the nature of morality may fail to represent how ordinary people think. Yet studies rely exclusively on hypotheses derived from this discipline, *not* a richer foundation in psychological research.

2.6 Conclusion

The previous sections provide concrete and specific reasons why responses to the disagreement paradigm may fail to accurately reflect whether participants are realists or antirealists. Setting these issues aside, we may also take a bird's eye view towards the problem of empirically assessing folk

realism and antirealism. Philosophers have historically presumed that the way ordinary people speak and think is underwritten by an implicit commitment to one or another of competing metaethical accounts. It's not clear to me why they've been so convinced that there is a determinate, shared folk metaethics that the presumption that there is has gone virtually unchallenged in academic philosophy. Yet even if the brick and mortar of everyday claims like "Alex is an asshole" and "Sam shouldn't lie" best fits some realist or antirealist semantic theory, it's not obvious we should expect to readily determine what that theory is by asking people questions about metaethics. And it's even less clear that we should reasonably expect people to have explicit metaethical stances. If contemporary analytic metaethics has demonstrated anything, it is that metaethics is *complicated, subtle, and difficult*.

In our experience, David and I have found that students struggle to understand metaethical concepts even under optimal educational settings where they have ample opportunity to read about the relevant distinctions and receive instructions from specialists (Bush & Moss, 2020). David and I have also pointed out that philosophers routinely accuse one another of failing to accurately characterize particular metaethical distinctions, noting that if "trained professionals struggle to grasp these concepts, it is unclear why we should suppose that untrained laypersons should readily do so" (p. 7). In addition, we've outlined several criteria that, along with the complexity of metaethical concepts, cast serious doubt on the plausibility that ordinary people could be reasonably expected to interpret questions about metaethics as intended:

- (i) the relevant metaethical theories are complex and difficult to grasp*
- (ii) most people are unfamiliar with these distinctions prior to encountering them in studies*
- (iii) metaethical theories are generally abstract and distant from real world practical questions lay populations would be more familiar with and expect to be asked about*
- (iv) there are typically plausible non-metaethical interpretations of the questions posed to respondents* (p. 6)

Given these considerations, it may be *less* likely that people's responses reflect people's stances and commitments towards realism and antirealism than other unintended intended distinctions.

The many methodological problems highlighted in this chapter corroborate this claim. *Some* of the issues can be mitigated or avoided. Considered in isolation, it may seem the disagreement paradigm could survive any particular methodological limitation. Yet each problem would typically require extending the instructions, providing more response options, or otherwise expanding the length and complexity of the disagreement paradigm, which could result in a different set of methodological problems: the resulting “corrections” would result in a bloated monstrosity of a study, so laden with clarifications and instructions and complicated response options that even professional philosophers would balk, while ordinary people would be left with the task of taking what amounted to a brief introductory course in philosophy.⁴² I discuss efforts to devise *training paradigms* that do just this in the next chapter.

⁴² Indeed, this is precisely what has happened with efforts to design *training paradigms*: Pölzler (2021) reports that it took on average *fifty minutes* to complete the studies appearing in Pölzler and Wright’s (2020a; 2020b), and I don’t think their efforts were adequate (see **Chapter 3** and **Supplement 3**). In my own experience, it would take far more work than a few minutes for participants to develop adequate understanding of what’s at stake in the dispute between realists and antirealists.

CHAPTER 3: Critique of Alternative Paradigms

3.0 Introduction

In addition to the disagreement paradigm, researchers have employed several other methods to study folk metaethics. The primary alternative is to devise *scales* in which participants are asked to express their level of agreement or disagreement with a set of sentences. These sentences are intended to express various metaethical positions (e.g., realism, relativism, and so on) with an aggregate score in any given subscale reflecting one's overall endorsement of that particular metaethical position. *Training paradigms* are another prominent alternative. This approach involves instructing participants in the relevant metaethical distinctions, and may include exercises intended to reinforce and confirm that participants understand the relevant distinctions.

There are a variety of other paradigms as well, often appearing in only one or a handful of studies. I address all of these paradigms in **Supplement 3**. The extremely short version of that commentary is that, without exception, *all of these paradigms are invalid*. Here, I will only focus on scales and training paradigms. Each serves to highlight the distinct methodological shortcomings and alternatives to the disagreement paradigms. These can be summarized as follows: (1) scales consistently suffer poor face validity, and, even when they don't, participants still struggle to interpret items that appear on these scales as intended. Training paradigms seek to minimize unintended interpretations, however (2) it's not clear they succeed at this task, and if they did, this would simply trade the problem of ensuring stimuli are interpreted with another, equally serious problem: once participants have been trained in the relevant metaethical distinctions, we can no determine whether their responses reflect metaethical stances or commitments held *prior to* training, or whether the training *induced* participants to adopt determinate metaethical stances, a process I call *spontaneous theorizing*.

3.1 Scales

The most prominent alternatives to the disagreement paradigm are metaethics *scales*. Metaethics scales present participants with a series of short statements that are intended to represent one or more metaethical positions. Participants are then asked to express a stance towards each such statement. This typically involves the use of Likert scales, with participants expressing the degree to which they agree or disagree with each statement. For instance, Collier-Spruel et al. (2019) present participants with a series of statements that reflect belief in relativism. Here are a pair of examples:

- (1) *The viewpoint of one's culture determines whether their actions are morally right*
- (2) *An action is only morally wrong if a person believes it is morally wrong*

Participants are asked to express how strongly they agree or disagree with each of these statements on a scale of 1 to 5 (1 = strongly disagree, 5 = strongly agree). Responses to these items can be totaled to provide an aggregate score of degree of belief in relativism, with higher scores representing greater belief in relativism, and lower scores reflecting disbelief in relativism. Researchers can also assess average level of agreement for clusters of items that represent a particular metaethical position, which could provide insight into whether a particular population leans more towards or against relativism (or whatever metaethical positions are included in the scale).⁴³

At present, there are six genuine metaethics scales, and two “quasi-scales” (clusters of *ad hoc* items used in place of a full scale): the relativism subscale of the Ethics Position Questionnaire (*EPQ*, Forsyth, 1980), the Objectivism-Subjectivism scale (*TOS*, Trainer, 1983), a set of 3 items used by Yilmaz & Bahçekapili (2015b), the New Meta-Ethics Questionnaire (*NMQ*, Yilmaz & Bahçekapili, 2015a), a three-item moral objectivism scale (*MO3*, Sarkissian & Phelan, 2019), the Moral Relativism

⁴³ Not all scales assess use Likert scales. Trainer (1983) presented participants with a series of opposing pairs of metaethical statements that reflected realism and antirealism (Trainer used the terms “objectivism” and “subjectivism”) and asked participants to choose the one they favor.

Scale (*MRS*, Collier-Spruel et al., 2019), and the Folk Moral Objectivism scale (*FMO*, Zijlstra, 2019).⁴⁴ Details of each are summarized in **Table 3.1** (complete scales are available in **Appendix C**). To keep the tedium of an item-by-item analysis to a minimum, I provide a critique of Sarkissian and Phelan’s (2019) 3-item quasi-scale below. However, substantive problems often emerge only by analyzing specific items. These critiques appear in **Supplement 3**, where I also provide a general overview of the advantages and disadvantages of scales.

3.2 Problems of validity in metaethics scales: An example

For the sake of brevity, I will focus on one of the shortest scales, Sarkissian and Phelan’s (2019) 3-item moral objectivism scale (*MO3*). This allows me to present a focused discussion of the distinctive issues with each of these items rather than spreading myself thin by addressing all of the items on a larger scale. There are a few other reasons for selecting this particular scale: these items suffer many of the same problems as items that are not discussed in the main text. For extended discussion of problems with other scales, see **Supplement 3**.⁴⁵ In addition, these items were *explicitly* designed in response to methodological concerns about the poor validity of earlier items, namely those that appear on Forsyth’s (1980) *Ethics Position Questionnaire* (*EPQ*). The *MO3* appears to be an *ad hoc* set of measures devised in the midst of a project, without any explicit efforts made to validate the items beyond

⁴⁴ Notably, Forsyth’s *EPQ* (1980) and Trainer’s *TOS* (1983) predate the disagreement paradigm by several decades. Yet neither managed to spawn a literature specifically dedicated to the study of folk metaethics in a way that is well-integrated with the philosophical literature in the way contemporary research on the psychology of metaethics has managed. Like the fall of the Roman empire, any official start date to the present era of folk metaethics research will be largely a matter of convention, but if I had to choose a date, it would be with the publication of Goodwin and Darley’s 2008 publication, “The psychology of meta-ethics: Exploring objectivism.” While several articles preceded it, including the aforementioned Forsyth (1980) and Trainer (1983), as well as Nichols and Folds-Bennett (2003), Nichols (2004), and Wainryb et al. (2004), none of these seemed to attract the attention that Goodwin and Darley’s article did. The latter seems to have attracted enough interest for the topic to reach critical mass, and establish itself a sustained area of research.

⁴⁵ This should be qualified by acknowledging that some scales went through far more rigorous validation efforts, and resulted in items that suffered fewer problems than those discussed here. Nevertheless, these items still suffer poor validity (See **Supplement 3**), and participants still exhibited low rates of clear intended interpretation (see **Chapter 4**). Thus, while other items may be better, they’re still not valid. For comparison, a very accurate description of Gödel’s incompleteness theorems may be less prone to particular confusions than a very bad one, but it’s still plausible that most people would remain confused and incapable of understanding the theorems.

reporting Cronbach's alpha, which does not provide adequate (or even very good) evidence of validity (Sijtsma, 2009). In particular, in describing the virtues of the MO3, they state that it was “not based on previous measures,” adding that they “[...] were concerned that existing measures, such as the 5-item scale adapted from Forsyth (1980), did not capture metaethical views in a precise way, potentially limiting what we can infer from their use in this context” (p. 4).

Table 3.1

Summary of metaethics scales

Scale	Full name	Reference	Items	Construct
EPQ	Ethics Position Questionnaire	Forsyth (1980)	10	relativism
TOS	Trainer-Objectivism subjectivism scale	Trainer (1983)	17	objectivism, subjectivism
YB3	3-item face valid metaethics scale	Yilmaz & Bahçekapili (2015b)	3	objectivism, subjectivism
NMQ	New Meta-ethics Questionnaire	Yilmaz & Bahçekapili (2015a)	8	objectivism, subjectivism
MO3	3-Item Moral Objectivism Scale	Sarkissian & Phelan (2019)	3	objectivism, relativism
MRS	Moral Relativism Scale	Collier-Spruel et al. (2019)	10	relativism
FMO	Folk Moral Objectivism scale	Zijlstra (2019)	20	no truth, relativism, universalism, absolutism, divine command theory
JRT5	Objectivity of Morality scale	Johnson, Rodrigues, & Tuckett (2020)	5	objectivism, subjectivism, normativity

Unfortunately, the MO3 does not capture metaethical views in a precise way, either. Although their scale purports to measure “objectivism,” they do not offer an explicit description of what they take objectivism to mean. However, it is reasonable to infer that they have in mind roughly the same

conception as the researchers they cite, such as Goodwin and Darley (2008). Thus, the construct they have in mind seems to correspond to moral realism in that objectivism refers to the claim that there are stance-independent moral facts. I provide a brief summary of the problems with each of the three items here in **Table 3.2**.

Table 3.2

Summary of methodological problems with the MO3

Methodological problems by item
<p><i>There exists a single moral code that is applicable to everyone, regardless of any individual person's beliefs or cultural identity.</i></p> <ol style="list-style-type: none"> (1) Conflates universalism with stance-independence (2) Exhibits abstract norm ambiguity (conflating multiple ways of conforming to the same abstract moral rule with relativism) (3) "Applicable" may be interpreted in practical or normative terms (4) Use of the term moral "code" encourages unintended interpretations (5) Could be interpreted as a descriptive claim (6) Could prompt conflation with absolutism vs. contextualism
<p><i>If two people really disagree about a particular moral problem then at most one of them can be correct, since moral problems cannot have multiple correct answers.</i></p> <ol style="list-style-type: none"> (1) Exhibits problems associated with disagreement paradigm (2) Exhibits abstract norm ambiguity (conflating multiple ways of conforming to the same abstract moral rule with relativism). The wording of this item is especially vulnerable to this problem (3) Prohibits expressing metaethical pluralism by presuming intrapersonally uniform metaethical stance (4) Unclear what "really" means. Does not clearly convey fundamental moral disagreements, and is thus subject to attributing source of disagreement to nonmoral differences (5) Use of the term "problem" rather than "claim" may amplify abstract norm ambiguity (6) Could prompt conflation with absolutism vs. contextualism (7) Worded as a negation with use of "cannot," which may increase cognitive load
<p><i>It is possible to compare different cultures by a single, universal standard of moral rightness.</i></p>

- (1) Conflates universalism with stance-independence
 - (2) Completely lacks face validity: whether it is possible to compare different cultures by a single standard does not entail that that standard is true, much less stance-independently true.
 - (3) Unclear: it's not clear what it means to say that it's "possible" to "compare" different cultures by a "single, universal standard."
 - (4) May prompt conflation between metaethics and normative ethics
 - (5) May conflate realism intersubjectivity or constructivism
 - (6) May conflate realism with absolutism, close-mindedness, rigidity, and other undesirable traits
-

3.2.1 MO3 item #1

MO3 #1 <i>realism</i>	There exists a single moral code that is applicable to everyone, regardless of any individual person's beliefs or cultural identity.
---------------------------------	--

This item conflates realism with universalism. Universalism holds that there is a single standard of moral truth, but this concerns the *scope* of moral claims (i.e., who they apply to), not what makes them true. Unfortunately, *who* a moral standard applies to has nothing to do with whether that standard is stance-independent or not. A moral antirealist could endorse a normative moral rule, e.g., "it is morally wrong to steal," and endorse the claim that this rule applies to *everyone*. For instance, ideal observer theory holds that there is a single correct moral standard, but this standard depends on what an ideally rational and fully informed agent would endorse (Firth, 1952). Thus, there is one correct moral standard, but that standard depends on the stance of a hypothetical agent. Conversely, a moral realist could believe there are stance-independent moral facts, but at least some of those facts aren't universal. They might think that e.g., different (stance-independent) moral standards apply to different people. More generally, people could believe that there can be multiple correct standards, but that does not necessarily make the stance-dependent, since it is possible to relativize moral claims to stance-independent criteria (Joyce, 2015). In short, the distinction between whether there is one or multiple correct moral standards is orthogonal to the distinction between stance-independence and stance-

independence. Asking people whether there are *universal* moral rules cannot serve as a legitimate measure of realism.⁴⁶

Nevertheless, one might reason that universalism is an excellent proxy for realism. We may suspect this is the case by starting with the observation that we tend to be realists about other domains, such as science and math. Then, we might suppose, since we tend to think there is one stance-independent standard of truth in these domains, we may imagine that the believing there is only a single body of facts with respect to that domain is a good indication that the domain is not merely universally true but true in a stance-independent way. In short, if universalism tends to coincide with realism in other domains, it may plausibly do so with respect to realism as well. This is not an unreasonable assumption. Yet it requires us to make a theoretical assumption about the relation between universalism and realism that may be unwarranted, and at best leaves us with an indirect measure of realism. It is also possible people interpret the claim as descriptive. That is, it could be that people believe the item reflects the claim that, as a matter of fact, there are standards all societies accept, e.g., descriptive universals, such as rules against rape or murder.

Another issue is that it is unclear what “applicable” means. While philosophers may recognize that this is intended to convey that everyone’s actions are right or wrong in accordance with the same standards, and that praise, blame, and whatever else may be associated with morality is justified and in all of these cases, ordinary people may not know what it means, or may interpret “applicable” in a more forceful, literal, and (most importantly) *normative* way.

⁴⁶ This conflation is unfortunate given that Goodwin and Darley’s (2008) earlier research on folk metaethics explicitly drew attention to the distinction and made it very clear that they sought to avoid the conflation:

“[...] the question of whether ethical standards should apply to all cultures is a question about the *scope* of ethical standards, and is independent of the question of whether such standards and beliefs are objectively or subjectively true [...] our interest centers on this second question, which concerns the *source* of such beliefs or standards – whether they derive their truth (or warrant) independently of human minds (i.e., objectively), or whether, instead, their truth is entirely mind-dependent or subjective.” (p. 1341)

This is especially unfortunate since Sarkissian and Phelan discuss this exact paper in their article (see p. 2).

Yet another concern with this item is the use of the term “code.” Participants may interpret this as a statement about whether moral standards should be formalized, written down, or otherwise enshrined in our institutions or laws. If so, they will have interpreted “code” in a literal but unintended way. In addition, the notion of a moral “code” may evoke notions of an especially absolutist or dogmatic moral system, or of stark and inflexible rules that are insensitive to situational factors. Insofar as participants tend to hold a favorable or unfavorable attitude towards a rigid and absolutist moral system, this may prompt participants to express greater or lesser agreement with this item for reasons unrelated to metaethics. For instance, participants may interpret disagreement with this item to reflect particularism or sensitivity to context, even though such positions are consistent with agreement, since people could believe there are universal, but flexible and context-sensitive moral principles.

A related, but subtly distinct conflation could occur if participants interpret this statement to imply that there are universal means of complying with a particular moral principle or rule. This is the problem of *abstract norm ambiguity* discussed in chapter 2. For instance, there may be a general moral rule that we should “show respect to others.” However, the precise way in which we ought to do this may vary across cultures e.g., the proper way to show respect may be to bow, kneel, or shake hands. Participants may believe this statement would reflect the view that there is only one proper way of conforming to an abstract moral rule, and reject it on those grounds. Yet again, this is not what agreement with such a statement would entail.

Moral rules may also apply to some people or groups in virtue of their particular status, position, or relation to others. For instance, parents have special duties and responsibilities to their children, doctors have special duties to their patients, leaders have responsibilities over their subordinates, and so on. In other words, people may have *positional duties*, or duties that they possess in virtue of their status, position, or role within a particular social context Babushkina (2019). Such positional duties may be *universal* in a conditional sense, i.e., “if you are the parent of a child, then you

have an obligation to care for that child,” or “if you are a police officer, you have an obligation to apprehend someone committing a crime,” and so on. Such rules may be universal, but not necessarily apply to everyone, simply because not everyone occupies the same status, position, or role within a society. Yet consider the wording of the item: “There exists a single moral code that is applicable to everyone [...]” This *could* be interpreted as a view that we have only nonpositional duties, in that the same set of rules apply to all people everywhere, regardless of their particular role or relation to others.

This possibility may not be aided, and may even be exacerbated, by the latter half of the item, “[...] regardless of any individual person’s beliefs or cultural identity.” *Philosophers* may understand this to mean that a person’s moral beliefs, or the moral standards of their culture, are irrelevant with respect to the moral facts, because those facts are stance-independent, and thus not made true by the standards of individuals or groups. Yet is this how nonphilosophers without training would interpret this phrase? This is an especially relevant question given that laypeople are unfamiliar with stance-independence and such considerations would probably not be salient without additional instructions or context in the questionnaire that the item is included in. In other words, this phrasing may *seem* to be a straightforward remark about how the moral facts in question are not stance-dependent, but this may only be obvious if one has the requisite training to recognize the purpose of this phrasing.

To make matters worse, this isn’t even the standard phrasing philosophers themselves would typically use. What does it mean to say that the same moral code applies to everyone regardless of that person’s cultural *identity*? Philosophers may wish to specify the *standards* of that person’s culture, rather than their identity. More generally, it is simply unclear how individual beliefs or cultural identity are intended to relate to the applicability of the putative moral rules; it is left to the participant to understand that the intended relation is to serve as truth-makers for moral facts. This is a specific and sophisticated relation that would be difficult to convey, and it’s asking a lot to expect nonphilosophers to reliably interpret it this way.

3.2.2 MO3 item #2

MO3 #2 <i>realism</i>	If two people really disagree about a particular moral problem then at most one of them can be correct, since moral problems cannot have multiple correct answers.
--	--

This item is an abstract, scale-version of the disagreement paradigm. It is abstract in that it refers to the moral domain as a whole, rather than to concrete moral disagreements. As such, its most serious shortcoming is that it inherits many of the same methodological problems as the disagreement paradigm itself.⁴⁷

Another problem with this item is that it is extremely vulnerable to *abstract norm ambiguity*, since a realist could believe that a “moral problem” could have multiple correct answers. What distinguishes a realist from an antirealist is not whether they believe moral problems have one solution rather than multiple solutions, but whether they believe the solution(s) to a problem are made true in virtue of stance-independent moral facts. Moral realists are thus *not* obligated to believe that “moral problems” have only one correct solution. Instead, they may believe that a *range* of actions could be permissible given the same abstract moral standard. By failing to make explicit that the disagreement is about the truth status of a distinct moral claim, this item conflates fundamental moral disagreement with disagreement over the number of ways one could conform to a given moral standard. The latter possibility is by no means obscure or esoteric. Consider the following actions:

- (1) Building a bridge
- (2) Defeating an opponent in chess
- (3) Writing a computer program that solves a specific problem

⁴⁷ One small but clever potential advantage with this item is the inclusion of the phrase “really.” This could help to emphasize that it is referring to *fundamental* moral disagreement, rather than a disagreement over nonmoral facts or in some other way not reflecting moral stances that genuinely conflict with one another in terms of their substantive moral content. However, it is not at all obvious that untrained participants would pick up on the use of “really” and interpret in the intended way.

Even with well-specified goals that allow for one to unambiguously categorize attempted solutions as correct or incorrect, all of these examples allow for *multiple* solutions to be correct. Nevertheless, there are stance-independent facts about the correct and incorrect ways to achieve each of these goals. It simply isn't the case that if one believes that the facts in a given domain are stance-independent that one believes that, for any problem within that domain, there is only one correct solution. To suggest this is to favor a form of extreme rigidity that many people would reject, regardless of whether they were realists or not. Such apparent rigidity may have given participants the mistaken impression that this item was asking about *absolutism*, the belief that there are exceptionless moral rules. Such views may be unappealing both because they are inconsistent with people's normative moral standards and because they signal undesirable character traits (e.g. close-mindedness).

Finally, this item seems almost ideally worded to encourage abstract norm ambiguity. Rather than referring to people disagreeing about a moral claim, or issue, it describes people disagreeing about a moral "problem." A "problem" is *precisely the kind of term people would be inclined to use when describing something that could have more than one possible solution.*

Thus, not only does this item potentially conflate metaethics with the idea that there could be more than one way to comply with a moral rule, it seems to actively encourage the unintended interpretation.

3.2.3 MO3 item #3

MO3 #3 ***realism***

It is possible to compare different cultures by a single, universal standard of moral rightness.

Like item #1, this item conflates realism with universalism. Yet this item is much less appropriate as a measure of realism for the simple reason that it is *unclear* and, if literally interpreted, *is not about metaethics at all*. First, it is unclear what it means to "compare" cultures by a single standard of rightness.

This does not necessarily entail anything about whether different cultures have different moral standards, nor does it indicate whether those standards are stance-independently true.

In principle, one could simply choose to compare culture according to a single, uniform standard without that standard being “true.” Food or movie critics could develop criteria that allows them to compare different food or movies “by a single, universal standard.” Yet simply because you *can* compare different things according to a particular standard it does not make that standard *true*, much less *stance-independently true*. Even if everyone agreed to follow that standard, this would not entail stance-independence. If everyone agreed on what movies or pizza toppings were best, this also would not make it stance-independently true that they were the best, it could just mean that everyone has the same stance.

For this item to be valid, participants would have to interpret *comparison* in such a way that it refers to holding all cultures to the same standard in a vantage-point neutral way. That is, it is possible to compare all societies by a single universal standard *because there is in fact only one universal standard*. However, it could be interpreted to reflect the claim that a person or group could compare different cultures by *their* standards, which could be universal in scope. In other words, we could imagine that Alex endorses a universal standard of moral rightness, that is, a standard which applies to everyone. And Alex could then compare different cultures according to this standard to evaluate the degree to which they comply with it. Yet simply because Alex can compare different cultures according to this standard, it does not follow that this standard is stance-independently correct.

Second, whether it is possible to compare different cultures according to some standard is conceptually distinct from whether it is in fact the case that there is a stance-independently *correct* standard. After all, even an antirealist could agree that it is *possible* to compare cultures in this way, either because it is possible realism is true, or it is possible to compare cultures in this way whether or not realism is true. This item is open to a variety of such interpretations because it is ambiguous in

what sense it is “possible” to do what this statement asks. The intended interpretation is that it is possible to compare different cultures by a single universal standard because as a matter of fact *there is* such a standard.

Finally, references to a “single” standard could prompt associations with inflexibility, rigidity, and absolutism. That is, participants may associate or conflate universalism with the adherence to a moral standard that is insensitive or blind to local cultural context. This could also be reflected in a dogmatic adherence to broad and abstract moral principles, or with a generalist approach to normative morality, rather than a more particularist approach. In short, there may be a persistent threat of *normative entanglement*: the conflation between metaethical and normative considerations. In practice, it may be difficult to conceptually pull metaethical and normative apart in a way where one can present a sentence that has unambiguous metaethical content without also prompting associations with particular normative moral stances. And if those moral stances imply undesirable traits or attitudes, people may eschew response options for unintended reasons.⁴⁸

3.2.4 Concluding remarks on the MO3

All three items on the MO3 lack face validity and suffer numerous shortcomings that raise serious questions about whether they could serve as valid measures of moral realism. Items #1 and #3, in particular, are not face valid at all. Item #2 suffers numerous shortcomings in addition to simply being another incarnation of the methodologically inadequate disagreement paradigm. Given these concerns, it’s unlikely responses to these items provide reliable information about folk realism and antirealism. While some other scales fare better, they all suffer a variety of similar (and in some cases, distinct) shortcomings. They conflate unrelated distinctions with realism and antirealism, are prone to

⁴⁸ I address this possibility in greater detail in **Supplement 3**, where I discuss the role that signaling and reputation may be at play whenever items intended to convey realism are interpreted as expressions of absolutism, or are thought to convey this in addition to whatever else they may express.

be interpreted in a variety of unintended ways, present participants with a forced choice between items that don't reflect legitimate dichotomies, and so on.

3.3 Training paradigms

The primary methodological hurdle for both the disagreement paradigm and alternative paradigms is the difficulty of ensuring participants interpret questions about realism and antirealism as intended, and are given appropriate response options to express their positions. Unfortunately, conventional methods are not up to the task. Presenting people with standard survey items is insufficient, because conventional survey methods cannot adequately specify what they are asking in a way that ordinary people can reliably interpret as intended. This is because they do not provide adequate instruction and disambiguation to successfully minimize causes of unintended interpretations.

Adequately ensuring people interpret questions as intended would require extensive training and instructions to ensure participants could competently circumvent the numerous alternative interpretations that result from ambiguities, an inclination to interpret stimuli in more familiar ways (e.g., as questions about normative ethics or epistemology), and other potential factors, such as reputational concerns (e.g., endorsing realism may give the impression of being close-minded and intolerant).

One solution to this problem is to simply provide the requisite training and instruction. This is precisely what Pözlner and Wright (2020a; 2020b) have done in a series of innovative new paradigms, which I refer to as *training paradigms*. Pözlner and Wright have been at the forefront of identifying, documenting, and developing workarounds for the methodological shortcomings present in previous research on folk metaethics, and this shows in the quality and thoughtfulness that has gone into their studies. They have streamlined stimuli, improved instructions, minimized ambiguities, employed comprehension checks of the relevant metaethical distinctions, and provided converging evidence by comparing results across a diverse array of distinct paradigms. They have also

demonstrated a sophisticated understanding of the metaethical positions available to participants, and have included a wider variety of response options that reflect the array of metaethical positions previous studies failed to include. Yet the primary feature of this recent wave of studies has been the introduction of stimuli to minimize unintended interpretations by teaching participants about metaethics and inducing them to reflect on the nature of morality.

Since Pözlner & Wright's later studies correct for methodological shortcomings with Wright's (2018) earlier research, I will focus on Pözlner & Wright's training paradigms (I address Wright's 2018 article in **Supplement 3**.) Pözlner & Wright also explicitly intend their training paradigms to correct for the interpretive difficulties of earlier folk metaethics research. After describing the shortcomings of previous research, they state that "Our design starts with an explanation that purports to prevent participants from misreading its main tasks as asking for first-order moral intuitions [...] or other matters that are unrelated to the issue of moral objectivity [...]" (p. 59).⁴⁹

Their explanation consists of a description of the distinction between metaethics and normative ethics, and makes it clear that they are going to ask people about metaethics. Participants are then asked to "bracket" their normative moral standards and to "ignore these intuitions or put them to the side" (p. 59). This is not a reasonable demand to make, and it's not clear people are psychologically capable of complying with it. Participants are effectively asked to suppress their moral attitudes and values, e.g., their repugnance at murder, torture, genocide, and cruelty. I can't do this, and I doubt that most ordinary people can do so either. This brings to mind instances in which jury members are instructed to disregard information that comes up in a courtroom, even if it's incriminating or reflects negatively on the character of the defendant. If the prosecution slips and mentions inadmissible evidence that the defendant's DNA was found on the murder weapon, and the

⁴⁹ Notably, they explicitly acknowledge the possibility that participants "do not have any determinate intuitions about the existence of objective moral truths at all" (p. 58). This is a rare acknowledgment of the possibility of indeterminacy.

judge asks the jury to “disregard” this information, do we seriously expect the jury to be able to ignore this? I doubt it. And for the same reason, I doubt people can “bracket” their normative moral standards.

This is followed by presenting participants with “the least controversial and least biasing examples of normative and metaethical sentences we could think of” (p. 59). Finally, they “test and improve participants’ understanding of the normative/metaethical distinction” using a pair of comprehension checks:

- (1) *Distinction comprehension*: Participants must convey their understanding of the instructions by correctly choosing the response option, “Normative sentences express moral judgments and meta-ethical sentences make claims about the nature of morality itself” (p. 59)
- (2) *Exercise*: Participants must successfully classify a set of sentences as either normative or metaethical statements

These training exercises preceded five distinct paradigms. Each of these paradigms includes additional details and instructions distinct to that particular paradigm, which I will only discuss when relevant.

Finally, participants in all of the paradigms that they use receive detailed response options that attempt to clearly describe the relevant philosophical position. These response options also present participants with a richer variety of answers than previous studies. For instance, Pölzler and Wright (2020b) distinguish between “secular realism” and “theistic realism.”⁵⁰ Taking all of these modifications into consideration, Pölzler and Wright’s training paradigms (2018a; 2018b) represent the most comprehensive instructions to date. This includes:

- (1) *Disambiguating instructions*: Explaining what metaethics is and distinguishing it from other topics participants may conflate with metaethics (e.g., normative ethics)

⁵⁰ They describe secular realism as follows:

“When a person says that something is morally right or wrong, good or bad, etc. she intends to state a fact. Such facts exist – and they are independent from what anybody thinks about them. For example, an action that is morally wrong is wrong no matter what anyone thinks. So it would still be wrong even if you yourself, or the majority of the members of your culture, thought that it is not morally wrong.” (Pölzler and Wright, 2020b, p. 60)

- (2) *Target instructions*: Explicitly telling people the study is about metaethics
- (3) *Response option instructions*: Explicitly describing different metaethical positions participants can endorse (e.g., realism, relativism, noncognitivism, and so on)
- (4) *Disambiguating examples*: Providing examples that illustrate the distinction between metaethics and non-metaethical topics. Other studies (e.g., Wright, 2018) may also include examples that illustrate metaethical distinctions (e.g., between propositional and nonpropositional sentences)
- (5) *Comprehension checks*: questions designed to ensure participants understand the relevant concepts and distinctions presented in (1)-(4). Participants who fail comprehension checks may be excluded from analysis.
- (6) *Training exercises*: Participants may engage in various training exercises to develop competence with the relevant concepts and distinctions. Participants who fail training exercises may be excluded from analysis.
- (7) *Detailed response options*: Participants are presented with detailed response options.
- (8) *Additional comprehension checks*: Participants may be asked additional follow-up questions to assess what particular questions were about
- (9) *Open response questions*: Participants were asked to explain their answers. Participants whose responses failed to demonstrate adequate relevance may be excluded from analysis.⁵¹

Given this many criteria, Pölzler and Wright pulled out all the stops, making a concerted effort to ensure participants interpret questions as intended. Training paradigms do not have to exhibit all seven of these characteristics, or any of these traits in particular. So long as the researchers conducting a study seek to cultivate competence in metaethics, a study qualifies as a training paradigm. Studies may involve minimal training (e.g., simply providing detailed instructions, but nothing else), or go through much more extensive efforts (e.g., requiring participants to pass a college course on metaethics). There is no bright line that categorically distinguishes training paradigms from everything else. Like many concepts, it's a matter of degree.

⁵¹ Open response questions or other comprehension checks that appear after the main measures of a study are not, strictly speaking, a necessary feature of a training paradigm. However, since they can be used to evaluate the effectiveness of training, I have included them. On their own, efforts to evaluate interpretation rates or assess whether participants understand metaethical concepts does not induce a *change* in their competence with those concepts. The essential characteristic of any training paradigm is that it employs some method for improving or modifying the way participants understand metaethical concepts, typically for the purpose of enhancing the probability that they will interpret questions about metaethics as intended.

Finally, note that training paradigms are consistent with a variety of measures that follow the training portions of the paradigm (and that precede comprehension checks or other measures that follow the main measures). In this case, Pölzler and Wright employed modified versions of the disagreement paradigm, along with several novel paradigms, including a “theory task,” a “metaphor task,” and a “truth-aptness” task (I discuss these in **Supplement 3**).

The question is: were Pölzler and Wright’s efforts successful? As you’ve probably already guessed, I believe the answer is a decisive, if regrettable, *no*. There are three main reasons why training paradigms fall short. The first two are serious, but survivable. However, the third reason is fatal to all training paradigms:

- (1) *Inadequacy*: There are good reasons to believe that existing training paradigms fail to ensure adequate reflection, competence, and emphasis on semantic considerations
- (2) *Irrelevance*: Competence in the relevant distinctions may have little or no significant causal influence on measures of the participant’s metaethical stances/commitments
- (3) *Spontaneous theorizing & philosophical induction*: Successful training yields a sample whose responses may have been formed by the training, rather than reflecting their metaethical stances/commitments prior to participant in the study.

I discuss problems (1) and (2) in **Supplement 3**. Although both present such serious challenges to existing paradigms that each independently undermine the validity of existing studies, space constraints limit me to discussing (3), *spontaneous theorizing* and *philosophical induction*, because they present *insurmountable* obstacles to the validity of training paradigms.

3.3.1 Spontaneous theorizing

Suppose for the sake of argument that training paradigms successfully enable participants to interpret questions about metaethics as intended.⁵² The goal of research on folk metaethics is to discover the

⁵² If the need for participants to *reflect* on their reactions is also considered a necessary condition, then we can assume this task succeeds as well. This *is* a requirement that Pölzler and Wright agree to for their purposes (see Pölzler & Wright, 2020a) since they accept Kauppinen’s account of the necessary conditions for philosophically relevant measures (see Kauppinen, 2007). It may not be necessary for some research goals, however. For instance, Pölzler (2018a) points out that the failure of Goodwin and Darley’s (2008) studies to meet Kauppinen’s criteria “should not be understood as a criticism of Goodwin and Darley. After all, they did not aim at providing semantically relevant evidence in the first place. Their

metaethical stances/commitments of *ordinary people*, i.e., people *without* philosophically significant training. How could we possibly achieve this goal by recruiting ordinary people and then training them in philosophy? Granted, we might suppose that they haven't received *significant* training. But there is an important difference between those who receive (a) no training, (b) some training, and (c) a lot of training.

Suppose I am correct about folk metaethical indeterminacy: that in the absence of formal philosophical training, people have no determinate metaethical stances or commitments. And suppose that, like the various interpretations of quantum mechanics, the way that ordinary people speak and think about morality does not contain any implicit commitment to realism or antirealism. If so, any attempt to measure how ordinary people think that *changes* their psychology by *causing* them to have determinate stances or commitments would not be a measure of how ordinary people speak and think outside the experimental context; it would be a measure of how they speak and think once they are placed in an experimental context that, by design, *induces* them to have determinate stances and commitments. And in the absence of evidence to the contrary, we cannot know whether the measures obtained via training paradigms represent what the participant thought prior to participating in the study, or whether their pattern of responses were caused by the study itself. Thus, we could not justifiably use the results of such studies as evidence of how “ordinary people” think about metaethics unless we could be sure that the stimuli employed in our studies wasn't causing them to develop or at least superficially appear to express determinate metaethical stances and commitments.

Unfortunately, there is no feasible way to do this. Here's why: on the one hand, we could employ conventional, non-training measures, then assess how participants interpreted questions after the fact to test whether they interpreted questions as intended. However, this is exactly what I've done

interest was rather with illuminating the prevalence and causes of ordinary people's intuitions about the moral objectivism/subjectivism distinction” (p. 4925).

(see **Chapter 4** and **Chapter 5**). The results of these studies reveal that people consistently fail to interpret questions as intended due to ambiguities, conflation, and other factors that render ordinary people incapable of reliably interpreting conventional questions about metaethics (i.e., questions that lack disambiguating instructions, training, and so on) as intended.^{53,54}

On the other hand, we could provide precisely those clarifications and disambiguations that form one of the foundations of training paradigms. Yet the very act of providing such clarifications is precisely what could *cause* participants, who are supposed to be “ordinary people,” to think about the issues at hand in ways that they hadn’t prior to participating in the study by causing them to reflect on metaethical concepts, develop competence in those concepts and distinctions that appear in measures, and so on. If this occurs, we would be unable to distinguish responses that reflect how that person thought *prior* to participating in the study from responses that reflect how the participant thinks *as a result of participating in the study*. I refer to this phenomenon as *spontaneous theorizing*. Spontaneous theorizing occurs *whenever a participant who held no determinate stance or commitment prior to participating in a study is induced to develop or express a stance or commitment due to the experience of participating in the study itself*.

For example, as Gill (2009) observes, most ordinary people have probably not considered whether mathematical Platonism is true. That is, they have not given much thought to whether our mathematical discourse refers to abstract objects that exist independent of how we think, speak, or act (Linnebo, 2018). And it is doubtful that everyday mathematical thought and language commits people to a particular account. Yet if we gave people a survey that required them to express agreement with either Platonism or anti-Platonism, we’d *necessarily* get *some* pattern of responses that would

⁵³ Even if participants seem to have interpreted questions as intended, such responses could in principle be confabulatory or have failed to reflect the actual reasons why participants answered as they did. However, this would be true for a wide variety of research and represents a more general methodological concern. Of course, that doesn’t mean it isn’t a legitimate concern. In this case it’s moot since people don’t typically interpret standard metaethics paradigms as intended anyway.

⁵⁴ It’s worth noting that even before these studies were conducted, Pölzler and Wright had already outlined enough methodological worries that they felt the need to devise training paradigms.

suggest *everyone* endorsed either Platonism or anti-Platonism. Even if participants held no determinate stance or commitment prior to the study, the very fact that they are now prompted to do so may cause them to reflect on and form such a view. That is, even if ordinary people have no determinate stance or commitment with respect to a particular philosophical distinction, the very act of asking participants may *cause them to develop a determinate stance*. Any study that *causes* participants to hold a particular view cannot be used as evidence that they already held that view, and it would be extremely difficult to empirically demonstrate whether the study itself caused the view or whether they held the view prior to participating in the study. This leaves any efforts to evaluate folk metaethics in a catch-22: *either* we don't provide training, in which case our measures aren't valid because participants don't interpret questions about metaethics as intended, or we do provide training, in which case we can no longer be sure that responses aren't the result of *spontaneous theorizing*.

3.3.2 Philosophical induction

Spontaneous theorizing is not even the most serious, or only, problem with training paradigms. There is a second, even more damning catch-22: if your attempt to train participants fails then you won't have a valid measure of their metaethical stances or commitments. However, if you succeed, then your participants are no longer ordinary people, and thus no longer members of the sample population they're supposed to represent. This is because adequately training participants to understand the relevant philosophical concepts and distinctions necessarily involve inducting them into the categories and distinctions *as they are understood in academic philosophy*, and thus requires them to think *within the framework of contemporary analytic philosophy*.⁵⁵ In other words, training paradigms succeed only insofar as

⁵⁵ and more specifically in accordance with mainstream metaphilosophical presuppositions implicit in the structure and framing of the metaethical concepts and distinctions presented in instructions and exercises. I develop on this concern in **Supplement 3**.

they succeed at *philosophical induction*, i.e., *inducing participants to become philosophers*, if poorly trained novice philosophers.

Of course, people who have just spent a few minutes of training are very different from academic philosophers who've spent years studying philosophical topics. But this is not a problem for my account: is this training good enough, or not? If it's not good enough, then we're back to square one. If it is good enough, then what does that entail other than that these people have crossed the threshold to be inducted into the hallowed halls of "philosophers"? Even if we wish to reserve this moniker for people who have passed an even greater threshold, that is still not a problem. We might think of participants who successfully pass the tasks presented in training paradigms not as philosophers *or* ordinary people, but something in between: *quasi-philosophers*. Training paradigms don't fail only if participants are philosophers, all that's necessary is that they are no longer ordinary people. If participants in these studies are something in between, then we still face a serious problem: the responses of quasi-philosophers may systematically differ from ordinary people (and philosophers, for that matter). And since we have no extraneous way to measure the responses of ordinary people as a point of comparison, we have no way to know how significant this deviation is. Measuring the responses of quasi-philosophers is thus never going to serve as an appropriate method for drawing inferences about ordinary people. What we'd be left with is a set of measures for a sample of people who were so changed by the study they participated in that our findings would lack any external validity. If the only way to measure someone's views is to teach them philosophy, and an ordinary person is, by definition, someone who hasn't been taught philosophy, then ordinary people are a population that is inaccessible to folk metaethics research. After all, you can't study the psychology of "people who have never seen *The Matrix*" by having people who hadn't seen *The Matrix* before they came into the lab come in, watch it, then answer questions about it.

This pair of catch-22s demonstrate that training paradigms were doomed from the start. The training paradigm requires researchers to causally interact with the participants in a way that (a) may cause relevant psychological changes in the sample and (b) *requires* changing the sample in such a way that the sample no longer represents any meaningful population outside the context of the study (and thus strips the paradigm of external validity). Pölzler and Wright correctly recognize that previous research on folk methods suffers fatal methodological shortcomings. Unfortunately, their cure is worse than the disease.

3.4 Conclusion

I have raised a variety of methodological worries about the current state of research on folk metaethics. These are not trivial concerns that can be cast aside and ignored. The current direction researchers in folk metaethics have taken suggests that they are inclined to agree. At least some researchers have introduced a variety of methods to circumvent the shortcomings of earlier work. Unfortunately, these efforts don't succeed, and indeed, *can't* succeed, for the reasons I've just outlined.

Yet so far my critiques have remained purely theoretical. I have yet to present any empirical work that would support these criticisms or support indeterminacy. That is the goal of the remaining two chapters. In chapters 4 and 5, I embark on an effort to demonstrate that most people do not clearly interpret questions about metaethics as intended. These rates are high enough to cast serious doubt on the validity of existing folk metaethics research. In addition, I present the results of a study that reveals how easy it is to conduct a study that could give the misleading impression that I've identified genuine patterns of folk philosophical thought, even when no such patterns plausibly exist. This highlights the possibility that research on folk metaethics is a byproduct of the forced choice design employed by most research.

CHAPTER 4: **Open Response Paradigms**

4.0 Introduction

In the preceding chapters, I argued that existing research on the psychology of metaethics is too riddled with methodological shortcomings to provide compelling evidence that ordinary people hold determinate metaethical stances or commitments. While I am confident these criticisms are sufficient to justify this conclusion, I have yet to provide any empirical evidence that would support folk indeterminacy. The remaining chapters seek to rectify this.

In this chapter, I assess the results of open response data. This data consists of written responses that reveal how participants interpreted questions about metaethics. The goal of these studies is to determine the proportion of participants who interpreted questions as researchers intended, i.e., as questions about moral realism or antirealism. If most participants do interpret these questions as intended, this would provide significant evidence that these studies are valid, and that people's responses to these questions genuinely reflect their metaethical stances or commitments. Conversely, if most people don't interpret questions about moral realism and antirealism as researchers intend, that would suggest that these studies are not valid.

Even if low rates of intended interpretation do not decisively undermine the validity of existing paradigms, widespread failure to interpret questions as intended calls for an explanation. While there are many plausible reasons why so many participants would appear to interpret questions about metaethics in unintended ways, indeterminacy offers a straightforward explanation for low intended interpretation rates: people do not understand the questions because they have no metaethical stances or commitments with respect to realism and antirealism. Thus, this chapter serves both to reinforce the case made in previous chapters that existing metaethics paradigms are invalid, and to offer empirical evidence of metaethical indeterminacy.

4.1 Rationale for using open response data

Given the methodological limitations of qualitative research (Guerin, Leugi, & Thain, 2018; Willig, 2008), coupled with its poor reputation among more quantitatively oriented social scientists (Kvale, 1994; Povee & Roberts 2014; Queirós, Faria, & Almeida, 2017; Rahman, 2016) it is reasonable to question an extensive reliance on open response data. After all, my conclusions are largely based on coding items based on my own judgment. This may introduce an unacceptable degree of bias and subjectivity into my results (e.g., confirmation bias, Nickerson, 1998). In addition, I am not blind to my own hypotheses, and awareness of my vulnerability to cognitive biases may do little to mitigate their impact (Aczel et al., 2015; Compen et al., 2022, Fischhoff, 1982; Welsch, Begg, & Bratvold, 2007; cf. Savitsky & Gilovich, 2003). Qualitative data can also be vague. Researchers may be hesitant to draw firm conclusions based on the rough impressions they or anyone else has of a set of written responses. In short, qualitative data may seem too error-prone and crude to be worth serious consideration.

These concerns are not without merit. Concerns about my findings are especially plausible since my results are likely to come as a shock. When I initially set out to assess open response data, I expected many participants to interpret questions about metaethics in unintended ways. But I never expected the rates to be so low. Across all studies, *most* participants did not interpret questions about metaethics as intended. In some cases, the percentage of participants who demonstrated a clear understanding of what they were asked is in the single digits. I am confident these findings challenge anyone convinced that ordinary people have determinate metaethical standards, yet skeptical readers may believe my assessment is too uncharitable, and that I have grossly underestimated the degree to which ordinary people correctly interpreted questions about metaethics.

I readily acknowledge this concern. Perhaps my method of coding is flawed, or there is some other reason why these findings tell us little about what, if any, metaethical stances and commitments characterize folk morality. Yet I believe the data will speak for itself. I have made all of the data readily

accessible, and invite readers to examine it. While I cannot guarantee that everyone will agree with how I coded each response (in fact, I am sure they won't), I am confident that few people familiar with the relevant metaethical distinctions will come away with the impression that there is little cause for concern.

Of course, confidence in my conclusion hardly justifies relying so extensively on qualitative data. Why, then, have I devoted an entire chapter to these results? First, the data I present in this chapter is not intended to provide an independent and decisive case for or against any particular hypothesis. Rather, my findings primarily function to check the validity of paradigms that are themselves subject to quantitative analysis. Insofar as my findings have additional implications, they should be taken only as suggestive evidence to be considered in conjunction with theoretical considerations and other sources of evidence.⁵⁶ Whatever its limitations, qualitative data can be used alongside quantitative data to *triangulate* on and potentially reach the same conclusions (Morses, 1991; Olson, Haralambos, & Holborn, 2004). When this occurs, quantitative and qualitative data can work in tandem to provide mutually corroborating evidence for the same conclusions. As such, I do not propose to supplant the rigor of quantitative data, but to complement it.

Second, my central argument is that most people consistently fail to interpret questions about metaethics as intended. Analysis of open response data can provide evidence for or against this hypothesis, whatever its limitations may be. And if I am correct, this is not a trivial concern. That participants understand what they are being asked is a *necessary* condition for validity. After all, if people do not understand what they are being asked, their responses will not reflect the psychological

⁵⁶ Thus, these findings serve an auxiliary role alongside other sources of evidence. I do not wish to understate the relevance of these findings, however. In principle, findings across the social sciences should converge on the same conclusions if those conclusions are genuinely robust. We should expect findings in fields as disparate in their methodology as history, anthropology, psychology, and neuroscience to converge or at least be mutually intelligible and consistent with one another. If anthropological data overwhelmingly suggested that the vast majority of people were moral realists, but the methods used by social psychologists and experimental philosophers indicated that most people were moral antirealists, this would call for an explanation.

phenomena researchers intend to measure. If open response data could provide even modest evidence that a paradigm or an entire body of empirical research were predicated on invalid measurement tools, such findings could have considerable value in directing subsequent efforts to further evaluate the validity of these measures.

Third, I rely on a large and diverse (though not demographically diverse) body of data, and conduct some quantitative analysis on the data itself, insofar as I estimate proportions of intended and unintended interpretations. This circumvents two common shortcomings with qualitative data: that findings are often based on comparatively small datasets (such as a handful of interviews) and the lack of any form of quantitative analysis. Lastly, I present a series of studies in **Chapter 5** which do take a quantitative approach to assessing how participants interpret metaethics stimuli.

With these caveats in mind, I concede that open response data may prove unhelpful in many cases. However, its use is warranted when there is theoretical justification for suspecting extremely low rates of intended interpretation for quantitative research. In such circumstances, the most efficient path to assessing validity isn't *necessarily* to conduct more quantitative studies. We could imagine a scenario where, for instance, the reason for a reliably recurring pattern in a given dataset is due to an ambiguity in how people interpret a question that researchers didn't notice. Simply asking people how they interpret the question could potentially reveal the presence of that ambiguity more quickly, easily, and decisively than running additional quantitative studies without attempting to directly pinpoint the problem. Indeed, inattention to interpretation may be invisible to many exclusively quantitative approaches, since such approaches can reliably yield data with psychometric properties that don't raise suspicion. Thus, I don't employ qualitative methods as a *replacement* for quantitative methods, but as an *auxiliary* means of assessing the validity of particular quantitative methods, methods which remain the primary way to address empirical questions.

In most cases, there may be little reason to worry that a significant proportion of participants would wildly misconstrue the nature of a task. Competent native speakers of a language typically have little trouble understanding questions or following instructions that are simple, straightforward, and deal with familiar phenomena. When major concerns arise, researchers are (hopefully) able to spot them, and when they overlook potential problems, these problems are (at least under ideal conditions) caught during the peer review process or after publication. In addition, researchers can (and frequently do) run their designs by colleagues, present them in lab meetings, or conduct pretests that can catch any serious issues with the interpretability of their stimuli. And where we recognize the potential for confusion or misunderstanding, we often provide clarificatory instructions or include comprehension checks that assess whether participants interpreted features of a study as intended, which can improve validity (Meisters, Hoffman, & Musch, 2020). In short, the way psychologists conduct research already builds checks against unintended interpretations into the process, and where these fail, the self-correcting nature of science will eventually reveal and allow us to correct for interpretative deficiencies.⁵⁷

These platitudes represent an overly optimistic picture of the social sciences built more on its presumptive ideals than how it actually functions. It is possible that in many cases the aforementioned methods aren't employed or are inadequate even when they are widely used, e.g., the efficacy of peer review remains a matter of serious contention (Goldbeck-Wood, 1999; Justice et al., 1998; Kelly, Sadeghieh, & Adeli, 2014; Mavrogenis, Quaile, & Scarlat, 2020; Smith, 2006; 2010; Tennant & Ross-Hellauer, 2020). There is also an irony in treating quantitative methods as the *sine qua non* of research when the validity of quantitative approaches often relies on the informal procedures (e.g., lab

⁵⁷ Even so, few of these concerns are self-consciously designed with the goal of ensuring proper comprehension. While standard procedures may mitigate concerns about interpretation to some limited extent, they aren't typically *designed* to do so. Quite the contrary. There is a remarkably cavalier attitude among social scientists, at least within social psychology and, astonishingly, experimental philosophy, with ensuring adequate comprehension on the part of participants. Simply put, researchers rarely bother to check whether participants interpret questions as intended.

meetings) and the subjective, *qualitative* impressions of researchers, who frequently employ intuitions, hunches, and subjective experience in lieu of formal methods for establishing the validity of their measures.

In many ways, quantitative methods serve as the public face for methods that, when we pull back the curtain, often depend on non-quantitative and informal procedures that draw on the subjective judgments of researchers. Indeed, in many (or perhaps even most) cases researchers don't use quantitative approaches to validate their measures, or at least don't report doing so. Instead, they frequently employ measures developed "on the fly," without addressing *any* formal steps relevant to validation.⁵⁸ Even though formal validation procedures are available for e.g., scales, a recent assessment of the scales reported in articles published in the *Journal of Personality and Social Psychology (JPSP)* in 2014 found that:

Roughly, half of these scales (46%) included no reference to previous validation, appearing to have been developed on the fly. α was the only psychometric information reported for half of these scales which had no previously published validity evidence, and 19% had no accompanying psychometric information. (Flake, Pek, & Hehman, p. 5)

These findings suggest that even researchers publishing in premiere journals frequently ignore the validity of their measures. This is not a minor concern. As Flake and Fried (2020) observe, lack of transparency "prevents the evaluation of all aspects of a study's validity: its internal, external, statistical-conclusion, and construct validity" (p. 457). They add that "recent research on commonly used measures in social and personality psychology showed that measures with less published validity evidence were less likely to show strong evidence for construct validity when evaluated in new data" (p. 457). Lack of validity evidence thus plausibly serves to conceal poor validity and enhance opportunities for researchers to (inadvertently or otherwise) engage in *questionable measurement practices*

⁵⁸ And I haven't even begun to assess the role subjectivity plays in researcher degrees of freedom, which have allowed p-hacking to persist for so long (Wicherts et al., 2016).

(QMPs), i.e., “decisions researchers make that raise doubts about the validity of the measures used in a study, and ultimately the validity of the final conclusion” (p. 458). An uncritical reliance on exclusively quantitative methods, coupled with a culture that places little emphasis on validity, does not put us in a strong position to reject qualitative and mixed methods approaches as soft or inadequate, especially when the very purpose of those methods is to evaluate the validity of the measures used in quantitative research, measures that are routinely devised using informal and non-quantitative procedures.

If so, researchers may have overlooked serious methodological shortcomings, such as high levels of interpretative variation and unintended interpretations using particular paradigms, or with respect to particular types of questions. Psychology is in the midst of a replication crisis (Nelson, Simmons, & Simonson, 2018; Open Science Collaboration, 2015; Wiggins & Christopherson, 2019), and researchers continue to grapple with numerous methodological shortcomings related to design and data analysis; it is not outlandish to propose that there may be yet another one: a widespread failure to ensure participants interpret stimuli as intended, an instance of a broader but oft-ignored problem crisis of measurement and validity (Flake & Fried, 2020; Lilienfeld & Strother, 2020; Schimmack, 2021).

Such concerns may be irrelevant to a great deal of research that presents participants with comparatively simple, straightforward, and familiar considerations. Unfortunately for researchers who study the psychology of metaethics, there is nothing simple, straightforward, or familiar about moral realism and antirealism. The previous chapters provide more than ample reason to conclude people struggle to think in metaethical terms. Indeed, even researchers themselves struggle to clearly and consistently describe the metaethical landscape, or to devise paradigms that adequately operationalize metaethical concepts in their stimuli.

As argued in Bush and Moss (2020), this is because metaethics is unlike many other subjects social scientists present in surveys, and its differences render it uniquely vulnerable to low rates of intended interpretation. There are four main factors that contribute to an increased likelihood that participants will interpret questions about metaethics in unintended ways:

(i) the relevant metaethical theories are complex and difficult to grasp

(ii) most people are unfamiliar with these distinctions prior to encountering them in studies

(iii) metaethical theories are generally abstract and distant from real world practical questions lay populations would be more familiar with and expect to be asked about

(iv) there are typically plausible non-metaethical interpretations of the questions posed to respondents, and participants may be more likely to interpret surveys as posing these more prosaic questions rather than abstruse metaethical questions (p. 7)

Given these characteristics, when we compare research on folk metaethics to conventional psychological research, it might come as more of a surprise if most participants *did* interpret questions about metaethics as intended. Although I do not consider any of these factors especially controversial, critics may object that moral realism is not difficult to grasp and that people are familiar with it. David and I discuss why we do not believe this is likely in Bush and Moss (2020).

4.2 General methods and procedures

4.2.1 General methods

The studies presented in this chapter all share a common structure. In some cases, participants were first asked a question about their metaethical stance. They were then asked to explain *why* they answered this question the way they did or *how* they interpreted one or more of the questions they answered. In other cases, they were presented with a statement, and were then asked to explain what that statement means in their own words or to explain how they interpreted the statement. These questions were either presented in a survey alongside other questions, or as standalone questions (i.e.,

the entire study simply consisted of asking such questions). In all cases, participants were presented with a textbox and wrote out their responses, rather than using Likert scales or selecting from among a set of multiple choice options. All responses were collected in English, and were uncorrected—any typos, misspellings, poor grammar, or uninterpretable remarks were left as they were.

All studies consist of short and straightforward questions, and the typical responses reflect this simplicity, since most consist of a few words or at most a sentence or two. Note that only the *structure* of these questions was simple; since the content itself concerns metaethics, it remains a daunting and complicated task to adequately interpret what these questions are asking, or at least, that is what I hope to show in analyzing the results. In some cases, participants opted to write more than this, but this is atypical. Thus, most of the data consists of short written responses, in which participants explained their answer to a question or offered an interpretation of a question or statement. Here is a sample question and response:

Question: *In your own words, what does it mean to say that the truth of the moral claim "murder is morally wrong" is objective?*

Response: *You are saying it is an objective fact murder is wrong. An objective fact is one which is not based on human judgement or belief.*⁵⁹

A few datasets were collected by other researchers and were included in previously published articles. Wherever I analyze such data, I will make an explicit reference to whose data I am assessing, and which published articles are associated with that dataset. However, most of the datasets were collected by myself, either alone or in conjunction with collaborators. Unless otherwise specified, any data presented is my own and all analyses were my own.

To my knowledge, no previous studies have systematically analyzed open response data with the goal of assessing the proportion of participants who interpreted questions about metaethics as

⁵⁹ This is an excellent and clear response.

intended. While some researchers have included open response data in their studies (e.g. Goodwin & Darley, 2008; Sousa et al., 2021; Wainryb et al., 2014; Wright, Grandjean, & McWhite, 2013), and do use it to check interpretation rates, such data has rarely featured prominently in research results.^{60,61}

4.2.2 General procedures

All studies were analyzed in two ways. First, *interpretation rates* for each study were coded using the same method. This is the primary method of analysis and the central focus of this chapter. My goal was to determine the proportion of participants that interpreted questions (and other stimuli) used in research on folk metaethics as researchers intended. David Moss has a background in philosophy and social science. We have collaborated on projects in metaethics and folk metaethics and he has published independent research on methodological problems in folk metaethics research in the past (Moss, 2017). Given our extensive interactions over the course of my research, I consider David an extraordinarily competent coder who is familiar both with the relevant metaethical distinctions and is highly sensitive to the methodological concerns motivating this project, e.g., ambiguity, conflation, and the role pragmatics can play in influencing how people interpret stimuli.

Second, I developed a set of themes to provide richer qualitative assessment of the content of people's responses. My approach does not rigidly adhere to any formal methodology, and as such could be considered a generic qualitative approach (Kahlke, 2014). However, my approach most closely reflects *thematic analysis* that is similar in many ways to the *reflexive thematic analysis* approach pioneered by Braun and Clarke (2006; 2014; 2019; 2020; Clarke & Braun, 2013). This involved an

⁶⁰ One notable exception is Wainryb et al.'s (2004) extensive analysis of open response data designed to assess how children interpreted questions about metaethics. Unfortunately, this data was not available for me to assess directly, so I had to rely on the analysis presented in the article. As I discuss in **Supplement 2**, virtually all of the participants in these studies appeared to interpret questions about metaethics in unintended ways. This is consistent with my findings, but since I cannot assess the data myself, this conclusion should be interpreted with caution.

⁶¹ Even when researchers do appeal to open response data, they rarely do so in a systematic and thorough way. For instance, Sousa et al. (2021) appeal to open response data to support their explanation of their findings. However, they only highlight one or two examples to support a point, rather than appealing to systematic analysis of the data they collected.

iterative process of familiarizing myself with the general patterns and themes that emerged in each set of data, reviewing, refining, and naming those themes, then reviewing the data again in light of these themes to code each item in accordance with one of the named themes.

This type of analysis plays only a secondary role in the findings reported here. While thematic analysis of the particular ways participants interpreted questions about metaethics in unintended ways is both fascinating and worthy of further study, my primary concern is with the overall rate of intended and unintended interpretation. As such, I have relegated discussion of this data to **Supplement 4**. Here, my focus will focus on interpretation rates.

4.2.3 Coding and analysis for interpretation rates

All written responses were coded using a two-dimensional quaternary system. First, all items were coded based on whether they appeared to reflect the intended metaethical distinction (1) or not (0). Second, they were coded based on whether the response was clear (1) or unclear (0). This created four possible ways an item could be coded:

- 1 | 1 = Clearly intended interpretation
- 0 | 1 = Clearly unintended interpretation
- 1 | 0 = Unclear (leans towards intended interpretation)
- 0 | 0 = Unclear (leans towards unintended interpretation)^{62,63}

Judgments about how to code items in accordance with this system developed over the course of coding, and rely as much (or more) on experience and familiarity with the data that would be difficult to articulate as they do on general and definable characteristics that could be readily summarized. Nevertheless, I can state roughly how judgments were made. Items coded as clearly intended interpretations (1 | 1) needed to plausibly express a view, however inchoate or clumsily phrased, that

⁶² Participants who did not respond were coded as 0 | 0.

⁶³ Early versions of this coding scheme were initially proposed by David Moss.

seems consistent with the intended interpretation of the stimuli. There was no expectation of a sophisticated or detailed response. For instance, the validity of the disagreement paradigm depends on participants interpreting disagreements to result from genuine disagreement in moral values (rather than e.g., disagreements over nonmoral considerations). To highlight just how simple a clearly intended interpretation could be, the response “Different standards of morality” was coded as a clearly intended interpretation. I did my best to extend this same low bar for a clearly intended interpretation across all datasets.

In contrast, another participant proposed that the “Other individual misread the question.” This was coded as a clearly *unintended* interpretation (0 | 1). Clearly unintended interpretations represent any instance in which the participant expressed a coherent interpretation of the stimuli, but this interpretation was inconsistent with the intended interpretation. This could be due to conflating some other consideration with a metaethical one (e.g., interpreting the question to be asking about their first-order normative judgments), not accepting the information presented in the study as stipulated (e.g., if they are told another person judged that murder or stealing were acceptable, they propose that the other person misunderstood the question), understanding specific terminology in a way inconsistent with researcher intent (e.g. understanding the term “objective” to mean *unbiased*), or invoking metaethical concepts, but expressing an understanding of them that is inconsistent with the relevant metaethical distinction (e.g. stating that morality “is objective,” then adding that objectivism means that moral standards “apply to everyone” or are “clear and well-defined.”). An unintended interpretation is *not* a “misinterpretation.” Misinterpretation implies fault or error on the part of participants. Yet as much or more of the problem lies in the inherent ambiguity and underspecificity in the way questions were phrased, and in some cases, due to researcher error (e.g., researchers employed stimuli that formally conflated realism/antirealism with other concepts or distinctions).

The distinction between unclear intended and unintended interpretations can be a bit murkier. Roughly, unclear intended interpretations could be consistent with researcher intent, but are sufficiently unclear, internally inconsistent, or ambiguous to undermine confidence about how the participant interpreted stimuli. Unclear unintended interpretations include several types of response that are easy to classify: instances where the participant did not respond or did not offer a substantive response (e.g., writing “n/a” or “.”), incoherent or uninterpretable remarks, responses that clearly ignored the task (e.g., directing a remark towards researchers, such as “this study sucks!”), or were otherwise irrelevant or lacking in substance (e.g., “MORALITY”). Finally, some instances of unclear unintended interpretations leaned in the direction of suggesting an unintended interpretation, but were too confusing, vague, or underdeveloped to justify confident classification. These general remarks about how responses were coded are far from adequate to capture the full range of considerations behind coding decisions, and discussion of individual items is often the best way to highlight how and why particular decisions were made. I provide a more extensive discussion of the coding procedure, along with examples, in **Supplement 4**. With the exception of Study 1A, all of the coding reported here and in the supplements was conducted exclusively by myself.

4.3 General predictions

I had no initial predictions about the precise proportion of participants I expected to fall within each category. Since Study 1A was the first dataset that I analyzed, I didn’t know what to expect, so there was no particular prediction made in advance for those findings. However, the incredibly low interpretation rate initially obtained for that study (see Bush & Moss, 2020 and **Supplement 4**) led me to anticipate that fewer than half of participants would interpret questions as intended, regardless of the measure used. Since no previous research adequately estimated interpretation rates for the folk metaethics research, I had no previous data to rely on in offering precise estimations. And given the

wide variation in interpretation rates across studies, it would be difficult to make precise predictions about any particular measure.

However, the main reason I didn't set out with precise predictions is that it would be inappropriate to do so. Coders (such as myself) were not blind to expectations about interpretation rates. To code data with precise prediction in mind would risk biasing coding to conform to expectations. Thus, while I offer a nominal prediction that fewer than half of participants will interpret metaethics stimuli in a clearly intended way, due to the non-blinded nature of the coding it would be inappropriate to treat these anticipated interpretation rates as true "predictions." Since I am coding the data myself, there would never be a way to eliminate the risk that I am biased in such a way that I code to conform to a given "prediction." As such, statistical tests should be interpreted as *post hoc* and effectively exploratory, not reflections of an unbiased assessment of the data, and definitely not as predictions in the absence of observing the data, since this wasn't possible.

This is obviously less than ideal. To anticipate one criticism with this data: it's unclear whether it's feasible in practice to obtain blind coders with sufficient competence to adequately code the data. Insofar as this limits the inferential power of my analyses, this shortcoming is simply unavoidable. While one might suppose that blind coders would be better than nothing, I am not convinced this is the case. *Systematic* bias may result in interpretation rates that I'm not confident would accurately reflect the true proportion of clear interpretations. Nevertheless, future efforts can and should be made to train blind coders. The only other ways to shore up concerns about bias would require corroborating evidence, perhaps from a combination of adversarial coders conducting similar research and the use of conventional quantitative methods where genuine predictions about interpretation rates can be tested without concern. I take the latter approach in **Chapter 5**.

4.4 Summary of studies

Studies 1A-1C evaluate whether people presented with the disagreement paradigm interpret the source of disagreement as intended (i.e., as a difference in fundamental moral beliefs).

Studies 2A-2D present participants with a question and a response from a previous participant. Interpretation rates focus on how participants interpret the previous participant's *response* rather than the question the previous participant was asked. This allows us to artificially construct a clear articulation of an intended response to the initial question, with the participant's task merely serving to interpret a statement reflecting a metaethical position.

Studies 3A-3C explore interpretation rates in the context of a concrete moral decision related to charitable giving. A noncognitivism condition was added to broaden the range of metaethical views reflected in the questions.

Studies 4A and 4B evaluate interpretation rates for questions that explicitly ask about the participant's metaethical views using standard metaethical terminology, e.g., “objective” and “relative.”

Studies 5, 6, and 7 evaluate interpretation rates for the questions that appear on scales used to measure folk metaethical views. Study 5 assesses interpretation rates for the Moral Relativism Scale (MRS, Collier-Spruel et al., 2019). Study 6 assesses interpretation rates for scale items employed by Yilmaz and Bahçekapılı (2015a; 2015b; 2018). Finally, Study 7 addresses the items used in the FMO (Zijlstra, 2019). All three studies employ two distinct measures: asking participants *why* they answered the question the way they did, and asking them to *explain* what they think the question means. This provides a converging pair of measures of interpretation rate for each item.

4.5 Studies 1A & 1B: Source of disagreement in Goodwin & Darley (2008)

Methods

Studies 1A and 1B are a reanalysis of open response data collected by Goodwin and Darley (G&D, 2008). G&D gathered open response data for two of the three experiments they conducted,

Experiment 1 and Experiment 2. The goal of these studies was to evaluate whether ordinary people are moral realists or antirealists.⁶⁴ Although both studies used the disagreement paradigm, there are additional details to this study that I will describe to put my reanalysis of the open response data in context.⁶⁵

All participants rated their level of agreement with a set of 26 moral and nonmoral statements on a six-point Likert scale (1 = *strongly disagree*, 6 = *strongly agree*). Nonmoral statements were drawn from three additional categories: social convention, aesthetics, and “factual” statements (e.g., statements about science or geography). For each statement, they were then asked whether it was a:

- *True statement.*
- *False statement.*
- *An opinion or attitude.*⁶⁶

This was followed by having the participant perform an unrelated task. The goal of this phase was to give researchers time to select five responses (2 moral, and 1 social convention, aesthetic, and factual question). Once these items were selected, the study resumed, and participants were told that participants from a previous study using the same stimuli disagreed with each of these statements.

They were then asked to consider this disagreement and select one of the following responses:

- *The other person is surely mistaken.*
- *It is possible that neither you nor the other person is mistaken*
- *It could be that you are mistaken, and the other person is correct*
- *Other*

⁶⁴ They do not use this terminology. While they explicitly state that their goal is to distinguish a stance-independent conception of morality from a stance-dependent conception of morality, they refer to the former as “objective” and use a range of different terms to refer to the latter, including “non-objective,” “subjective,” “opinion-based,” and “preferential.” It is hard to fault the lack of precise and consistent terminology here, given that it is difficult to characterize a range of conceptually distinct non-objective views in any simple and straightforward way. While I believe “antirealist” is adequate, a reasonable person might have qualms with this terminology.

⁶⁵ Both studies also used an additional paradigm to generate a combined measure of folk metaethical belief. However, the additional measures are not relevant to this analysis. See Goodwin & Darley (2008) for details.

⁶⁶ This is not a valid measure of moral realism/antirealism. This question conflates realism/antirealism with cognitivism/noncognitivism (Pölzler, 2018b). It is also highly vulnerable to be interpreted as a question about normative ethics. It is also consistent with asserting something true or false that it is an “opinion or attitude.” Indeed, some uses of “opinion” are cognitivistic. As such, the “An opinion or attitude” option is not clearly mutually exclusive with the other response options.

The first option was interpreted as a realist response, while the second and third were interpreted as antirealist responses. Later in the survey, participants were asked to explain what they thought the source of the disagreement between themselves and the previous participant could be. Specifically, they were asked:

*Give us your thoughts about why it is that there is disagreement. What could be its source?*⁶⁷

Goodwin and Darley included this question because they were concerned that some participants might interpret the source of the disagreement in ways that would be inconsistent with the interpretation necessary for their response to reflect a metaethical stance or commitment.⁶⁸ It will come as no surprise that I agree with their concerns. This is a laudable step to ensure that participants interpret the question as intended, and more researchers should employ questions like this. They also report reassuring results:

When individuals did not interpret the disagreement in a bona fide way [...] we excluded such data from all foregoing analyses. In fact, only seven out of a total of 102 responses were excluded on these grounds.

If their analysis of participant explanations is correct, this would indicate that less than 7% of participants interpreted the disagreement in a way that would threaten the validity of the disagreement paradigm. If so, there would be little cause for concern.

Unfortunately, my reanalysis of their findings suggests that they underestimated the rate of intended and unintended interpretations. However, we must pause to consider what would constitute a “bona fide” interpretation of the source of disagreement. Goodwin and Darley (2008) provide a handful of examples, then remark more generally that they excluded “[r]esponses in which participants

⁶⁷ At least, this is what I infer from analyzing their raw data. Geoffrey Goodwin was kind enough to share their data and their original coded analysis of the open responses, and this was the question that appeared in that file.

⁶⁸ As they put it: “We were wary of the fact that some individuals might interpret ethical disagreement as indicating that the disagreeing other person might have been thinking of extraordinary extenuating circumstances, or that they had misread the question, or that they may not have understood the words used in a conventional way, and so on [...] The list of such possible caveats to interpreting the prima facie ethical disagreement is large.” (p. 1347)

said they needed more information about the context of the events were also excluded for this reason” (p. 1348, footnote 5). Here is one concrete example of a response that they excluded:

A difference in perception of a situation in which gunfire was opened on a crowded city street. I was thinking gunfire from terrorists/criminals; other person may have thought gunfire from police officers to catch a criminal.
(p. 1348, footnote 5)

Finally, they state they did not exclude “responses in which participants said that perhaps the disagreeing other was operating with a different sense of “morally wrong/bad.” While they do not explicitly state what counts as a “bona fide” interpretation, it is not hard to determine what type of interpretation they have in mind, or at least *ought* to have in mind: *fundamental moral disagreements*, i.e., disagreements about basic moral values, rather than differences in nonmoral beliefs or standards (Bush, 2016; Caven, 2015; Moody-Adams, 2009). As Caven (2015) puts it:

Fundamental moral disagreement involves two agents who are situated in the same context, possess the same non-moral understanding and are both free of errors of inferential reasoning, yet come to different judgements concerning what is the morally best course of action, all-things considered. (p. 3)

For instance, Alex and Sam could disagree about whether a particular killing was justified because Alex believes the killing was an act of self-defense, and Sam believes it was an act of premeditated murder. Both agree that killing in self-defense is morally justified, but that premeditated murder is not. Yet they could disagree about whether the particular event that took place was an act of self-defense or an act of premeditated murder. If so, Alex and Sam do not have a fundamental moral disagreement, because they share the same moral stance towards self-defense and premeditated murder. They just disagree about what actually happened (i.e., the nonmoral facts).

If participants attribute the source of the disagreement between themselves and a previous participant to something other than a fundamental moral disagreement, then there is no way to be sure whether their response reflects a metaethical stance or commitment. For instance, if participants thought that the previous participant who ostensibly disagreed with them was not thinking of the same

moral issue as they (the participant) were, then whether the previous participant's judgment was "correct" or "mistaken" would have no bearing on realism and antirealism. After all, if I think premeditated murder is wrong, and you inform me that someone else thinks killing in self-defense is not wrong, then both the judgment that they are mistaken *and* the judgment that neither of us is mistaken are consistent with realism and antirealism. In short, participants *must* have regarded the disagreement between themselves and a prior participant to be a fundamental moral disagreement. If they did not, then their response does not reflect a metaethical stance or commitment.

Participants. Responses were initially collected from the participants who participated in Experiment 1 ($n = 50$) and Experiment 2 ($n = 76$) in Goodwin and Darley (2008).⁶⁹ Experiment 1 consisted of 50 Princeton University undergraduates (28 females, 22 males) who participated in the study for course credit, while Experiment 2 included 76 students drawn from two populations: (1) 71 Princeton University undergraduates (49 females, 29 males) who participated for course credit and a second sample of five theological seminary students who participated for \$10 (these samples were collapsed into a single sample for analysis).⁷⁰ No other demographic details were provided.⁷¹ Experiment 2 likewise included two responses from each participant, resulting in a total of 152 responses. This resulted in a total of 252 responses across Experiments 1 and 2 (corresponding to Studies 1A and 1B, respectively).

⁶⁹ Details about responses are based on what was reported in Goodwin and Darley (2008).

⁷⁰ The rationale for including five seminary students was to increase the number of participants likely to attribute their moral standards to a divine source.

⁷¹ Respondents in Experiment 1 answered two open response questions. In Goodwin and Darley's original article, they report 102 responses, and appear to list 51 subjects. However, only 50 subjects are listed in the data. Upon examination, participants are numbered 1 through 51, but respondent #23 is missing from the dataset. Thus, it is unclear whether there were 50 participants who each responded twice, resulting in 100 responses, or whether there were 51, resulting in 102 responses. Since two of these responses were not available to us, I only coded and analyzed the 100 available responses. Goodwin and Darley make no explicit reference to this inconsistency in the data reported in their original article, though they reference having only 50 participants, yet later report 102 written responses (they do reference "missing data" in a note under Table 1, on p. 1347).

Procedure. I do not have complete access to the original study materials and did not conduct the original study myself. As such, this reanalysis of the data was conducted based on data collected following the procedures presented in Goodwin and Darley (2008), and my inferences about what that procedure likely entailed. From the way the study is described in the text, and from the materials I received, it would appear participants filled out printed surveys, and that their responses were subsequently transcribed. After completing the primary measures in the study, participants appear to have completed additional follow-up questions. The specific question analyzed here was included among these questions and appeared as a question separated from others by a bullet, and followed by several empty underlined lines in which participants could write a response, like this:

- Give us your thoughts about why it is that there is disagreement. What could be its source?

Written responses were presumably transcribed. The original data I received was in an Excel spreadsheet, so I presume the written responses were either transcribed directly to it or transcribed elsewhere and then copied and pasted to it. Interpretation rates for Experiment 1 and Experiment 2 were analyzed separately.

Measures. Interpretation rates for all items in Study 1A were coded by myself and David Moss. Interpretation rates for all items in Study 1B were coded exclusively by myself. Participants provided a written response to the question outlined above. All responses were coded in the two ways described above. Each item was coded as either an intended (“1”) or unintended (“0”) interpretation and as being clear (“1”) or unclear (“0”).

These codes were then combined to form one of four codes:

$1|1 = \text{Clearly intended interpretation}$

$0|1 = \text{Clearly unintended interpretation}$

$1|0 = \text{Unclear intended interpretation}$

$0|0 = \text{Unclear unintended}$

A second coder (David Moss) was recruited and conducted the same analysis of the data for Experiment 1. A clearly intended interpretation was one that indicated the participant attributed the source of disagreement to a *fundamental moral disagreement* (see above; see also Bush, 2016; Caven, 2015; Moody-Adams, 2009). A clearly unintended interpretation attributed the source of disagreement to something other than a fundamental disagreement, such as the other person misunderstanding the question or imagining a different moral situation than the participant. Unclear interpretations ($1|0$ or $0|0$) failed to clearly convey either an intended or unintended interpretation.

After initial coding, I reviewed the data to identify sources of disagreement. We attempted to resolve all disagreements via discussion. In some cases, we reached an agreement and coding disputes were resolved. In other cases, we maintained our disagreement, and in those cases responses were not changed by either of us. No other coders participated in the analysis of Experiment 2. Thematic analysis for both Experiment 1 and Experiment 1 was conducted using the method described above, though neither David nor any other coders were involved in this process.

Study 1A Results

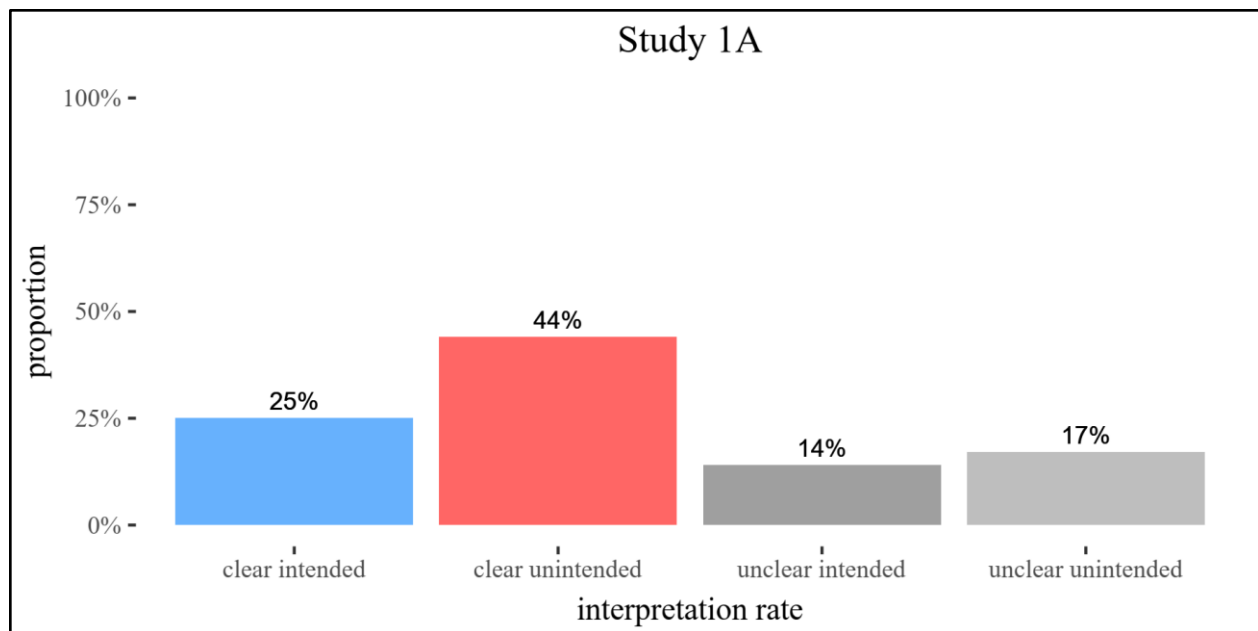
Interpretation rates. Overall, 25% ($n = 25$) provided clear intended interpretations. A one sample proportion test without continuity correction was conducted to determine whether the proportion of responses coded as clearly interpreted was significantly less than 0.5, and it was significantly less than 0.5, $\chi^2(1, N = 100) = 25, p < 0.001$.⁷² With 95% confidence, the percentage of responses coded as

⁷² I report a chi-squared test statistic (χ^2) because all analyses were conducted in R using `prop.test()`. Although `prop.test()` conducts a chi-squared test, χ^2 is equivalent to the square of the Z-statistic. I provide the following example in chapter

clearly interpreted was less than 32.72%. In addition, 44% ($n = 44$) of participants responded in a way that was coded a *clear unintended* response, while the remaining 31% ($n = 31$) of responses were *unclear*. David Moss coded 20% ($n = 20$) of responses as clear intended, 57% ($n = 57$) as *clear unintended*, and 33% ($n = 33$) as *unclear*. Results may be seen in **Figure 4.1**.

Figure 4.1

Interpretation rates for Study 1A



Interrater reliability. Initial coding was conducted independently by myself and one other coder, David Moss. David Moss is an experienced researcher competent in both qualitative research and metaethics who has collaborated with me on numerous projects and is familiar with the literature both on the psychology of metaethics and critiques of the assumption that folk metaethics is uniform and determinate (as expressed by e.g., Gill, 2009 and Loeb, 2008). David has also previously conducted

five as well as here: the formula for a one proportion Z-test is $Z = (p - p_0) / \sqrt{p_0(1 - p_0)/n}$. For instance, the Relativism 1 condition in Chapter 5, Study 1A, this would be $Z = (0.24 - 0.5) / \sqrt{0.5(1 - 0.5)/50} = -3.676955$. $(-3.676955)^2 = 13.52$. Radziwill (2015) explicitly commented on this fact, observing that a one sample proportion test in R technically conducts a chi-squared test, noting that “The prop.test function doesn’t even do a z test. It does a Chi-square test, based on there being one categorical variable with two states (success and failure)!”

independent qualitative research on folk metaethics (in the form of interviews). David and I have worked together extensively on the psychology of folk metaethics and normativity more generally (see Moss & Bush, 2022). In fact, David proposed the coding scheme used here. Nevertheless, coding with a non-adversarial collaborator has the disadvantage that we share similar expectations about the data and are thus likely to be biased in the same direction.

After coding, results were compared, and disagreements were resolved through discussion. Some disagreements remained unresolved. After attempting to resolve disagreements, interrater reliability was assessed via Cohen's Kappa (McHugh, 2012). I calculated interrater reliability in both a collapsed and uncollapsed form. The collapsed form involved converting all responses that were clearly unintended and unclear into a single category and distinguishing this category from clearly intended interpretations. This allowed me to assess agreement specifically on which items were instances of clearly intended interpretations, while ignoring disagreements about whether responses that were not clearly intended failed to be categorized this way because they were clearly unintended, unclear intended, or unclear unintended. This resulted in a binary coding system (1 = *clear intended*, 0 = *not clear intended*). The uncollapsed condition did not collapse any of the four categories when comparing interrater reliability (1 = *clear intended*, 2 = *clear unintended*, 3 = *unclear unintended*, 4 = *unclear unintended*). Interrater reliability was calculated using the irr (version 0.84.1) package in R. Interrater reliability for uncollapsed coding of responses was moderate $K = 0.759$, (84% agreement). Interrater reliability for collapsed coding of responses was high $K = 0.857$ (95% agreement).⁷³

Study 1B Results

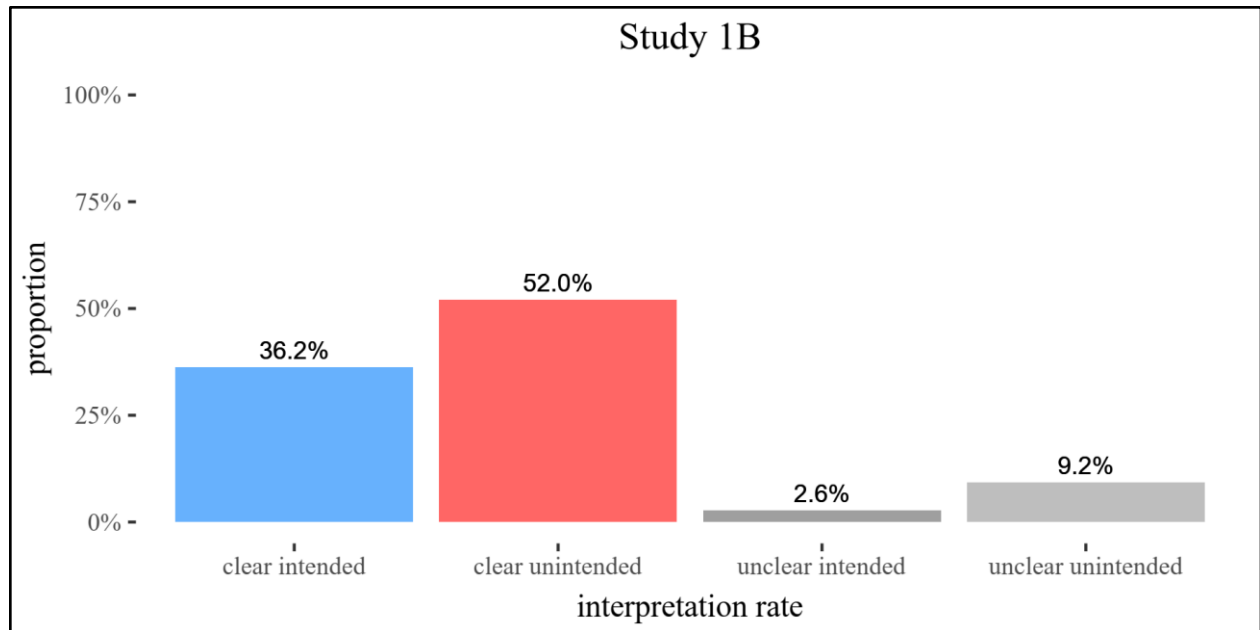
Interpretation rates. 36% ($n = 55$) of participants interpreted the source of disagreement in a *clear intended* way. This was significantly less than 0.5, $\chi^2(1, N = 152) = 11.61, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 42.79%. 52% ($n =$

⁷³ I report “moderate” and “strong” following the guidelines outlined in McHugh (2012).

79) interpreted in a *clear unintended* way, and the remaining 11.8% ($n= 18$) of responses were *unclear*. Results may be seen in **Figure 4.2**.

Figure 4.2

Interpretation rates for Study 1B



Discussion

The proportion of *clear intended* interpretations was very low in both studies. Both coders judged less than half of participants to clearly interpret the question as intended, and both coders judged around half of participants in both samples to have interpreted the source of disagreement in a way that was clearly unintended. Even if it was charitably assumed every participant whose response was unclear interpreted the source of disagreement as intended, this would still indicate that about as many people did not interpret the source of disagreement as intended as those who did. This is an extremely serious methodological shortcoming with the results of this particular study. This is because interpreting the source of disagreement as intended is *necessary* for the validity of the disagreement paradigm. If the participant does not regard the disagreement between themselves and another person to result from

different moral standards, then their responses to the main measures cannot tell us whether they think that there is a stance-independent moral fact about the moral issue in question. Nor does it tell us that they think there is no stance-independent moral fact. Rather, their response is effectively a response to a different question entirely. As a result, their response simply cannot tell us one way or another about their metaethical standards or commitments.

Our goal in reanalyzing Goodwin and Darley's findings was to evaluate the accuracy of their analysis, which purportedly revealed that only a small number of participants interpreted the source of disagreement in a way that would challenge their findings. In particular, they reported that about 7% of participants (7 out of 102) in Study 1A did not interpret the source of disagreement as intended. The present findings suggest that the true proportion is almost certainly much higher than this. Indeed, I can provide more than 7 examples of participants who very clearly interpreted the source of disagreement in an unintended way. In fact, I can provide *dozens* of such examples, and this would reflect what may be a *lower bound* on the proportion of participants who interpreted the source of disagreement in an unintended way, since this isn't even accounting for the unclear interpretations. This reanalysis of this data strongly suggests that Goodwin and Darley underestimated the degree to which participants attributed the source of disagreement to something other than a genuine difference in moral belief.⁷⁴

Overall, these findings support the conclusion that a substantial proportion of participants did not interpret the source of disagreement as intended in the studies conducted by Goodwin and Darley (2008). This is consistent with the more general concern that, when participants are presented with

⁷⁴ I had no precise predictions about the overall proportion of people who would not interpret the source of disagreement as intended. While I expected the number to be substantial, I had no distinct hypothesis (e.g., "more than half"). The expectation (which was mine, not David's) that fewer than half of participants would interpret questions as intended arose as a result of the coding for this study, analyzing subsequent open response data, and further consideration about the implausibility of high rates of intended interpretation. This also resulted in a reanalysis of the original data in light of the greater experience I developed while analyzing other datasets.

versions of the disagreement paradigm similar to the one employed by Goodwin and Darley, that many will not interpret the source of disagreement in a way that is necessary for the disagreement paradigm to be valid. In short, these findings challenge the validity of at least some forms of the disagreement paradigm.

4.6 Study 1C

It's possible that the low clear intended interpretation rates in Study 1A and 1B were due to idiosyncratic features of the participants who were sampled or to the methods employed by Goodwin and Darley (2008). To address this, I assessed interpretation rates using a new version of the disagreement paradigm and found the same general results in a new sample. This sample differed from Study 1A and 1B in that I sampled participants on Amazon's Mechanical Turk, rather than sampling from a student population. In addition, the materials were presented in an abstract, simplified form with minimal instructions. The goal was to eliminate any potential sources of ambiguity introduced by the instructions employed by Goodwin and Darley, and by focusing on an abstract, unspecified moral disagreement rather than on specific, concrete moral disagreements about particular issues (e.g., abortion), I minimized the risk of normative confluences influencing participant responses. Interpretation rates for all items in Study 1C were coded exclusively by myself. In spite of these changes, the proportion of participants coded as *clear intended* was 23% ($n = 23$), which was significantly less than 0.5, $\chi^2(1, N = 100) = 29.16, p < 0.001$. See **Supplement 3**, section **S4.4.1** for full details.

4.7 Study 2: Interpreting other people's responses to metaethics stimuli

Another way to assess whether people interpret questions about metaethics as intended is to evaluate how they interpret other people's responses to questions about metaethics. One advantage to adopting this approach is that the participant does not have to interpret the question about metaethics directly, but is instead presented both the question and a response to that question that did interpret the

question as intended. This provides additional context that minimizes potential ambiguity or uncertainty about what the question is asking and provides a direct and explicit articulation of a metaethical stance. For comparison, suppose you asked a participant:

Do you like to go to the bank?

One problem with this question is that it is ambiguous. Some participants may interpret this as a question about financial institutions, while others may interpret it as a question about riverbanks. If we then asked the participant why they answered the question the way they did, and they say something like:

I like to go to the bank because it means I've got enough money to take some out!

If we were interested in whether people like going to riverbanks, such a response would be unhelpful, since this participant interpreted the question to refer to financial institutions rather than riverbanks. When participants respond this way, we cannot know whether they like going to riverbanks or not. Yet suppose we instead presented participants with the following:

A respondent was asked the following question:

Do you like going to the bank?

John: "No. I don't like going to the bank. I usually have to wait in line and it's really boring. Plus, I can do most of my banking online, anyway."

In your own words, what do you think the respondent means in the statement above?

It would be difficult for participants to *not* interpret John's response to refer to financial institutions rather than riverbanks. As such, we could be confident that *if* people understand what banks (understood as financial institutions are) are, and *if* they understand what it means to not like going to the bank, then they are likely to understand John's response. In short, providing another person's response can help disambiguate what a question is asking, and by focusing on the *response* to the

question, rather than the question itself, we may increase our chance to ensure that participants are in a good position to understand the concepts we are asking about.

Study 2 assesses whether people understand specific expressions of moral *realism* and moral *relativism* (a specific form of antirealism). *Realism* refers to the belief that there are stance-independent moral facts, while *relativism* refers to the claim that moral facts are only true or false relative to the standards of individuals or groups.⁷⁵ Interpretations of question and response pairings reflecting realism and relativism were each tested with two distinct question-response pairings.

One pair of questions present a direct question about realism or relativism and a person's response to the question. A second pair of questions present a version of the disagreement paradigm, and either a realist or relativist response to it. In both cases, participants were asked to explain in their own words what they think the *response* to the question means. Thus, I did not ask participants to explain what they think the question is asking, nor did I ask them how they would respond to the question. Rather, my exclusive focus was on how they interpret someone else's response to the question, with such responses intended to reflect either realism or relativism. For all four measures, I expected a majority to not interpret the question in a clearly intended way.

Methods

Participants. Participants consisted of 449 adult US residents on Amazon's Mechanical Turk (208 males, 191 females, 1 other, 53 unreported, $M_{\text{age}} = 34.6$, $SD_{\text{age}} = 10.9$, age range = 19-70). Since the *direct* and *disagreement* conditions are conceptually distinct, results from each will be analyzed separately.

⁷⁵ As always, relativizing moral standards to individuals and groups is a simplification. It is always possible for moral standards to be thought of as true or false relative to some other standard of evaluation, such as different religious systems or different species. While such forms of relativism are conceptually possible, they are probably not common in practice. These various ways moral claims can be relativized are also not necessarily mutually incompatible. While someone could think that moral claims can only be relative to individual standards, they could think that they can be relativized to the standards of both individuals *and* groups, or even individuals, groups, species, religious systems, and so on. One could, for instance, think moral standards can be true or false relative to any real or hypothetical standard of evaluation in a profligate way, that is, without consideration for what type of evaluative standard it is. I, for one, find this *more* plausible than more narrow construals of relativism.

Procedure. All participants were randomly assigned to one of four conditions. Two conditions represented realism, and two conditions represented relativism. In addition, two conditions presented a direct question about realism or relativism followed by a response, and two questions presented a version of the disagreement paradigm followed by a response. This resulted in four conditions:

1. *Direct realism*
2. *Direct relativism*
3. *Disagreement realism*
4. *Disagreement relativism*

After receiving instructions that informed participants that they would be given an anonymized response from a participant in a previous survey, they were then presented with a question and a response, and asked to explain in their own words what the previous respondent meant by their answer to the question. In other words, they were *not* asked to interpret the question, but the other person's *response* to the question). In all conditions participants were presented with the following instructions:

Instructions

Recently, a diverse range of American citizens participated in a survey about moral attitudes conducted by the Pew Research Center. You will now be presented with the question and response of one of these participants. The question and respective answer will be selected at random. The name of the respondent was changed to preserve his or her identity.

Please read carefully. Afterwards, you will be asked a series of questions about the respondent. There are no right or wrong answers, please just provide your thoughtful first response.

Participants were then told that the respondent was asked to respond to a question. Participants were given that person's response and told to think about the person who answered the question. After they were presented with these instructions, participants were asked a series of questions (not reported here). Later in the study, they were asked the following question:

In your own words, what do you think the respondent means in the statement above?

Since this question was presented on a page that appeared after the initial instructions were presented, reminder text featuring the question and response were presented at the top of the screen. The specific

wording used for each of the four conditions (*direct realism*, *direct relativism*, *disagreement realism*, and *disagreement relativism*) is featured on **Table 4.1**. All of the answers “John” provided were written by the research team, and were not actual responses written by previous participants.

Measures. Interpretation rates for all items in Study 2A and Study 2B were coded exclusively by myself. All participants were presented with a text box where they could write a response. Although other stimuli were included and other measures were collected in the original study, the only measure analyzed here consists of the written responses participants provided. All responses were coded following the procedures outlined in the introduction. As such, there were two measures: *interpretation rates* and *thematic analysis*. Interpretation rates were coded by judging whether each response was clear or unclear, and whether it reflected an intended or unintended interpretation. Themes were also coded for each response.

Results

As expected, most participants did not clearly interpret any of the responses to questions about realism or relativism as intended. Across all conditions, the proportion of participants coded as *clear intended* was less than half: direct realism 5.4% ($n = 6$), direct relativism 12.4% ($n = 14$), disagreement realism 1.8% ($n = 2$), and disagreement relativism 25.9% ($n = 25.9\%$). The proportion of participants coded as clearly interpreted was significantly less than 0.5 for direct realism, $\chi^2(1, N = 112) = 89.29, p < 0.001$, for direct relativism $\chi^2(1, N = 113) = 63.94, p < 0.001$, for disagreement realism $\chi^2(1, N = 112) = 104.14, p < 0.001$, and for disagreement relativism $\chi^2(1, N = 112) = 26.04, p < 0.001, 95\%$. With 95% confidence, the percentage of responses coded as clearly interpreted for the direct realism condition was less than 10.03%, while it was less than 18.39% for direct relativism, 5.25% for disagreement realism, and 33.21% for disagreement relativism. These results indicate that most participants did not clearly interpret the question as intended. In addition, many participants interpreted all four responses in a *clear unintended* way: direct realism 48.2% ($n = 54$), direct relativism

46.0% ($n = 52$), disagreement realism 28.6% ($n = 32$), disagreement relativism 33.9% ($n = 38$). All results are summarized in **Figure 4.4**. Overall, participants appeared less likely to provide a clear intended interpretation for items expressing realism than for relativism, and more participants offered unclear responses for the disagreement conditions than for the direct conditions.

Table 4.1

Description of conditions for direct realism/relativism and disagreement realism/relativism

Direct realism	Direct relativism
<p>The respondent was asked the following question:</p> <p>Do you think there is a single correct answer as to whether individual moral statements are true or false?</p> <p>John:</p> <p>“There is one standard of moral truth by which all people should be judged”.</p> <p>Think about the person who answered this question.</p>	<p>The respondent was asked the following question:</p> <p>Do you think there is a single correct answer as to whether individual moral statements are true or false?</p> <p>John:</p> <p>“There is no single standard of moral truth. Different societies must be judged by different moral standards”.</p> <p>Think about the person who answered this question.</p>
Disagreement realism	Disagreement relativism
<p>The respondent was asked the following question:</p> <p>When two people disagree about a moral issue, do you think they can both be correct, or must at least of them be incorrect?</p> <p>John:</p> <p>“When people disagree about a particular moral issue there can be at most only one correct answer”.</p> <p>Think about the person who answered this question.</p>	<p>The respondent was asked the following question:</p> <p>When two people disagree about a moral issue, do you think they can both be correct, or must at least of them be incorrect?</p> <p>John:</p> <p>“When people disagree about a particular moral issue each can be correct according to their own moral standards”.</p> <p>Think about the person who answered this question.</p>

Figure 4.4

Interpretation rates for Study 2A (by condition)

Figure 4.4.1 Study 2A: Direct | Realism

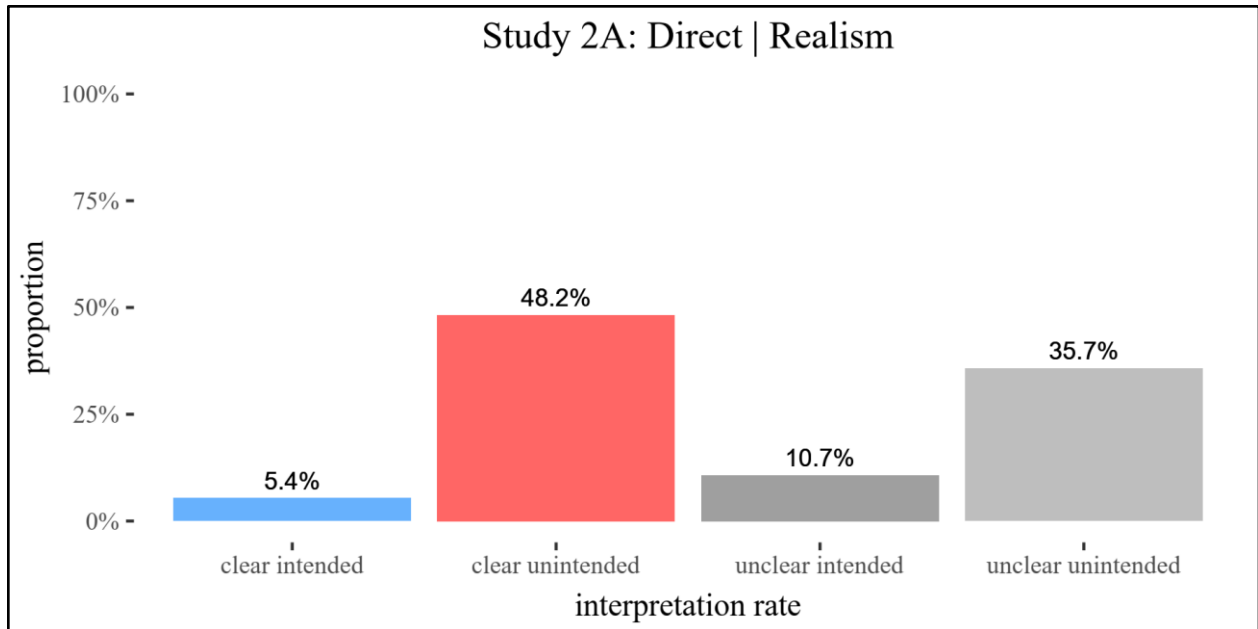


Figure 4.4.2 Study 2A: Direct | Relativism

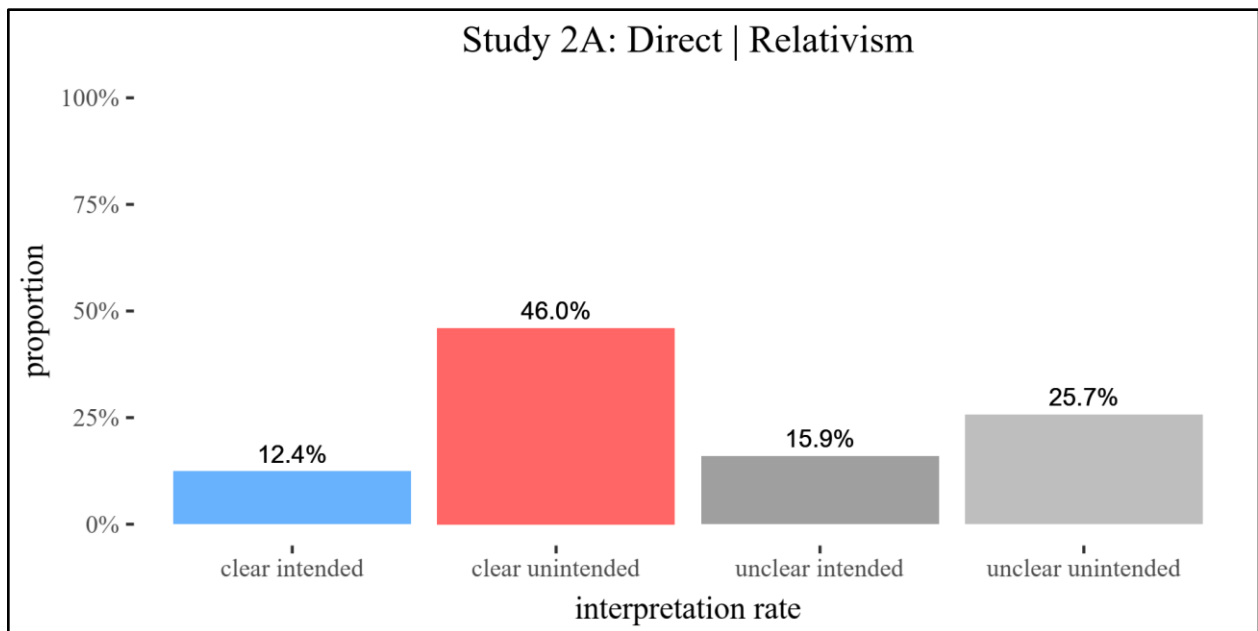


Figure 4.4.3 Study 2B: Disagreement | Realism

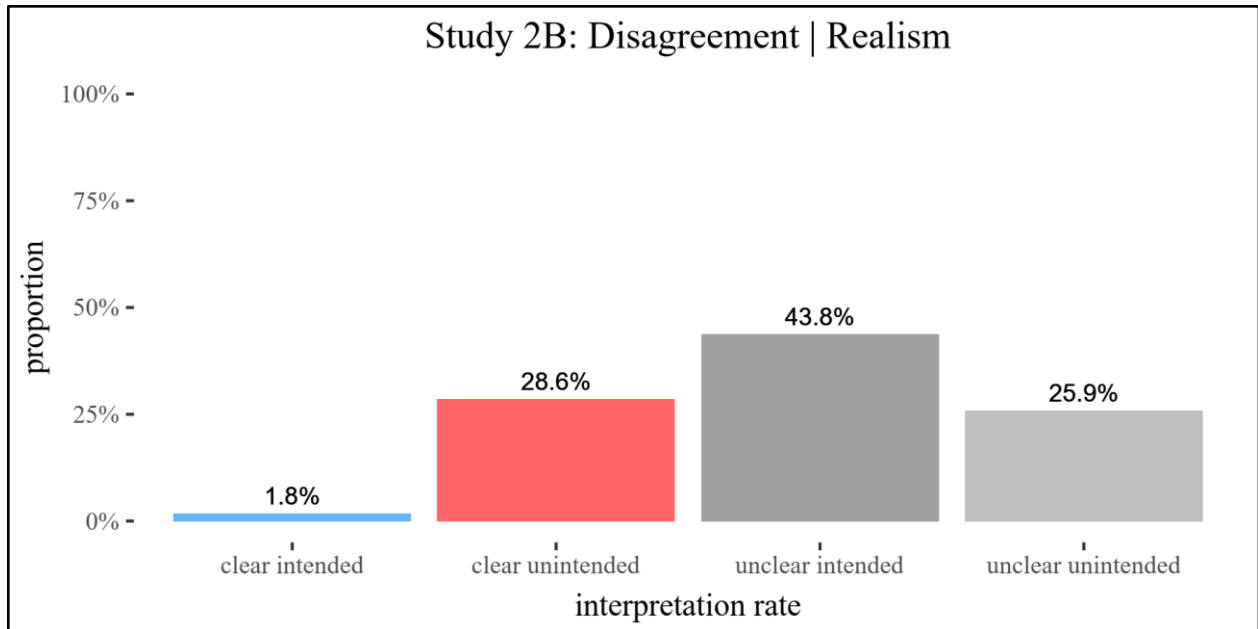
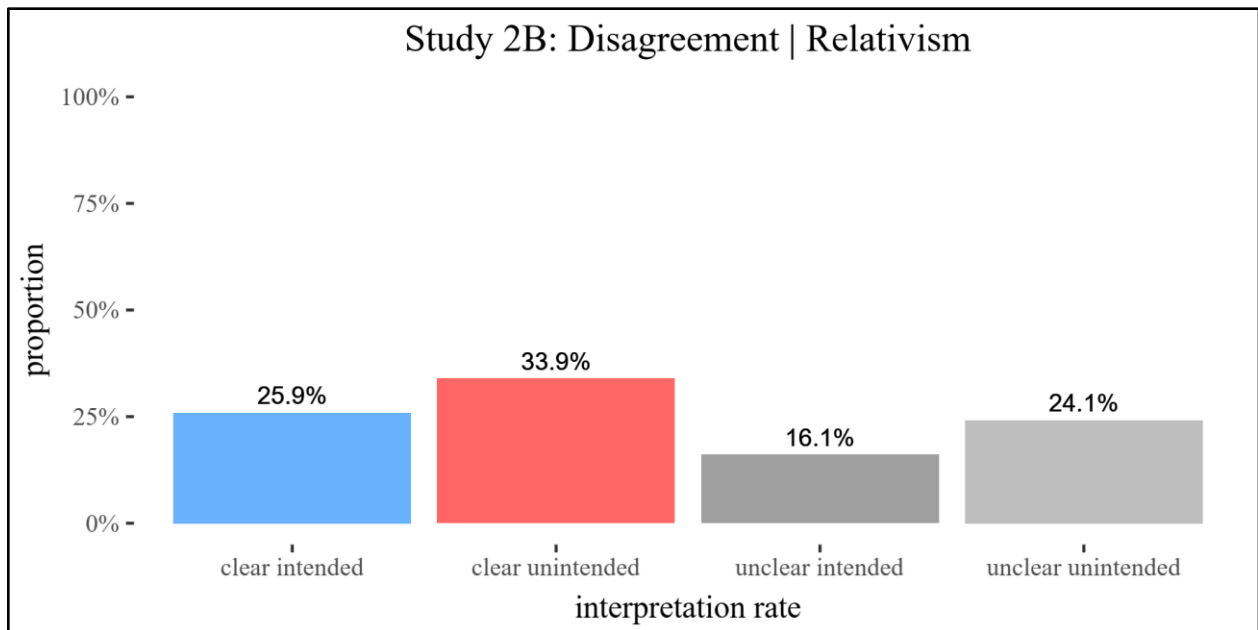


Figure 4.4.4 Study 2B: Disagreement | Relativism



Discussion

Overall, clear intended interpretation rates were very low, ranging from 1.8% - 25.9%, while clear unintended interpretations were very high, ranging from 28.6% - 48.2%. Even if every unclear interpretation were an instance of an intended interpretation, it would still turn out that about a quarter to half of participants across conditions did not interpret what they were asked as intended. While there are some reasons for why this may have occurred that are the result of researcher error, there are nevertheless numerous instances in which participants clearly interpret expressions of metaethical stances in ways that have nothing to do with the intended metaethical stance, or about metaethics at all. This is not what we'd expect if we were asking questions that were easy to interpret. These findings indicate that when you present people with another person's response to a question about metaethics, many participants struggle to interpret that response as an expression of the metaethical stance it was intended to reflect.

Nevertheless, these results suffer from several serious shortcomings. First, the higher rate of clear unintended for the direct conditions may be due to at least two factors. First, the direct realist condition may be due to a methodological problem with the item, which may have inaccurately expressed *universalism* and various *normative* concerns rather than *realism*. Likewise, participants may have interpreted the first part of John's response in the *direct relativism* condition as descriptive for understandable reasons. After all, it states that "There is no single standard of moral truth." This *could* plausibly be read as a descriptive claim. In both cases, the low rates of intended interpretations may be due to misoperationalization rather than the inherent ambiguity that may be present even if realism and relativism were expressed more accurately. As such, the *direct* conditions provide at best only limited evidence. Ironically, I failed to accurately represent realism in a study that was motivated in part by criticism of other researchers for failing to operationalize realism.

There are also shortcomings with the disagreement conditions. In the realist condition, many participants restated John's response without elaborating or expanding on the remark. Such responses do not present a clear intended interpretation, but this may be due to inherent limitations in asking people what a response means. Suppose you were given someone's response, and asked, "what do you think this respondent means in the statement above?" An understandable reaction is to attempt to *restate* the position as clearly as you can, rather than *interpret* it in a way that draws out the background assumptions or intentions of the speaker. If so, you'd be complying with the formal instructions of the task in a way that is not diagnostic of whether you interpreted it as intended. After all, if John thinks that if two people disagree that at least one person must be mistaken, to respond by saying that he thinks that there's only one correct answer is *completely accurate*, yet it tells us nothing about whether you understood the remark to convey a realist stance towards the issue. Strangely, the *repeat* theme was not very common in the relativist condition.

One way of circumventing these interpretive difficulties would be to present participants with a variety of specific and directed questions that provide a more comprehensive and focused way to probe participants for how they interpreted a statement. For instance, you could ask participants to match a statement to one or more statements that most closely reflect what it means or ask participants how much they agree or disagree that someone who made that statement meant to convey a variety of claims, or is likely or unlikely to have particular beliefs or attitudes.

There are two more shortcomings with these conditions. One is that they only describe morality in abstract terms, rather than referring to any specific moral issues. One rationale for considering specific, concrete moral issues is that the added context may facilitate enhanced comprehension. After all, people are probably more familiar with judging specific moral issues rather than thinking about morality as a whole. A second shortcoming is that the preceding studies have, so

far, only considered realism and relativism. However, relativism represents only one form of antirealism, which excludes both error theory and noncognitivism.

Study 3 addresses these shortcomings by presenting metaethical statements in a concrete context and including a noncognitivism condition. In addition, I used statements that more accurately convey the central distinction between realism and relativism in a way that does not conflate realism with universalism or normative claims.

4.8 Study 3: Charity & noncognitivism

This study assesses interpretation rates in the context of a study that was otherwise focused on assessing attitudes about charitable giving. My goal was to assess interpretation with respect to a concrete moral issue. I used several other measures that were not open response questions, which I discuss in **Chapter 5**. Discussion here will only concern the three open response questions presented to participants towards the end of the study.

The primary rationale for employing questions in a charitable context was to assess interpretation rates with respect to a concrete moral issue, to include an item that reflected noncognitivism (in addition to items reflecting realism and relativism), and to present these statements in a simple, alternative format that may prompt a higher rate of intended interpretations and, if it did not, would show that changes in how one conveys metaethical positions does not substantially alter the low rate of clear intended interpretations.

There is considerable precedent for asking questions about noncognitivism. As Beebe (2015) notes, early formulations of the disagreement paradigm did not include a response option for noncognitivism. As a result, early studies on folk metaethics required participants to choose between response options that presupposed cognitivism, even though this may not reflect how ordinary people think about specific moral issues or morality in general. Yet when Beebe included a noncognitivist option alongside the standard realist and relativist responses to the disagreement paradigm, many

participants opted for this response. It was even the most common response for some moral issues! More recently, Davis (2021) found that noncognitivism was the most common response using a novel, more sophisticated version of the disagreement paradigm, averaging 34.2% of responses across the five subdomains of the Moral Foundations Questionnaire (p. 17; Graham et al., 2011), while Pölzler and Wright (2020a, 2020b) found that noncognitivist responses were common across a variety of novel paradigms, including 23% of participants in an abstract version of the disagreement paradigm, as well as 30% of respondents in their “theory” task, 3% in their comparison task, and anywhere from 9%-35% across a variety of concrete moral issues when presented with a truth-aptness task.

While these rates do not comprise a majority of respondents, they show that noncognitivist response options are often favored by a substantial minority of participants. As such, there is little justification for excluding them using a forced choice paradigm that presupposes cognitivism. In this study, I added an additional noncognitivism condition to assess whether participants interpreted a claim about noncognitivism as intended. I included additional realism and relativism items as well, in order to assess whether different ways of conveying realism and relativism would result in greater clear intended interpretation rates.

Methods

Participants. Participants consisted of 367 adult US residents on Amazon’s Mechanical Turk (166 males, 134 females, 1 other, 66 unreported, $M_{\text{age}} = 37.4$, $SD_{\text{age}} = 11.7$, range 19-73).⁷⁶

Procedure. All participants participated in all conditions. Participants were given a brief set of instructions describing the purpose of the study:

Instructions

In this study you will be asked a series of questions about charities. Please read the questions carefully and answer using the options provided. There are no right or wrong answers, please just provide your thoughtful first response.

⁷⁶ One participant reported an age of 5.

Participants were then asked to respond to a series of questions about their level of agreement with a range of statements reflecting metaethical stances towards charitable giving. Participants were presented with a comprehension check to assess their understanding of the notion of something being “objectively true.” Next, all participants were presented with the three open response questions analyzed here, in order of realism, relativism, and then noncognitivism. The three questions may be seen in **Table 4.2**. Order was not varied across participants. Finally, demographic data was collected.

Table 4.2

Wording for conditions in Study 3

Condition	Wording
Realism	In your own words, what does it mean to say that “it is a fact that some charities do more good than others, not a matter of personal beliefs or values”?
Relativism	In your own words, what does it mean to say that “the truth about which charities do the most good depends on the beliefs and values of each individual”?
Noncognitivism	In your own words, what does it mean to say that “there is no fact of the matter about which charities do the most good”?

Measures. Interpretation rates for all items in Study 3A, Study 3B, and Study 3C were coded exclusively by myself. Measures consisted of written responses to the three open response questions.

All responses were coded for interpretation rates and themes.

Results

As expected, most participants did not clearly interpret any of the responses to questions about realism, relativism, or noncognitivism as intended. 9.0% ($n = 33$) provided a clear intended response in the realism condition, 11.4% ($n = 42$) provided a clear intended response in the relativism condition, and 0.5% ($n = 2$) in the noncognitivism provided a clear intended response in the noncognitivism

condition. The proportion of participants coded as clearly interpreted was significantly less than 0.5 in the realism condition, $\chi^2(1, N = 367) = 221.32, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 14.17%. It was also significantly less than 0.5 in the relativism condition, $\chi^2(1, N = 367) = 218.23, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 14.46%. Finally, it was significantly less than 0.5 in the noncognitivism condition $\chi^2(1, N = 367) = 359.04, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 1.63%. 42.8% ($n = 157$) expressed a clear unintended response in the realism condition, 43.1% ($n = 158$) provided a clear unintended response in the relativism condition, and 41.7% ($n = 153$) in the noncognitivist position. Results can be seen in **Figure 4.5**. Overall, these findings support the hypothesis that ordinary people did not understand questions about metaethics as intended.

Figure 4.5

Interpretation rate for Study 3 (by condition)

Figure 4.5.1 Study 3A: Realism

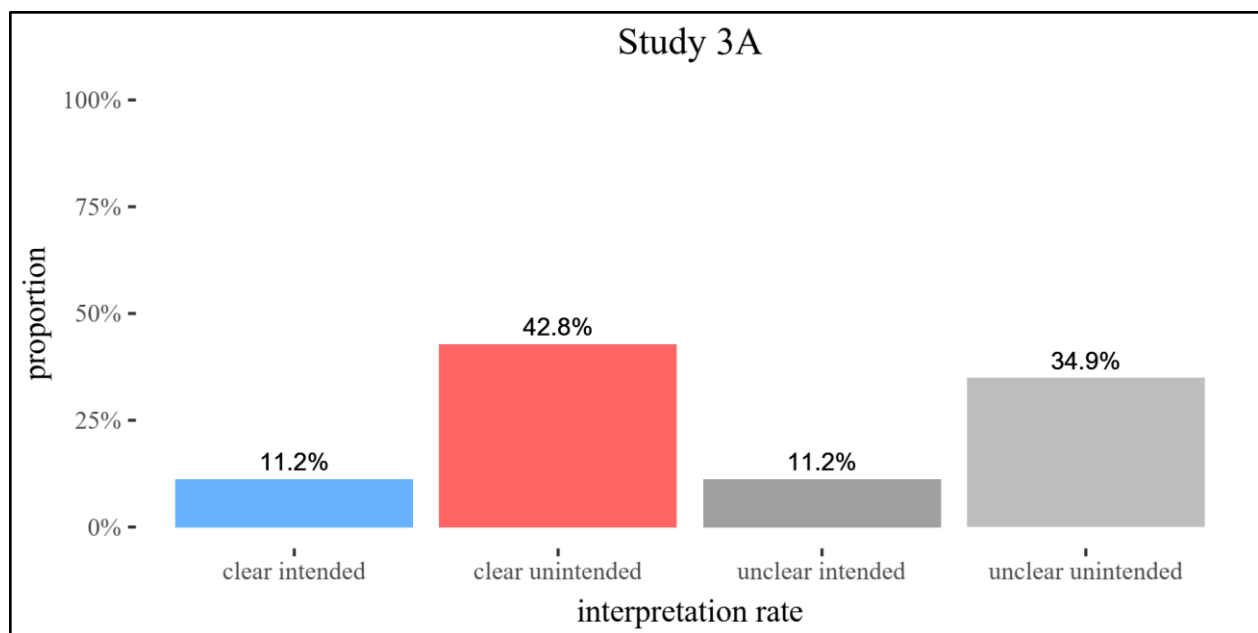


Figure 4.5.2 Study 3B: Relativism

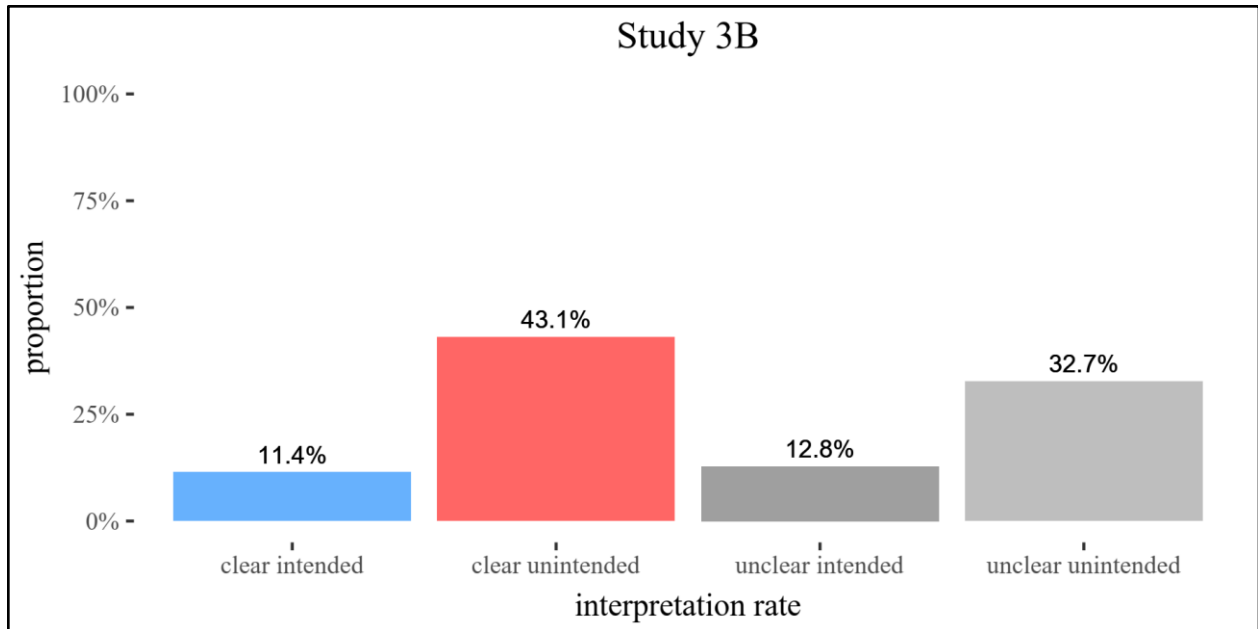
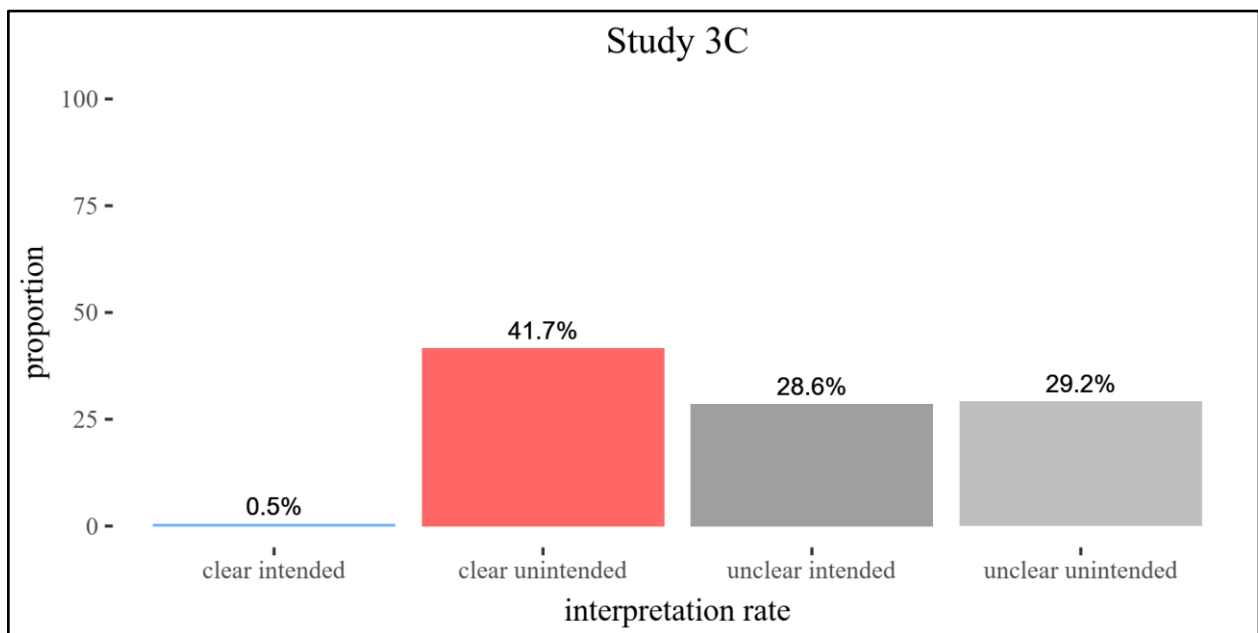


Figure 4.5.3 Study 3C: Noncognitivism



Discussion

Clear intended interpretation rates were very low, while clear unintended interpretations very common. In all three conditions, more than 40% of participants exhibited clear unintended interpretations. Even if every participant coded as having an unclear interpretation interpreted questions as intended, these findings still support the conclusion that most people struggle to interpret questions about metaethics as intended. This provides additional support for the conclusion that ordinary people struggle to understand questions about metaethics as intended. In addition, these findings are consistent with the possibility that most ordinary people have no determinate metaethical stances or commitments, since this could explain why people. Nevertheless, there are significant limitations with these findings.

4.9 Study 4: Direct realism and relativism

One way to assess whether ordinary people are moral realists or antirealists is to directly ask them, using clear and explicit metaethical language. Fisher et al. (2017) take this approach in their third experiment (pp. 1127-1128). Fisher and colleagues report a series of studies which purportedly demonstrate that people who engage in cooperative interactions with other people were less likely to endorse moral realism than participants who engaged in competitive interactions with other people. They found that participants assigned to the cooperative condition had lower mean realism scores. This would indicate that engaging in a cooperative interaction caused participants to be less likely to endorse moral realism. However, Fisher et al. recognize that participants may not have interpreted the disagreement paradigm as intended, so they designed a study with a different measure of metaethics that simply directly asked participants whether, for a series of concrete moral actions, there was an objectively true answer as to whether those actions should be allowed.

This is a laudable step in the right direction, since they explicitly acknowledge the possibility that questions about metaethics may be susceptible to unintended interpretations that could threaten

the validity of the studies in question. The primary dependent variable used in their previous studies was a version of the disagreement paradigm adapted from Sarkissian et al. (2011). Participants were told: “Earlier studies show that people take opposite positions on the issue of [issue] Given that people have opposite views, at least one side must be wrong” (p. 1123). Participants were asked to express their level of agreement (1 = Strongly disagree, 7 = Strongly agree). Fisher and colleagues note that people may have agreed with the statement, not because they are relativists, but because there could be multiple ways of conforming to the same objective moral standards. As they put it:

[I]t may not be clear that we are asking about the metaphysical issue as to whether there is an objective truth about the question under discussion. Instead, participants might interpret the measure as simply asking whether this is a question for which there is only one correct answer versus multiple different correct answers. (For example, if the question had been “Please name a prime number between 1 and 10,” there would be multiple different correct answers, but any given answer would still be objectively true or false.) (p. 1127)

Their solution is to assess whether they find similar results using a *direct paradigm*. *Direct paradigms* are any paradigms that explicitly ask participants about their metaethical views using metaethical language, e.g., “Do you think there are objective moral facts?” or asking people whether they agree or disagree that “moral realism is true.” Most studies appropriately avoid questions like these because their validity depends on participants understanding technical philosophy terms in a way consistent with how philosophers understand these terms.⁷⁷ In this particular instance, Fisher et al. ask participants the following question:

“Consider the following question: ‘Should [topic] be allowed?’ Please tell us whether you think there is an objectively true answer to this question.”

[1 = definitely no objective truth, 7 = definitely an objective truth] [p. 1127]

⁷⁷ Arguably, these studies also rely on the questionable assumption that philosophers themselves agree on the meaning of these terms. They don’t. However, there may be enough overlap that this would render such questions at best only a bit noisy, not necessarily meaningless or invalid

Note the explicit use of the term “objectively true.” Whether this is a valid measure of folk metaethical belief turns on whether participants understand “objectively true” in the intended respect, which Fisher et al. make quite clear: a *metaethical* respect regarding the *metaphysical status* of moral truth. At the start of the paper, they also offer some account of what they mean by *objective*. “There is a correct answer made true by the features of the topic, and any other answer must be wrong” (p. 1120). This is not an ideal characterization of objectivism, but it is clear enough that they are referring to *moral realism*. Thus, whether their measure is valid would turn on whether people understand “objective truths” about morality to refer to stance-independent normative facts about what is morally right or wrong.

Fisher and colleagues ran their version of the *direct paradigm* and reported that they replicated their previous findings. Specifically, they found that the mean level of agreement that the moral issues had “objectively true” answers was significantly lower in the cooperative condition than in the interaction condition. Thus, they found a significant effect in the same direction as their earlier studies, with participants assigned to the cooperative condition expressing reduced endorsement of moral realism. Given that this study found a significant mean difference in the same direction predicted by their previous studies, Fisher and colleagues concluded that participants interpreted their earlier measures as intended, which resolves any concerns that participants did not interpret the disagreement paradigm as intended. Specifically, they state that their result “replicates our main finding from the previous experiments and suggests that participants were correctly interpreting the original measure” (p. 1128).

There are several serious problems with this claim. First, note that they found a statistically significant difference in the same direction across a handful of studies. But it does not follow that because several studies had a statistically significant result in the same direction that *therefore* those studies must be measuring the same thing, nor does it follow that they are measuring the intended psychological construct. For comparison, imagine if I ran two studies. In one study, I asked if people

prefer A over B. In a second study, I ask if people prefer X over Y. If I predict that people would prefer A in the first study, and X in the second study, does it follow that therefore A and X are measures of the same thing? No. While two measures yielding similar results in line with one's predictions provides some minimal corroborating evidence that both measures are valid, such evidence cannot be decisive, nor even very strong. The problem is that, in the absence of extraneous reasons to consider the measures valid, it is possible that two measures that reliably yield similar results are measuring the same phenomenon, or two closely correlated phenomenon, but that the phenomenon they're measuring isn't the phenomenon researchers are intending to measure. In short, two measures cannot pull one another up by their mutual bootstraps alone. To illustrate why, suppose you ran the following two studies:

Study 1	
Condition A	Condition B
Please rate your agreement with the follow statement:	Please rate your agreement with the follow statement:
<i>It is usually good to be nice to people.</i> [1 = Strongly disagree, 7 = Strongly agree]	<i>It is usually good to be mean to people.</i> [1 = Strongly disagree, 7 = Strongly agree]
Study 2	
Condition A	Condition B
Please rate your agreement with the follow statement:	Please rate your agreement with the follow statement:
<i>Pizza tastes good.</i> [1 = Strongly disagree, 7 = Strongly agree]	<i>Dirt tastes good.</i> [1 = Strongly disagree, 7 = Strongly agree]

Imagine you ran both of these studies. What would you predict about the mean scores for level of agreement? Chances are you'd expect mean agreement in **Condition A** in both studies to be significantly higher than mean agreement in **Condition B**. Yet these studies clearly measure something different. Even if the size of the difference in both cases were about the same, this wouldn't be good

evidence that these questions all measure the same psychological construct. Just the same, the fact that Fisher et al. (2017) show that two studies yield similar results *does not* demonstrate that those studies were measuring the same thing.

Furthermore, even if those studies were measuring the same thing, it does not follow that because the studies yielded similar results that they were measuring the *intended* construct. It could simply be that people interpreted the disagreement paradigm in unintended ways, *and* that they interpreted the *direct paradigm* in unintended ways. And since the reasons *why* people do not interpret questions about metaethics as intended is *not random*, both studies could simply recapitulate the same background problem of unintended interpretation. Fisher et al. seem to have made a catastrophic error in reasoning: they forgot that two wrongs don't make a right.

Setting these methodological shortcomings aside, which suggest that Fisher and colleagues are very far from having demonstrated the validity of their measures, we can simply assess whether participants interpret claims about moral objectivity in a way consistent with Fisher et al. (2017). Such findings would demonstrate not only that one purported demonstration of the validity of metaethics paradigms is mistaken, but that direct paradigms are likely going to be uniformly invalid. That is, if ordinary people reliably interpret references to morality or moral claims being “objectively” true in ways that are inconsistent with moral realism, this would suggest that directly asking people about moral realism using terms like “objective” or “objectively” isn't a valid way of measuring folk metaethical stances or commitments.

In the following study, I sought to refute Fisher et al.'s claims by testing whether ordinary people understand the term “objective” with respect to claims about moral truth in a way consistent with their studies and more generally in a way consistent with moral realism. I tested this by testing the hypothesis that a majority of people would not provide a clear intended interpretation of what it means for the truth of a moral claim to be objective. I also tested whether people would understand

the claim that moral truth is “relative,” as well, to head off direct paradigms that directly ask about moral relativism, as well. I was also curious whether participants might respond differently to different moral issues, so for exploratory purposes I included a pair of concrete moral issues. This resulted in six conditions: two abstract conditions which asked whether the truth of a moral claim was “objective” or “relative” in the abstract, without specifying any particular moral issues, and four additional conditions, where I varied both whether I asked about whether the truth of a moral claim was “objective or “relative,” with respect to the either the claim that “murder is morally wrong,” or that “abortion is morally wrong.” Also note an advantage over the previous study, where I did not explicitly use the term “moral.” By using “moral” in these studies, it would be less likely for participants to not interpret the issues in question as moral issues. This is especially important for the “abortion” condition, since Wright, Grandjean, and McWhite (2013) found that only 51% of participants classified “1st trimester abortion” as a moral issue (7% categorized it as a social issue, and 41% classified it as a personal issue).

Methods

Participants. Participants consisted of 300 adult US residents on Amazon’s Mechanical Turk. 459 participants initially participated in the study. However, 159 participants did not complete the study and were dropped from analysis.⁷⁸

Procedure. Participants were asked a single open response question and presented with a text box. Across all conditions participants were asked what it means to say that something is “objective” or “relative” in their own words. Wording for all conditions may be seen in **Table 4.3**. No other data was collected.

⁷⁸ This very high rate of noncompletion is likely due to participants being asked to complete an open response question with no other stimuli. My intent was for participants to write only a short response which should take less than a minute, but the size of the text box and background assumptions about expectations of the task may have led participants to believe I required more significant input. This may undermine data quality.

Table 4.3*Wording for stimuli in Study 4 (by condition)*

Condition	Wording
Abstract Realism	In your own words, what does it mean to say that moral truth is objective?
Abstract Relativism	In your own words, what does it mean to say that moral truth is relative?
Concrete Realism Murder	In your own words, what does it mean to say that the truth of the moral claim "murder is morally wrong" is objective?
Concrete Relativism Murder	In your own words, what does it mean to say that the truth of the moral claim "murder is morally wrong" is relative?
Concrete Realism Abortion	In your own words, what does it mean to say that the truth of the moral claim "abortion is morally wrong" is objective?
Concrete Relativism Abortion	In your own words, what does it mean to say that the truth of the moral claim "abortion is morally wrong" is relative?

Measures. Interpretation rates for all items in Study 4A and Study 4B were coded exclusively by myself. Interpretation rates and thematic analysis were coded for all responses. A clear intended interpretation for realist conditions would consist of understanding the notion that moral truth is “objective” in a way consistent with moral realism, i.e., that there are stance-independent moral facts. A clear intended interpretation of the relativist conditions would entail recognizing the notion that moral claims can be true or false, but only relative to the standards of different evaluative standards (such as those of different people or cultures).

Results

Across all concrete conditions, a majority of participants did not provide a clear intended interpretation: abstract|realism 14% ($n = 8$), abstract|relativism 18.2% ($n = 8$), concrete|murder|realism 18% ($n = 9$), concrete|murder|relativism 6.1% ($n = 3$), concrete|abortion|realism 15.8% ($n = 9$), concrete|abortion|relativism 2.3% ($n = 1$). Aggregating

across conditions, 11.2% ($n = 22$) of participants interpreted the meaning of objective or relative as intended. In the abstract|realism condition, the proportion of participants coded as clearly interpreted was significantly less than 0.5, $\chi^2(1, N = 57) = 29.49, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 23.24%. In the abstract|relativism condition, the proportion of participants coded as clearly interpreted was significantly less than 0.5, $\chi^2(1, N = 44) = 17.82, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 29.49%. In the concrete|murder|realism condition, condition, the proportion of participants coded as clearly interpreted was significantly less than 0.5, $\chi^2(1, N = 50) = 20.48, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 28.50%. In the concrete|murder|relativism condition, condition, the proportion of participants coded as clearly interpreted was significantly less than 0.5, $\chi^2(1, N = 49) = 37.74, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 14.36%. In the concrete|abortion|realism condition, the proportion of participants coded as clearly interpreted was significantly less than 0.5, $\chi^2(1, N = 57) = 26.68, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 25.26%. In the concrete|abortion|relativism condition, the proportion of participants coded as clearly interpreted was significantly less than 0.5, $\chi^2(1, N = 43) = 39.09, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 9.77%. A majority of participants provided a clear unintended interpretation for all four conditions: murder|realism 60% ($n = 30$), murder|relativism 65.3% ($n = 32$), abortion|realism 63.2% ($n = 36$), and abortion|relativism 74.4% ($n = 32$). Aggregating across conditions, 65.3% ($n = 130$) of participants interpreted the meaning of objective or relative as intended. These results indicate that a majority of participants did not interpret any of these items as intended, and a substantial number clearly interpreted them in unintended ways. Results are summarized in **Figure 4.6**.

Figure 4.6

Interpretation rates for study 5 (by condition)

Figure 4.6.1 Study 4A: Abstract | Realism

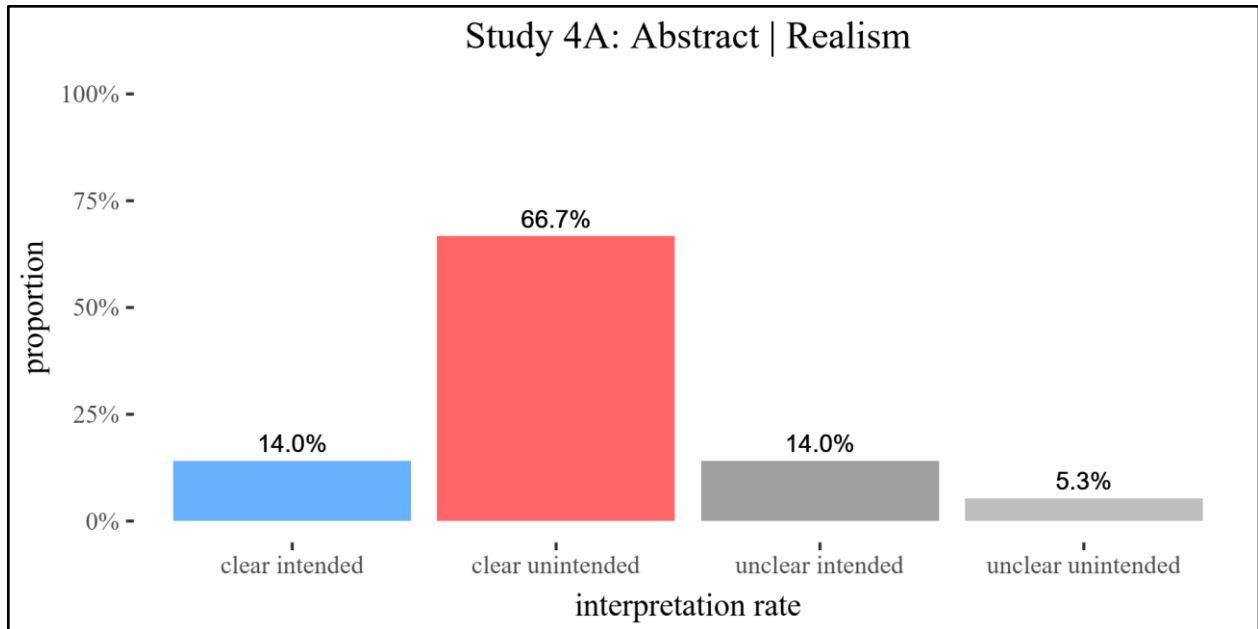


Figure 4.6.2 Study 4A: Abstract | Relativism

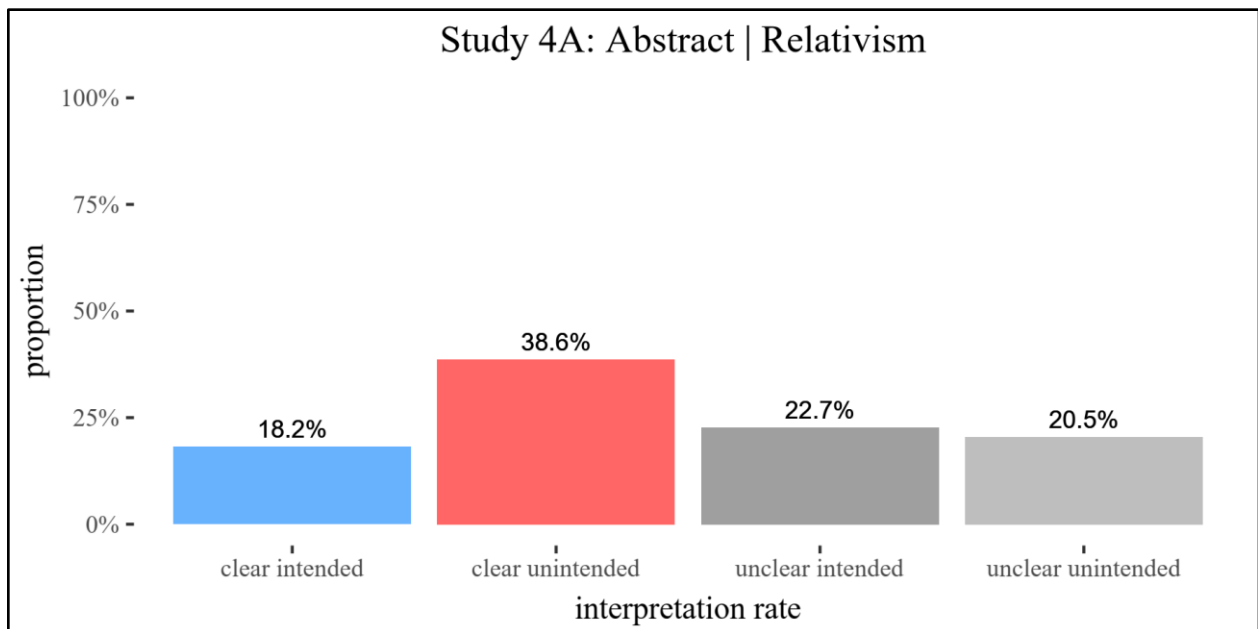


Figure 4.6.3 Study 4B: Concrete | Realism | Murder

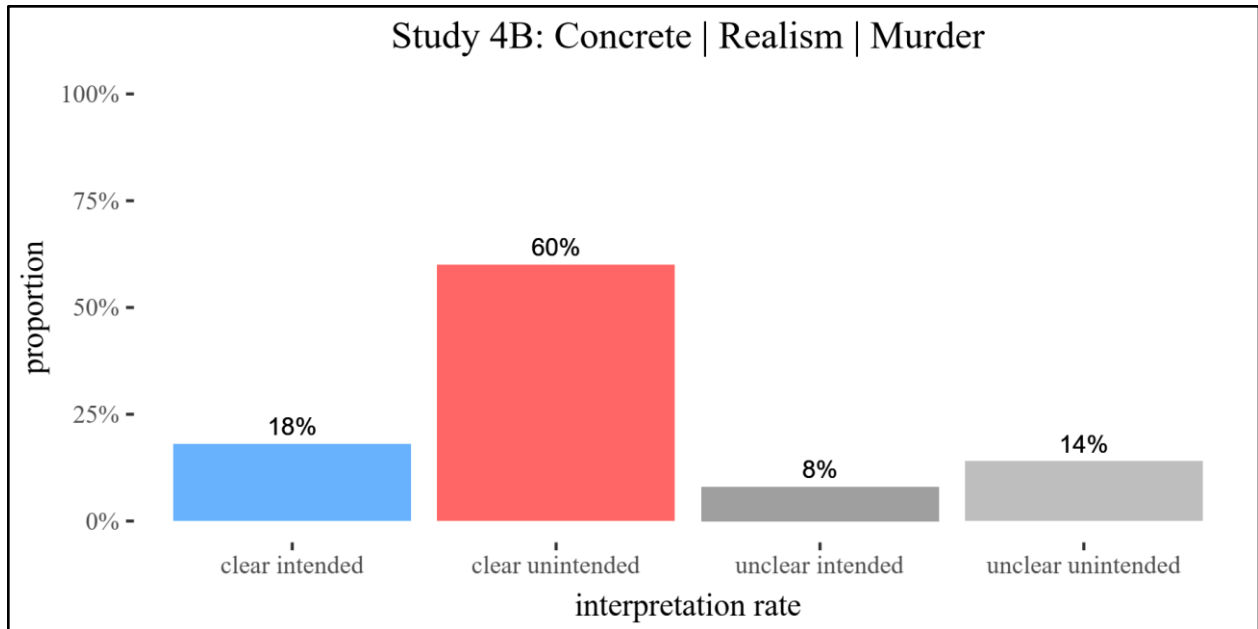


Figure 4.6.4 Study 4B: Concrete | Realism | Murder

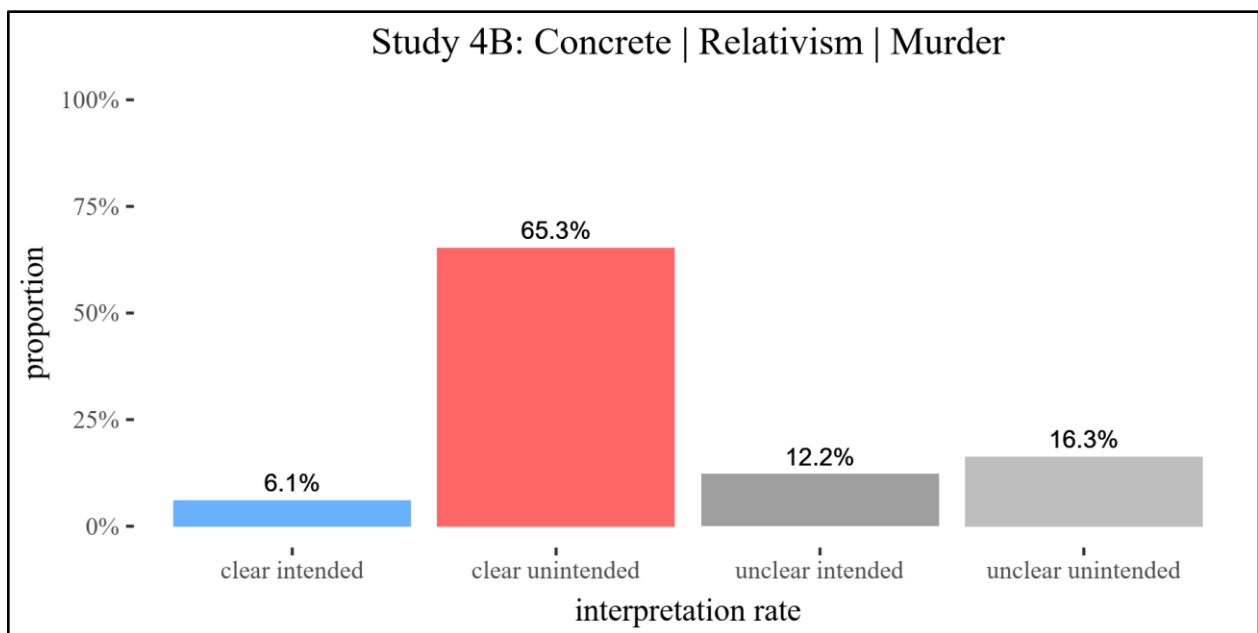


Figure 4.6.5 Study 4B: Concrete | Relativism | Murder

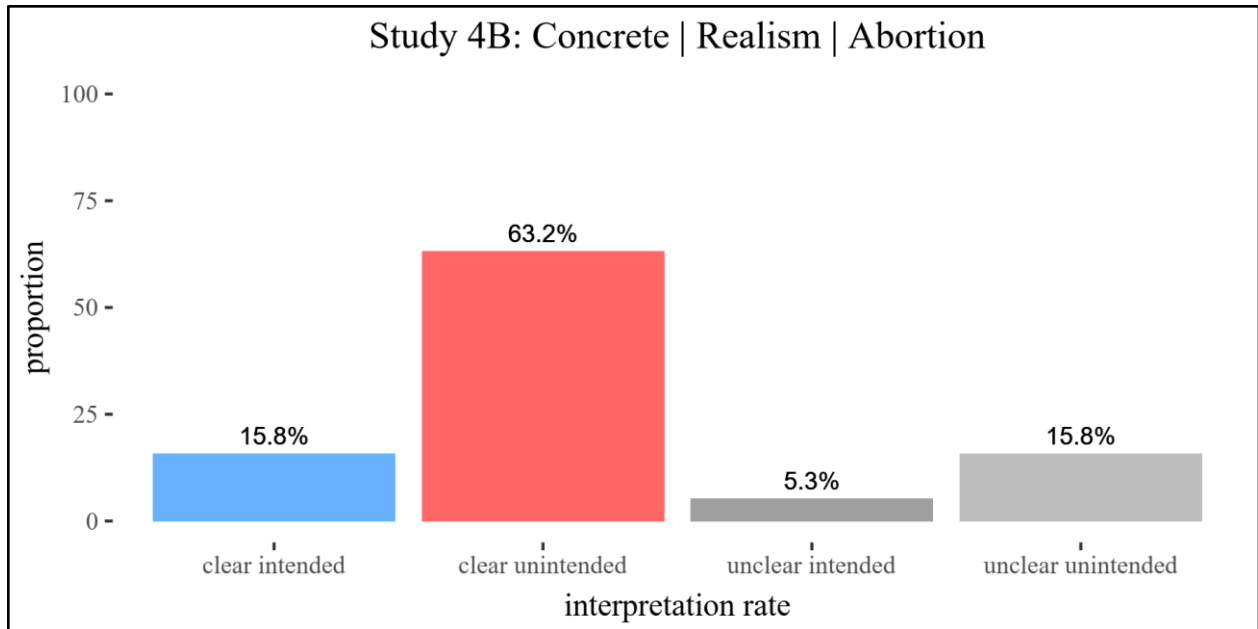
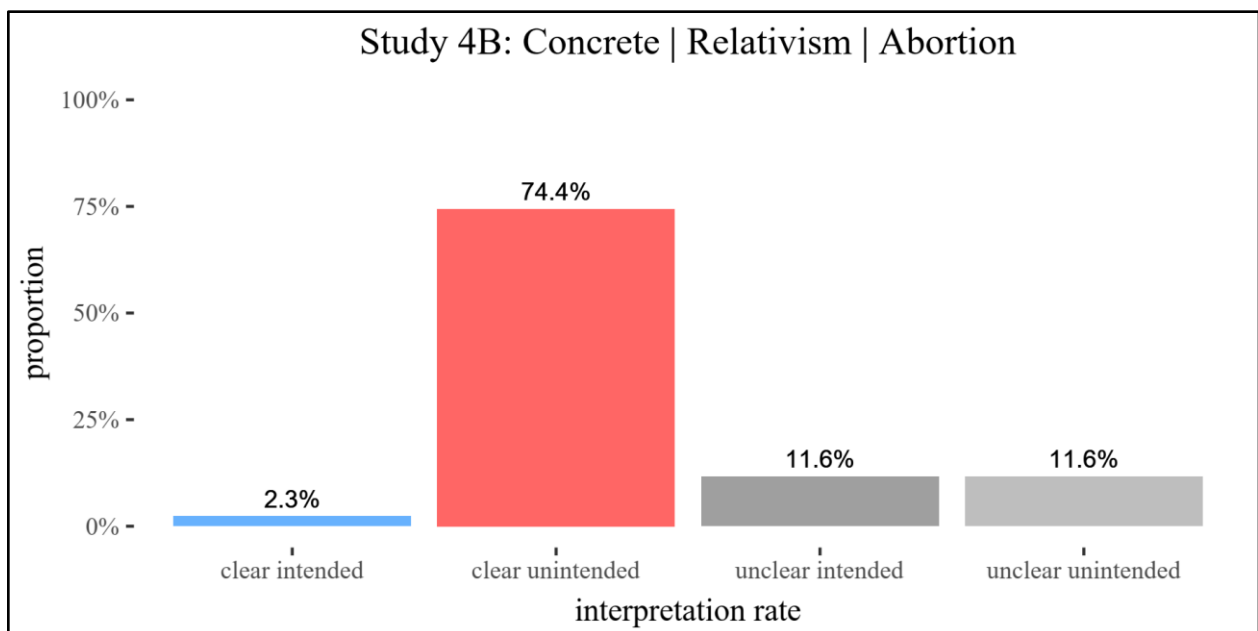


Figure 4.6.6 Study 4B: Concrete | Relativism | Abortion



Discussion

Across all conditions, the proportion of participants who interpreted direct questions about what it means for moral truth to be objective or relative exhibited incredibly low clear intended interpretation rates. The proportion of clear intended interpretation rates ranged from 2.3% - 18.0%. Conversely, clear unintended interpretation rates were extraordinarily high, even compared to analysis of other open response questions. Clear unintended interpretation rates ranged from 65.3% - 74.4%. Thematic analysis corroborates these findings, lending support to the low rate of clear intended interpretations and high rate of clear unintended interpretations being attributable to a coherent and plausible pattern of specific unintended interpretations. For instance, many participants appeared to misread the term “objective” to mean “relative” or “subjective,” while others conflated realism with absolutism or a rigid view of morality (as indicated by the *black and white* theme), or being biased. The most common themes across all three relativism conditions likewise support precisely those confluences I have identified as plausible candidates for ways in which ordinary people conflate relativism with other considerations: in all three conditions, the *descriptive* and *context* themes were in the top three most common themes. This suggests that participants are not merely responding *randomly*, completely clueless as to what I was asking them. Rather, it points to a reliable, recurring pattern of active conflation between relativism and other concepts. This is precisely what we should expect if people do not understand direct questions about relativism.

4.10 Study 5: Interpretation of the Moral Relativism Scale (MRS)

Collier-Spruel et al. (2019) constructed the Moral Relativism Scale (MRS) to, not surprisingly, measure beliefs about moral relativism. The MRS consists of ten items intended to reflect moral relativism or the negation of moral relativism. A participant’s overall score is interpreted as the degree to which they endorse moral relativism. The MRS has many virtues. The paper which presents the scale documents that the scale was validated over the course of nine studies with over 3,200 participants. It

was measured alongside numerous other constructs with an eye towards demonstrating its predictive validity. The final pool of items loaded well onto a single factor ($\lambda \geq 0.60$), and any items with a 10th grade reading level or higher were dropped to ensure wording was not too complicated for some readers. In addition, the items were evaluated by a panel of 11 experts in moral philosophy and psychology and only items that passed muster were retained.⁷⁹ Sure enough, the items exhibit far greater face validity than previous scale items. The efforts to develop this scale are a remarkable piece of work and could serve as a model for many of the steps critical to devising a high quality scale.

In spite of its virtues, I am skeptical that the items on this scale are interpreted as intended often enough for it to serve as a valid measure of folk relativism (see **Supplement 3**). Here, my only interest is in evaluating whether participants interpreted items on this scale as intended. There are several differences between the approach taken here and the approach taken in previous studies where I collected my own data (as opposed to Study 1, where I reanalyzed existing data). First, I used, the closest approximation, the precise wording of the questions used in Collier-Spruel et al., in order to assess how participants interpreted precisely those questions actually used in the scale. Second, I employed two measures: both a request to explain *why* they answered the question as they did, and to explain what they believe the items used in the scale mean. This provided two separate measures of interpretation rate.

Finally, Collier-Spruel et al. (2019) developed two scales, a moral relativism scale, and a moral tolerance scale, in order to evaluate the relationship between relativism and tolerance. I collected the same type of data for the 10-item tolerance scale as well. My goal was to analyze interpretation rates

⁷⁹ I do wonder whether the people described were, in fact, experts. We are told that the expert panel included professors, postdocs, and graduate students who studied philosophy and psychology and that are all “researchers of morality” (p. 4). This does sound like a promising panel. Yet as I have demonstrated *ad nauseum* throughout this dissertation, metaethics is *difficult*. Expertise in moral philosophy and psychology does not necessarily confer expert judgment about the degree to which a given item plausibly reflects relativism. In my experience, even professional philosophers conflate or misconstrue relativism. I fear that even researchers more competent than I am lack the requisite degree of discernment, borne not of general knowledge but of direct and constant experience with considering the validity of metaethics measures to judge the quality of an item.

for the tolerance items. However, after an initial perusal and a second look, it became apparent that virtually no participants appeared to interpret questions about tolerance in unintended ways, nor to explain what the tolerance items mean in unintended ways. Perhaps it would be rhetorically persuasive to analyze and compare the results of the tolerance items, but I am not sure this would be appropriate. While some responses were unclear, there were almost no straightforward instances in which people appeared to interpret questions about tolerating other people in unintended ways. Yet because I am not as experienced or knowledgeable about the concept of “tolerance,” I am not sure that I have the experience necessary to evaluate such responses. This suggests to me that proper open response analysis may require deep immersion in the academic literature and considerable experience analyzing open responses. Since I lack both types of experience, I opted not to analyze the results. However, results are available to anyone who wishes to have a look, and I suspect (or, rather, *expect*) that most people will come away with the impression that participants have no problem interpreting questions about tolerance. This serves, in an indirect way, to reinforce the validity of the methods used here. It could have turned out that open response questions provide such ubiquitously low quality indications of the degree to which participants interpreted questions as intended that there would be no distinctive shortcoming in how people responded to questions about metaethics. But this brief foray into how people interpret other questions provides hints that this is probably not the case. The primary prediction was that fewer than half of participants across all conditions and in aggregate would interpret questions about relativism as expected, and that this would be reflected both in their explanations for why they answered as they did, and their explanations for what they interpreted each item on the MRS to mean.

Methods

Participants. Participants consisted of 156 adult US residents on Amazon’s Mechanical Turk. The complete study included items from both the moral relativism subscale (MRS) and the moral tolerance

scale (MTS) consisted of 353 participants. However, only the results of participants in MRS conditions were analyzed. No demographic data was collected.

Procedure. All participants were asked to rate their level of agreement with an item on the MRS on a 5-point scale. Wording was as follows:

Please rate how much you agree or disagree with the following statement.

One's own culture determines whether that person's actions are "right" or "wrong."

☐ (1) strongly disagree ☐ (2) Disagree ☐ (3) Neither agree nor disagree ☐ (4) Agree ☐ (5) Strongly agree

Next, they were asked to explain why they chose the response they did to the multiple choice question, and were then asked to explain what the item they were asked means. In both cases they were given a text box. Order of the two open response questions was not varied. The two questions appeared as follows:

Please explain why you chose this response.

---{page break}---

In your own words, please explain what this statement means:

[Item] [Example: **One's own culture determines whether that person's actions are "right" or "wrong."**]

Measures. Interpretation rates for all items in Study 5A and Study 5B were coded exclusively by myself. Measures consisted of level of agreement with the item (1 = strongly disagree, 5 = strongly

agree), and written responses to the two open response questions. Results of the two measures are analyzed separately.

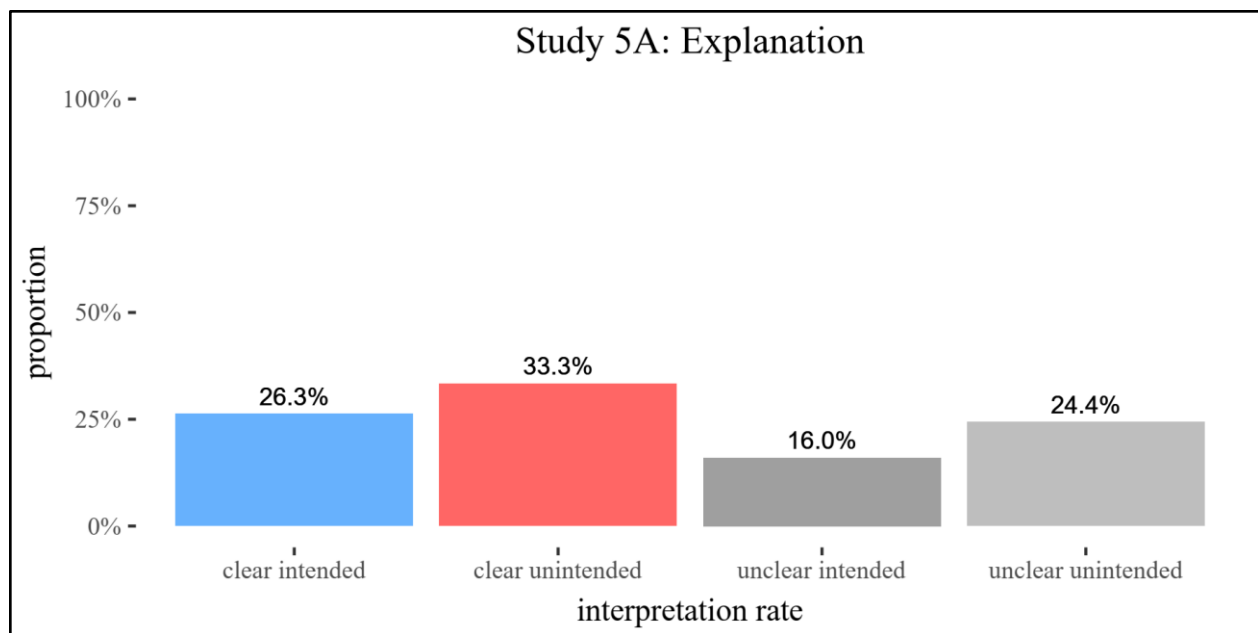
4.10.1 Study 5A: Explanation

Results

As expected, most participants did not interpret items in a clear intended way when summing all conditions. Overall, 28.2% ($n = 44$) of participants provided a clear intended interpretation. This was significantly less than 0.5, $\chi^2(1, N = 156) = 35.10, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 32.45%.⁸⁰ Results are summarized in **Figure 4.6**.

Figure 4.6

Interpretation rates for Study 5A: Explanation



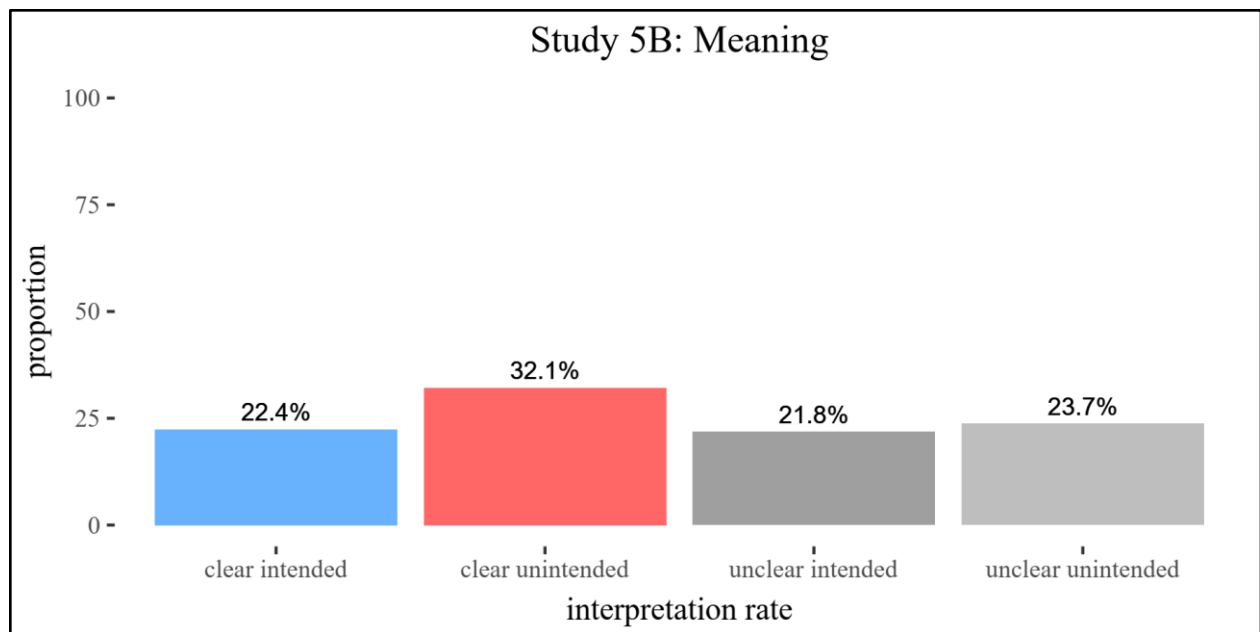
⁸⁰ I aggregated responses across all items rather than analyzing each separately because there were too few participants per item (14-17 per item). Future studies should assess response rates for individual items.

4.10.2 Study 5B: Meaning

As expected, most participants did not interpret items in a clear intended way when summing all conditions. Overall, 21.8% ($n = 34$) of participants provided a clear intended interpretation. This was significantly less than 0.5, $\chi^2(1, N = 156) = 47.41, p < 0.001$. With 95% confidence, the percentage of responses coded as clearly interpreted was less than 28.37%. Across all items, a majority of participants did not interpret those in a clear intended way, with the proportion of participants who provided clear intended responses ranging from 5.9% (item 1, $n = 17$) to 42.9% (item 6, $n = 14$). Overall, clear unintended interpretations were high, with 32.1% of participants across all items expressing a clear unintended interpretation ($n = 50$). Clear unintended interpretations were moderate to high. Results for individual items are available on **Figure 4.7**.

Figure 4.7

Interpretation rates for Study 5B (aggregated)



Discussion

Consistent with previous studies, most participants did not tend to interpret items on the MRS in a clear intended way. However, they did perform *better* than previous studies, and as we will see, for all subsequent studies. This may be attributed to the many steps taken to enhance the validity of the items used in the MRS. Notably, the MRS is the only scale to have undergone a rigorous validation process and the benefits may be revealed in the improved rate of clear intended interpretations]. Nevertheless, clear intended rates were still very low, and clear unintended interpretation rates were still very high. If anything, the low rate of clear intended interpretation rates even following such strenuous efforts to create a valid set of measures is a *stronger* indication that participants struggle to interpret such questions as intended, and provides somewhat better support for metaethical indeterminacy.

4.11 Study 6: The New Metaethics Questionnaire (NMQ)

Study 6 employs the same methods as Study 5 to assess items on a 3-item scale (YB3) that appears in Yilmaz & Bahçekapili (2015b) and the New Meta-ethics Questionnaire (NMQ) employed in Yilmaz & Bahçekapili (2015a). Interpretation rates for all items in Study 6 were coded exclusively by myself. The same pattern of extremely low clear intended rates for all conditions (which ranged from 11.4% to 22.5%), and all proportions of clear interpreted interpretations were significantly less than 0.5. For full discussion, see **Supplement 4**, section **S4.4.2**.

4.12 Study 7: Folk Moral Objectivism scale (FMO)

Study 7 employs the same methods as Study 5 and Study 6 to assess items on the Folk Moral Objectivism (FMO) scale that appears in Zijlstra (2019). Interpretation rates for all items in Study 7 were coded exclusively by myself. The same pattern of extremely low clear intended rates for all conditions (which ranged from 12.4% to 19.8%), and all proportions of clear interpreted interpretations were significantly less than 0.5. For full discussion, see **Supplement 4**, section **S4.4.3**.

4.13 General discussion

Over the course of this chapter, I have endeavored to provide a comprehensive body of evidence in support of two conclusions:

- (1) Most participants do not understand questions about metaethics as intended
- (2) This could be explained by widespread folk metaethical indeterminacy: people do not understand questions about metaethics because they don't have metaethical stances or commitments

It is far easier to provide evidence for (1) than for (2). Studies 1A through 1C reveal that people frequently do not attribute descriptions of a moral disagreement to differences in moral standards, even though this is not only the intended interpretation, but a necessary condition for the validity of the paradigm. Study 2 showed that when people are asked to assess other people's interpretations of questions about metaethics, many still fail to interpret those questions as intended. This study included both a response to a version of the disagreement paradigm, and direct questions about realism and relativism, expanding the range of questions in addition to providing a novel approach to assessing how people interpret questions about metaethics. One advantage to this study is that it circumvents the concern that questions about metaethics are ambiguous. By providing a response to such questions that was specifically designed to reflect an intended interpretation, and to thereby reflect the kinds of metaethical rationale consistent with an intended interpretation, the fact that participants still reliably fail to interpret these questions as intended provides converging evidence that participants do not interpret questions about metaethics as intended.

The low rate of clear intended interpretations and the comparatively high rate of clear unintended interpretations recurred across all studies: Study 3 demonstrated that participants struggle to interpret questions about metaethics in the concrete context of charitable giving, Study 4 showed that participants did not interpret the terms "objective" and "relative" as intended in both concrete and abstract cases, suggesting that studies that employ explicit metaethical language are likely failing

to reliably prompt intended interpretations. Finally, Studies 5, 6, and 7 show that participants do not interpret the items used in a variety of scales designed to assess folk metaethics as intended. When asked to explain why they agreed or disagreed with items intended to reflect their metaethical stances or commitments, the reasons given by most participants did not clearly indicate that they interpreted these questions as intended. Supporting this low rate of intended interpretations, asking participants what the items on these scales meant likewise revealed that few participants could clearly articulate what these statements meant.

Taken together, these findings represent multiple, converging lines of evidence across a wide variety of measures, all of which suggest that most ordinary people do not interpret questions about metaethics as intended. This conclusion is supported by thematic analysis, as well, which revealed recurring themes among unintended interpretations that made sense of and could readily account for why interpretation rates were so low. For instance, many participants appeared to conflate questions about relativism with descriptive claims about the differences in the moral standards of different people or cultures. This is precisely the kind of unintended interpretation we would expect of people engaging with the stimuli but failing to interpret it as intended (rather than, e.g., random, disorganized responses that might instead suggest disinterest or inattention).

At the very least, these findings challenge the validity of the disagreement paradigm, explicit paradigms (e.g., Fisher et al., 2017), and existing scales used to assess folk metaethics. Nevertheless, there are considerable limitations to these findings. Before we can even address the degree to which these results support (2) metaethical indeterminacy, we must first acknowledge and, where possible, address these limitations.

Limitations

Given the unusual nature of my analysis, I have opted to provide a more extensive section on limitations than is typical. See **Supplement 4**. Here, I will provide a brief outline of the main limitations of the studies reported here.

- (1) *Generalizability*. All studies were conducted on US participants, mostly on MTurk. This significantly limits the generalizability of the results reported here.
- (2) *Lack of demographic data*. In several instances I did not gather demographic data. This prohibits analysis of potential demographic differences within samples.
- (3) *Potentially poor response quality*. Most studies were conducted in contexts in which participants may have had suboptimal incentive to engage with content. This may have resulted in reduced response effort, which could compromise the quality of the data.
- (4) *Risk of bias and error*. With the exception of Study 1A, all data was coded exclusively by myself. This introduces significant risk of bias, idiosyncrasy, and error.
- (5) *Introspective access*. Participants may have competence with metaethical concepts but are unable to convey that competence in a research context. If so, these findings may overestimate the number of unintended interpretations.
- (6) *Accounting clear intended interpretation rates*. A significant minority of participants do offer clear intended interpretations. This must be explained in a way consistent with indeterminacy for indeterminacy to remain a viable hypothesis.
- (7) *Tension between indeterminacy and validity*. Low clear intended interpretation rates could be explained by *poor validity* or *indeterminacy*. However, the less valid a study, the worse indeterminacy serves as an explanation. Since I argue for both, there is some tension in my explanation of findings.
- (8) *Untested paradigms*. Interpretation rates for some measures have not yet been tested. In addition, some of those that have been tested had few participants per item.

Future directions

Like the limitations outlined above, I will only briefly reference future directions here. See **Supplement 4** for a fuller discussion.

- (1) *Acquire additional coders*: The most important step is to identify people competent to code the same datasets. Especially adversarial coders disinclined to share my biases.
- (2) *Improve generalizability*: All studies were of MTurk participants or college students in the United States. Future research should be directed at non-WEIRD populations and subpopulations of interest, e.g., insular religious communities.
- (3) *Improve rigor*: Participants could be assessed under more rigorous conditions, e.g., greater incentives, less time constraints, detailed instructions, or following training paradigms.
- (4) *New or more detailed questions*: small changes to the present method could be made, e.g., researchers could ask different questions or request longer and more detailed responses.
- (5) *Larger sample size for scale items*: Validity of scale items should be tested with a larger number of responses per item to better estimate interpretation rates for individual items.
- (6) *Comparison to expert populations*: The same measures could be tested among expert populations to assess whether interpretation rates substantially improve.
- (7) *Explore nonmoral domains*: Similar methods could be used to assess interpretation rates for realism/antirealism about other domains, including aesthetics, personal preferences (in e.g., food, music), social conventions, prudential norms, epistemic norms, and descriptive claims (e.g., science, math, religion), among others.
- (8) *Employ richer qualitative methods*: Researchers could conduct a richer variety of qualitative measures of folk metaethics, such as interviews.
- (9) *Employ corpus methods*: Researchers could employ corpus methods to assess large bodies of real world text. This could be used to assess e.g., the proportion of metaethical language used, the contexts in which metaethical claims are made, and whether people employ metaethical language consistently and in a way consistent with academic philosophy (see Chartrand, 2022 for discussion on corpus methods in experimental philosophy). This has recently been done (Stojanovic, 2022)

4.14 Conclusion

In this chapter, I introduced a novel method for assessing interpretation rates for questions about metaethics. This mixed methods approach involves the acquisition of open response data, coupled with a coding scheme designed to assess the overall rate at which participants appear to clearly interpret an item in an intended or unintended way. Across all studies, significantly fewer than half of participants interpreted questions about metaethics as intended. Extremely low rates of intended interpretations were found for every major scale and paradigm tested, as well as a variety of novel paradigms introduced by my colleagues and me. This suggests that existing methods for measuring folk metaethical stances and commitments are not valid. If most people do not know what they are being asked, and if a sizable minority, often accounting for around half of the participants, appear to have clearly interpreted what they were asked in an unintended way, results will consist largely of noise.

I have argued that there are two main factors contributing to the low rate of intended interpretations: many of the stimuli and questions are ambiguous or exhibit low face validity. As a result, many of the measures used to assess folk metaethics are not valid measures of folk metaethics. Second, I have argued that ordinary people don't have determinate metaethical stances or commitments, and that this partially explains the low rate of clear intended interpretations. Even correcting for ambiguity and reducing the poor face validity will be insufficient: people continue to interpret questions and statements about metaethics in a variety of unrelated ways even under conditions in which I have sought to clarify the relevant metaethical concepts. Far from existing in tension with one another, a large part of the reason researchers struggle to convey metaethical concepts in ordinary language, and a large part of the reason why interpretative variation, and unintended interpretations, remain high is *because* people lack the conceptual resources to respond to questions about metaethics appropriately. That items routinely fail to adequately convey metaethical concepts may be due, in part, to correctable researcher error. But such errors are themselves a

symptom of a deeper issue: metaethical distinctions are sophisticated, technically, metaphysically top-heavy notions that are simply absent from the way ordinary people speak and think.

Unfortunately, qualitative data is unlikely, on its own, to convince optimists about the prospects of empirical folk metaethics or anyone else who may be skeptical of my approach. In the next chapter I turn to a handful of quantitative approaches to shed further light on folk metaethics, which go some way in providing corroborating evidence in support of metaethical indeterminacy and the invalidity of existing measures of folk metaethics.

CHAPTER 5: **Multiple Choice & Indeterminacy Paradigms**

5.0 Introduction

My goal in this chapter is to provide converging evidence that most ordinary people don't interpret questions about metaethics as intended. This chapter is divided into two sections. First, I employ multiple choice questions and Likert scale items designed to assess how people interpret questions about metaethics. This approach provides a straightforward quantitative alternative to assess interpretation rates for questions about metaethics. The second section presents participants with a set of questions about nonmoral issues: in particular, their views on how to interpret quantum mechanics and mathematical platonism. The goal of these studies is to illustrate that people can appear to express a determinate philosophical stance towards an issue even when they plausibly lack such a position. Furthermore, these studies illustrate *spontaneous theorizing* by illustrating how forced choice paradigms create the artificial appearance of a determinate position where none existed prior to participating in the study.

5.1 Multiple choice paradigms

In the previous chapter, I assessed interpretation rates using open response questions. Written responses provide a rich body of data that can provide insights not only into interpretation rates, but the particular ways participants interpreted various metaethics stimuli. Yet there are limitations to this approach.

One shortcoming with open response questions is that people may have an implicit competence with metaethical concepts, but the structure and wording of open response questions is insufficient to render those concepts salient in a way that manifests in people's responses. Using open response questions to evaluate how people interpret questions about metaethics potentially suffer

from the very problem they are ostensibly designed to assess: people may interpret open response questions in unintended ways. If so, participants coded as having unclear or unintended responses may be competent with the relevant metaethical distinctions, but fail to demonstrate this. Ironically, then, open response studies may suffer from the same shortcoming as the measures they were designed to debunk.

It is also possible that ordinary people may lack the relevant vocabulary to clearly convey the relevant metaethical concepts, even if they have them. For instance, a person might recognize the distinction between stance-independent and stance-dependent facts, but lack a clear or recognizable terminology to express this. In study 2 of the previous chapter, I asked participants to explain what a *response* to a question about metaethics meant, rather than the question itself. This goes some way in mitigating these concerns, since the response uses precisely the kinds of vocabulary I (and hopefully other competent coders) would consider a clear way of conveying the relevant metaethical distinctions. Yet this doesn't go far enough. There are no guarantees that people will interpret a response to questions about metaethical questions as intended, either.

There is no easy solution to these problems. Nor is any one method likely to decisively settle the matter. Our best bet is to provide mutually corroborating lines of evidence using a variety of distinct methods and theoretical approaches that converge on the same conclusions. To this end, I developed a simple and straightforward set of studies that eschew coding open response in favor of measures subject to conventional quantitative analysis. In all studies, participants were presented with a *target statement* that reflected a realist or antirealist metaethical position. Then they were presented with a set of statements. *One* of these statements matched the intended meaning of the target statement (the *matched statement*), while five statements did not match the meaning of the statement, but were designed to appear as plausible candidate meanings (the *decoy statements*). These decoy statements were designed to reflect the common ways ordinary people are inclined to interpret statements about

metaethics in unintended ways. Matched and decoy statements are presented in three formats. In studies 1A through 1D, I used a multiple choice question. Finally, in study 1E, I use a Likert scale that allows participants to express how much they agree or disagree that both the matched and decoy statements match the meaning of the target statement.

The rationale behind these studies is simple: if people interpret statements reflecting metaethical distinctions as intended, we should expect them to be able to match the meaning of these statements with other statements that convey the same meaning as the target statement. If only a small portion of people correctly match the target statement with the matched statement, this suggests that people struggle to interpret the target statement as intended. All target statements were drawn directly from stimuli used in previous studies, were adapted from the stimuli used in previous studies, or were designed to convey to my satisfaction the relevant metaethical distinctions. Thus, if few people match the target statement with the matched statement, this challenges the validity of existing measures and raises doubts about the ease with which researchers can measure folk metaethical stances and commitments using conventional survey methods.

5.1.1 Study 1A: Moral relativism scale (MRS) items

Currently, the Moral Relativism Scale (MRS) is the most rigorous and well-developed metaethics scale (Collier-Spruel et al., 2019). Collier-Spruel and colleagues chose from a pool of items rated by a team of experts for accurately representing relativism (or its negation) and that exhibited conventional psychometric virtues, e.g., loading well onto a single factor, had a high Cronbach's alpha, and that exhibited adequate predictive and convergent validity, since scores on the MRS were associated with tolerance and other predicted associations, and with other measures of relativism, respectively (but see Maul, 2017). This made it an ideal candidate for an initial attempt to assess interpretation rates using a multiple choice paradigm.

The MRS consists of ten statements. Participants express how much they agree or disagree with each statement on a five-point scale. Seven of the items express relativism in some form, while three items express realism, universalism, or both, and are reverse-coded, such that a low score on these items is interpreted as a greater relativism score. A composite of one's response to all ten questions may be taken as one's overall endorsement of moral relativism. If these items are valid measures of relativism, then participants presented with items on the MRS should be able to match the meaning of those items with another statement that also expresses relativism (or the rejection of relativism). I predicted that participants would struggle to match items from the MRS with their intended interpretation, and that fewer than half would choose the correct response.

Methods

Participants. Participants consisted of 201 adult US residents on Amazons' Mechanical Turk. I intended to recruit 200 participants from Amazon's Mechanical Turk. 206 participants began the study. However, 5 did not complete the study, and were dropped from subsequent analysis.

Procedure. Across all conditions, participants were first presented with a set of instructions that explained what they'd be asked to do in the study. These instructions appeared on a page without stimuli to encourage participants to read the instructions carefully. Then they were randomly assigned to one of the conditions. Each of these conditions repeated the same instructions, which explained that they would be presented with a statement and asked to choose the statement that had the same meaning, *not* to choose a statement they agreed with the most.

Then all participants were presented with a statement in quotes and bold text. This statement, and the response options that followed, differed across conditions. They were then asked to choose the statement that "is the best interpretation of the meaning of the statement," and were given a set of six statements to choose from. Order for statements was randomized. Two conditions presented a

statement expressing relativism, and two expressed statements intended to express the rejection of relativism. Here is a sample item (the correct answer is underlined):

Below, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

"The viewpoint of one's culture determines whether their actions are morally right."

- ☐ The truth about whether an action is morally right can only be judged according to the standards of different cultures
- ☐ People from different cultures can be justified in holding different moral beliefs
- ☐ Each culture adapts its views about what is morally right to fit with their traditions and meet local needs
- ☐ Since each culture has different views about what is morally right, a person's actions will be judged as moral by some cultures and immoral by other cultures
- ☐ Different cultures have the right to hold different moral beliefs, so it isn't acceptable to impose our moral values on them
- ☐ The attitudes of each person's culture influence whether their actions are judged as moral or immoral

As this sample item illustrates, participants could choose one of six response options. One item reflects the intended interpretation (the *correct* response), while the other five do not reflect an intended interpretation (*incorrect* responses). The correct and incorrect responses are unique to each item to provide more context and appear as relevant as possible to the statement used. The *incorrect* responses were intended to reflect a variety of confluations I anticipated would be most plausible for the target statement. For instance, in order from top to bottom, the five *incorrect* responses listed above include: (1) epistemic claim that we're *justified* in holding contrary moral beliefs (*epistemic conflation*), (2) the notion that different people conform to the same moral rules in different ways (*evaluative standard ambiguity*), (3) a descriptive claim about how people will be judged (*descriptive conflation*), (4) a normative claim

about people being entitled to their beliefs (*normative conflation*), and (5) a causal or genealogical claim. Although not originally anticipated, (5) was a very common theme that emerged from analyzing open response questions. The same attention was given to tailoring appropriate alternatives to all other statements, in both this and all subsequent studies.

There are some difficulties with providing an accurate response option for the reverse-coded items that were intended to express the rejection of relativism. However, these problems were due to the inherent misoperationalization of these items. Both items were ambiguous and conveyed more than one concept. These were the items selected:

"There are moral rules that apply to everyone regardless of personal beliefs."

"The same moral standards should be followed by people from all cultures."

The first statement appears to convey *both* universalism *and* realism, since it describes moral rules "applying to everyone" (i.e., universalism) and doing so "regardless of personal beliefs" (i.e., realism). Unfortunately, *neither* the rejection of universalism *nor* the rejection of realism necessarily entail relativism.^{81,82} The second item conveys both universalism and a *normative* claim about what people "should" do. Rather than design a "correct" response option that accurately conveyed the full content of these items, I opted for response options that expressed realism since this is the standard

⁸¹ The contrast to realism, understood as stance-independence, would be stance-dependence, but stance-dependence and relativism are not identical and don't entail one another. Conversely, universalism could be construed as the view that there is a single correct moral system rather than multiple correct moral systems. This may be a more appropriate contrast, and future studies should assess interpretation when the correct response conveys universalism exclusively. Unfortunately, *if* that were the intended contrast, then this item isn't face valid because it includes a remark suggesting realism. Note that this item was approved by a panel of experts, despite conflating realism and universalism. This isn't a critique. I'd like to think I'm an expert and I've conflated realism and universalism in my own research as well. Rather, I want to draw attention to the fact that people who invest their professional careers in this topic *still* conflate these distinctions. Are we really expecting participants with no experience and little incentive to think carefully about this topic to avoid these confluations?

⁸² Note also that agreeing with this item is consistent with cultural relativism, since it only states that rules apply regardless of personal beliefs, not the standards of one's culture. The other statement conveys a rejection of cultural relativism. It's worth noting that rejecting subjectivism doesn't entail universalism or realism, since you could be a cultural relativist, while rejecting cultural relativism does not entail universalism or realism, since you could be an individual subjectivist. Thus, neither of these items accurately captures a proper continuum between relativism and the rejection of relativism.

contrasting concept presented in the literature. However, an argument could be made that the optimal contrast would be some form of universalism (or more specifically some type of non-indexicalism).⁸³

Our primary goal was to assess whether most participants would choose the correct response. This sets the threshold at 50%, which is remarkably low. If only half of the participants in a study interpreted an item as intended, this would raise serious questions about whether the measure in question was valid. Of course, it is possible that the “correct” response is ambiguous or confusing or difficult to interpret. If so, participants who interpreted the original statement as intended may still choose an incorrect response. By setting the bar very low, I hoped to buffer against such errors. Nevertheless, I did my best to select a correct interpretation that accurately reflected the intended metaethical position even if the original statement did not do a great job of conveying that position.

Measures. The primary measure consisted of a single multiple choice question with six response options. The *correct* response option matched the meaning of the target statement. The remaining five *incorrect* response options did not match the meaning of the target statement. All responses were coded as either “1” = correct or “0” = incorrect. The four target statements may be seen below. Each set of six response options was individualized to the particular target statement.

Relativism 1

Different people can have opposing views on what is moral and immoral without anyone being wrong.

Relativism 2

The viewpoint of one’s culture determines whether their actions are morally right.

Realism 1

There are moral rules that apply to everyone regardless of personal beliefs.

Realism 2

The same moral standards should be followed by people from all cultures.

⁸³ E.g., I might have used something like “there is a single correct standard of moral values” as a correct response.

Results

A one sample proportion test without continuity correction was conducted for each of the four conditions to determine whether the proportion of participants who selected the correct response was significantly less than 0.5. For Relativism 1, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 50) = 13.52, p < 0.001$.⁸⁴ With 95% confidence, the percentage of correct responses was less than 35.10%. For Relativism 2, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 51) = 18.84, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 30.18%. For Realism 1, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 48) = 18.75, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 29.59%. For Realism 2, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 52) = 15.08, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 33.87%. The frequency and percentage of correct responses is featured on **Table 5.1**.

Discussion

Across all four conditions, less than half of participants successfully matched an item from the MRS with its intended metaethical interpretation. This held for items representing relativism and the rejection of relativism. In fact, participants did only marginally better than chance (i.e., 16.67%).⁸⁵

⁸⁴ As noted in chapter 4 (where I provide a similar comment), all one sample proportion tests were conducted in R using `prop.test()`. However, rather than conducting a one sample proportion test directly, `prop.test()` conducts a chi-squared test. Fortunately in these cases χ^2 is equivalent to the square of the Z-statistic one would obtain by conducting a one sample proportion test using the standard formula. To demonstrate this, note that the formula for a one proportion Z-test is $Z = (p - p_0) / \sqrt{p_0(1 - p_0)/n}$. For this study, this would give us $= (0.24 - 0.5) / \sqrt{(0.5(1 - 0.5)/50)} = -3.676955$. $(-3.676955)^2 = 13.52$, which is equivalent to the χ^2 reported in the main text.

⁸⁵ And for Relativism 1, the pattern of responses was not significantly different from chance. A chi-square goodness-of-fit test was performed to determine whether people were equally likely to choose each of the six responses. The pattern of responses was not significantly different from an equal distribution across all responses for Relativism 1, $\chi^2(5, N = 50) = 5.20, p = 0.392$. However, given the relatively small number of participants per condition this study is probably underpowered.

These findings strongly indicate that very few participants matched the selected items from the MRS scale with statements intended to accurately reflect the intended metaethical concept.

Table 5.1

Proportion of correct responses (by condition)

Item	Correct		Incorrect		Total
	%	n	%	n	
Relativism 1	24.0	12	76.0	38	50
Relativism 2	19.6	10	80.4	41	51
Realism 1	18.8	9	81.2	39	48
Realism 2	23.1	12	76.9	40	52

Note. $n = 201$.

5.1.2 Study 1B: Fisher et al.'s explicit measures

Fisher et al. (2017) attempt to assess folk realism by *explicitly* asking people whether morality is objective. Participants were presented with questions about a variety of topics and asked whether they think there is an “objectively true” answer to those questions. Here is the wording that they used:

Consider the following question: “Should [topic] be allowed?” Please tell us whether you think there is an objectively true answer to this question. (p. 1127)

Responses were measured using a 7-point Likert scale (1 = definitely no objective truth, 7 = definitely an objective truth).⁸⁶ Topics included a variety of issues, including “same-sex marriage, marijuana legalization, teaching evolution in school, abortion, and violence in videogames” (p. 1126). According

⁸⁶ Note that technically this is a measure of *confidence* in realism, not whether the participant endorses realism or not. It’s not clear that someone who doesn’t think realism is “definitely true” is *less of a realist* than someone who does. I remain puzzled that researchers uncritically treat degrees of epistemic confidence as an indication of “how much” you believe something. A person can be extremely committed to a view but be hesitant to maintain that it’s “definitely” true. For instance, a passionate religious believer may be open to doubt and choose a 6, while an ambivalent religious person could choose 7 because they don’t have any doubts. It’s not clear what it would mean to say the second person believes “more” or “more strongly.”

to Fisher et al., one advantage of this approach is that “it removes any ambiguity about whether we really are asking about whether there is an objective truth about the topic” (p. 1127).

Unfortunately, this is a shockingly naive approach to “removing ambiguity.” Far from removing ambiguity, asking people whether there is an “objective truth” to a moral issue presumes that participants will interpret this term in a way consistent with researcher intent and with one another. As I demonstrated in chapter 4, they do not. “Objective” is an ambiguous phrase that has multiple possible meanings. Far from eliminating ambiguity, Fisher et al. amplified it. The term “objective,” as it is used in philosophy, is a piece of technical jargon with a narrow meaning. Everyday uses of “objective” include notions such as “unbiased,” “capable of being measured using publicly available criteria,” and so on. The mistake Fisher and others make is to imagine that ordinary people’s conception of “objective” matches the technical way the term is used in academic philosophy.

The goal of this study was to assess whether people would reliably match the meaning of an adapted version of the stimuli Fisher et al. (2017) used. Since participants were explicitly asked to select the best interpretation of the meaning of a statement, it would have been unnecessarily complicated to use Fisher et al.’s exact stimuli, since it included a question within a broader statement. Asking them to interpret a statement within a statement would be unnecessarily cognitively demanding. “Consider the following question” also seemed redundant and unnecessary for assessing whether people interpreted Fisher et al.’s stimuli as intended. To resolve these concerns, I adapted the wording of the original stimuli to make the statement as clear as possible. I chose one of the moral issues that appeared in Fisher et al.’s study, frontloaded an explicit metaethical statement, and then brought in the concrete moral issue I selected and used the phrase “objectively true” to match the precise term used by Fisher et al. Although Fisher and colleagues did not employ a similar question for relativism, I also devised a novel question that explicitly asked about moral relativism as well. Once again, my goal was to assess whether most participants would match the meaning of these statements

to another statement that expressed realism and relativism, rather than a statement that conveyed some other unintended concept.

Methods

Participants. Participants consisted of 301 adult US residents on Amazon's Mechanical Turk (124 females, 176 males, 1 other, $M_{\text{age}} = 38.2$, $SD_{\text{age}} = 11.4$, age range = 19-74). I intended to recruit 300 participants from Amazon's Mechanical Turk. 302 participants began the study. However, one participant did not complete any of the measures, and was dropped from subsequent analysis.

Procedure. This study employed the same procedure as Study 1A. All participants were presented with the same instructions, which clarified that they would be asked to match the target statement with the statement that most closely matched its meaning. Participants were then randomly assigned to either the realism or relativism condition. They were given reminder instructions, the target statement, and a set of six response options, where they could select the statement that most closely matched the meaning of the target statement. Finally, participants were presented with the comprehension check as study 1B and I collected data on age and gender.

Measures. The primary measure consisted of a single multiple choice question with six response options. The *correct* response option matched the meaning of the target statement. The remaining five *incorrect* response options did not match the meaning of the target statement. All responses were coded as either "1" = correct or "0" = incorrect. In addition, a comprehension check was introduced to assess whether participants interpreted the task as intended (i.e., to select the statement that matched the meaning of the target statement). The A sample item with the correct response is underlined is presented below:

In response to the question on the previous page, which of the following were you doing?

- ☐ Selecting the single statement which I think best matches the meaning of the bolded sentence
- ☐ Selecting the single statement that I most agree with
- ☐ Selecting every statement that I agree with
- ☐ Selecting every statement which I think best matches the meaning of the bolded sentence

Explicit | Realism

“Moral truth is objective, so there is an objectively true answer to whether violence should be seen as an appropriate response in conflicts.”

Explicit | Relativism

“Moral truth is relative, so the truth about whether violence should be seen as an appropriate response in conflicts is relative.”

Finally, participants were given a comprehension check and demographic data (age and gender) was collected.

Results

A one sample proportion test without continuity correction was conducted for both conditions to test whether the proportion of participants who selected the correct response was significantly less than 0.5. For the realism condition, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 149) = 53.16, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 26.05%. For the relativism condition, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 152) = 96.61, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 16.09%. The frequency and percentage of correct responses is featured on **Table 5.2**. 80.1% ($N = 241$) of participants passed the comprehension check. Excluding participants who did not pass the comprehension check did not meaningfully change any results.

Table 5.2*Proportion of correct responses (by condition)*

Item	Correct		Incorrect		Total
	%	n	%	n	
Explicit Realism	20.1	30	79.9	119	149
Explicit Relativism	11.2	17	88.8	135	152

Note. $n = 301$.**Discussion**

Once again, most people did not interpret questions about metaethics as intended. In this case, participants did not interpret statements explicitly conveying the notion that morality is “objective” or “relative” in a way consistent with the metaethical positions these terms are supposed to correspond to. These results were found in a larger sample than previous studies, and were not meaningfully influenced by the failure of many participants to pass a comprehension check. This casts serious doubt on Fisher et al.’s claim that they successfully eliminated ambiguity by asking whether morality is objective. Although it was not a part of Fisher et al.’s measures, people also struggled to understand what it would mean to say that moral truth is relative. Even if people struggle to interpret explicit use of metaethical terms as intended, they may still interpret other, indirect ways of expressing realism and antirealism. In the next study, I assess an attempt to prime participants with realism and antirealism. Once again, my goal is to assess whether participants match the statement with its intended metaethical meaning. Of course, I only assessed interpretation rates for a single item (violence as a method of resolving conflicts), so it remains possible participants would interpret “objectively true” differently in reference to some other moral issue, or if the item were worded differently. However, absent any positive indication that people do interpret terms like “objective” and “relative” as intended, the onus remains on those who do believe participants interpret explicit questions about whether morality is objective or relative as intended to furnish supporting evidence.

5.1.3 Study 1C: Young and Durwin's measures

Young and Durwin (2013) conducted a study that assessed whether priming people with realism and antirealism would influence charitable giving. The details of that study are unimportant for our purposes. Our focus is whether the priming conditions that they used could plausibly serve as methods for inducing belief in realism and antirealism. Y&D employed two conditions with metaethics primes. In both conditions, a canvasser approached people on the street, pitched a charity, and attempted to solicit donations for the charity. In the course of this pitch, canvassers used either a realism or antirealism prime that consisted of asking participants a simple question:

Realism: “Do you agree that some things are just morally right or wrong, good or bad, wherever you happen to be from in the world?”

Antirealism: “Do you agree that our morals and values are shaped by our culture and upbringing, so there are no absolute right answers to any moral questions?” (p. 303)

One obvious problem with these items is that they are not face valid representations of realism or antirealism. The realism condition would be better interpreted as an expression of universalism. It could also be read as a descriptive claim, i.e., that people *consider* some of the same things to be right or wrong everywhere in the world. The antirealism item is much worse. It actually contains two claims: an uncontroversial descriptive claim about the origins of our moral standards that most people would probably agree with, coupled with the denial of “absolute” right answers to moral questions. Unfortunately, “absolute” does not clearly convey stance-independence, but could instead mean all sorts of things: an undeniable answer, an exceptionless moral rule, and so on. Nevertheless, it’s still possible that participants would match these priming conditions with the appropriate metaethical stance. The items used in this study were taken verbatim, with the sole exception of replacing the question mark with a period and treating them as statements rather than questions. Once again, my goal was to assess whether most participants would match the target statement to the correct metaethical position.

Methods

Participants. Participants consisted of 202 US residents on Amazon's Mechanical Turk (91 females, 110 males, 1 other, $M_{\text{age}} = 41.1$, $SD_{\text{age}} = 11.8$, age range = 22-71). I intended to recruit 200 participants from Amazon's Mechanical Turk. 202 participants completed the survey.

Procedure. I employed the same measures as those employed in Study 1B and Study 1C, using new target statements and response options.

Measures. I employed the same measures as Study 1B: all participants were given one target statement, and were asked to choose which of the six response options matched the meaning of the target statement. This was followed by a comprehension check and demographic questions.

Y&D | Realism

Some things are just morally right or wrong, good or bad, wherever you happen to be from in the world.

Y&D | Antirealism

Our morals and values are shaped by our culture and upbringing, so there are no absolute right answers to any moral question.

Results

A one sample proportion test without continuity correction was conducted for both conditions to test whether the proportion of participants who selected the correct response was significantly less than 0.5. For the realism condition, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 100) = 27.04, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 31.65%. For the antirealism condition, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 102) = 28.59, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 31.06%. The frequency and percentage of correct responses is featured on **Table 5.3**. 82.7% ($N = 167$) of participants passed the comprehension check. Excluding participants who did not pass the comprehension check did not meaningfully change any results.

Table 5.3*Proportion of correct responses (by condition)*

Item	Correct		Incorrect		Total
	%	n	%	n	
YB3 Realism	24.0	24	76.0	76	100
YB3 Antirealism	23.5	24	76.5	78	102

Note. $n = 202$.**Discussion**

Once again, most participants did not interpret questions about realism and antirealism as intended. Notably, a very high proportion of participants chose specific incorrect responses. In the realism condition, 37% of participants chose the response option “There are some actions that are considered morally wrong in all societies.” In the antirealism condition, 49% of participants chose the incorrect response option, “Each person’s views about what is morally right or wrong is shaped by their unique background and experiences, so we frequently reach different conclusions about moral questions.” Both of these claims are *descriptive*, with the latter including an additional claim about the etiology of moral belief: that we reach different conclusions *because* of our different backgrounds and experiences. This suggests that participants don’t find the target statements so confusing that they simply guess at their meaning, which would typically result in a more random response pattern. Rather, they appear to systematically favor incorrect response options. This lends support to the conclusion that people reliably conflate claims about metaethics with unintended concepts.

5.1.4 Study 1D: Bush and Moss’s face valid metaethical items

So far, all of the target statements were drawn from or based on existing folk metaethics paradigms. However, these items all suffer very poor face validity. As a result, we could explain the low rate of intended interpretations as an indication not that people don’t have determinate metaethical stances or commitments, but as evidence that researchers have consistently failed to properly operationalize

folk metaethics stimuli. To address this possibility, I adapted items that had been developed for an unpublished folk metaethics scale and used these as target statements. My goal was to craft items that would reflect more face valid representations of realism and relativism. I opted for relativism rather than a more general way of expressing antirealism since it seemed less plausible that people would interpret the denial of stance-independent moral facts as intended as readily as they'd interpret a statement expressing relativism. That is, I wanted to choose items that I thought had a reasonable chance of enjoying a higher rate of intended interpretation. I chose the following two items:

BeM | Realism

There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.

BeM | Antirealism

The truth of all moral claims can vary depending on the moral standards of different individuals or cultures.

Methods

Participants. Participants consisted of 201 US residents on Amazon's Mechanical Turk (88 females, 113 males, $M_{\text{age}} = 40.0$, $SD_{\text{age}} = 11.9$, age range = 19-81). We intended to recruit 200 participants from Amazon's Mechanical Turk. 201 participants completed the survey.

Procedure. I employed the same measures as those employed in previous studies, using new target statements and response options.

Measures. I employed the same measures as Studies 1B and 1C: all participants were given one target statement, and were asked to choose which of the six response options matched the meaning of the target statement. This was followed by a comprehension check and demographic questions.

Results

A one sample proportion test without continuity correction was conducted for both conditions to test whether the proportion of participants who selected the correct response was significantly less than 0.5. For the realism condition, the proportion of participants who selected the correct response was

significantly less than 0.5, $\chi^2(1, N = 100) = 4.00, p < 0.023$. With 95% confidence, the percentage of correct responses was less than 48.22%. For the antirealism condition, the proportion of participants who selected the correct response was significantly less than 0.5, $\chi^2(1, N = 101) = 64.96, p < 0.001$. With 95% confidence, the percentage of correct responses was less than 15.88%. The frequency and percentage of correct responses is featured on **Table 5.4**. 87.1% ($n = 175$) of participants passed the comprehension check. Excluding participants who did not pass the comprehension check, 41.8% ($n = 38$) of participants chose the correct response, which resulted in the realism condition no longer being significantly less than 0.5, $\chi^2(1, N = 91) = 2.47, p = 0.058$. With 95% confidence, the percentage of correct responses was less than 50.38%. No other results were meaningfully changed.

Table 5.4

Proportion of correct responses (by condition)

Item	Correct		Incorrect		Total
	%	n	%	n	
B&M Realism	40.0	40	60.0	60	100
B&M Relativism	9.9	10	90.1	91	101

Note. $n = 201$.

Discussion

Most participants in the relativism condition did not interpret the target statement as intended. Likewise, most participants in the realism condition did not interpret the target statement as intended, unless I removed participants who failed the comprehension check ($N = 10$). Once I did so, I could no longer reject the null hypothesis that the proportion of people who select the correct response is equal to or greater than 0.50. This suggests that it's possible that the true proportion of people who interpret the realism condition as intended is equal to or greater than half. I discuss this possibility in the general discussion. In short, this outcome is plausibly due to inadequate power to detect small

differences. With the upper bound of the confidence interval just barely over 50%, these results are not good evidence most people interpreted the item as intended. Overall, these findings are consistent with the conclusion that people do not reliably interpret questions about metaethics as intended. Even if around half of the participants did interpret one of the items as intended, such an item would still be a valid measure of realism. And if this is the best we can achieve, then efforts to convey realism to ordinary people via single-sentence measures are probably hopeless.

5.1.5 Study 1E: Likert scale paradigm

One shortcoming with multiple choice paradigms is that they require participants to choose just one of the response options available. There are at least three limitations with measuring interpretation rates this way. First, it is possible that if participants believe more than one of the response options matches the meaning of the target statements. Second, it is possible that participants believe *none* of the response options match the meaning of the target statement, but if so they have no way to express this. Third, this method does not allow us to assess *how well* one or more statements match the meaning of the target statement. It is possible, for instance, that some participants only slightly agree that a particular response option matches the meaning of the target statement, while others strongly agree that it does so. Requiring participants to express a categorical judgment prevents us from assessing such variability.

Study 1E addresses all of these shortcomings by presenting participants with a Likert scale that allows them to express level of agreement with each response option separately. This allows participants to judge that more than one of the response options match the meaning of the target statement, or that none of them do, and it also allows them to express level of agreement. This study drew on four items from the MRS (two realism, two antirealism) that were used in Study 1A and the two items I created that were used in Study 1D (one realism, one antirealism). These six items (three

realism, three antirealism) served as the target statements. All three antirealist conditions represented either cultural relativism or subjectivism.

In addition to assessing how well the response options originally devised for the multiple versions of these items matched the meaning of the target statement (the one correct *matched* response and five incorrect *decoy* responses), I also asked participants how well each of the other two alternative target statements reflecting the same metaethical position (i.e., either realism or antirealism) matched the meaning of the target statement, in order to assess whether participants would judge that items ostensibly intended to measure the same psychological construct would mean the same thing.⁸⁷ I will refer to these as the *alternative* statements.

It's possible that participants would be significantly more likely to judge that the target statement matched the alternative statements and the matched statement than the decoy statements. This would not be a good indication of the validity of the target statement, however. Such a tendency could be significant but marginal, and a marginal tendency to favor the intended interpretations of an item would not be an indication that a measure is valid, only that it performs better than chance. As such, my expectation was that there would be no reliable tendency, overall, for participants to demonstrate an *overwhelming* tendency to judge that the target statement would match the meaning of the alternative statements or the matched (i.e., correct) statement, when compared to the decoy (i.e., incorrect) statements. Instead, I expected the general pattern to be one in which participants

⁸⁷ Note that realism and antirealism could be depicted as complex constructs with multiple subdimensions (e.g., antirealism could be divided into noncognitivism, relativism, and error theory), so it's possible participants may not think that items that express e.g., individual subjectivism mean the same thing as cultural relativism. Such judgments would be reasonable, since individual subjectivism and cultural relativism *don't* mean the same thing. Notably, the scales typically used to measure people's metaethical attitudes don't draw such distinctions, and instead treat measures of e.g., cultural relativism and individual subjectivism as measures of the same underlying construct. For instance, Collier-Spruel et al. (2019) use the following two items as measures of "relativism":

- (1) *The viewpoint of one's culture determines whether their actions are morally right*
- (2) *Each person is the final authority on whether his or her actions are morally correct*

Taken at face value, these items are not consistent. If cultures determine moral truth, individuals can't be the final authority, and vice versa.

demonstrated no clear and decisive tendency to match the meaning of the target statement to the other target statements or the correct response, when compared to incorrect responses.

Methods

Participants. Participants consisted of 604 US residents on Amazon's Mechanical Turk (313 females, 288 males, 2 other, $M_{\text{age}} = 42.1$, $SD_{\text{age}} = 13.0$, age range = 20-80). I planned to collect 600 participants in total (100 participants per condition). A total of 604 participants completed the survey.

Procedure. Each participant was assigned to one of six conditions. In each condition, participants were presented with one target statement. They were asked to judge "the degree to which you agree or disagree that the additional statements match the meaning" of the target statement, for each of eight statements (1 = Strongly disagree, 7 = Strongly agree). For the realism conditions, these eight statements consisted of the other two realist target statements, the correct matching statement, and five incorrect statements. For the antirealism conditions, the eight statements consisted of the other two antirealist target statements, the correct matching statement, and five incorrect statements. For each condition, the correct and incorrect statements were designed to most closely reflect the meaning of that statement (for the correct statement) and to most closely reflect plausible unintended interpretations.

Measures. Measures consisted of level of agreement with the eight statements described above on a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree), a comprehension check, and demographic questions. The comprehension check appeared after the main measures and assessed whether participants interpreted the task as intended (i.e., that their level of agreement indicated how well each statement matched the meaning of the target statement, *not* how much they agreed with that statement).

Results

For each condition, I ran linear mixed model analysis in Jamovi (version 2.3.13) using the GAMLj module (version 2.6.6) to estimate differences in levels of agreement between the statements participants were asked to match to the target statement (this included both alternative target statements, the matched statement, and the five incorrect “decoy” statements). For instance, for condition 1 I included r_1 as the dependent variable, where r_1 is level of agreement that each of the statements matches the meaning of the target statement and added the target statement (c_1) as a fixed effect. I included participant (ID) as a random effect (random intercept only). The model specification was as follows: $r_1 \sim 1 + c_1 + (1 | ID)$. The same model was used for conditions 2 through 5. Using the GAMLj module, I also assessed the estimated marginal means and 95% confidence intervals for responses to each of the eight statements. Linear mixed models are often used for more sophisticated designs. However, as Magezi (2015) observes, “it is important to realize that the use of LMMs is by no means restricted to complex grouping designs, and can also be used for experimental psychology studies with a single grouping factor of participant or subject” (p. 2; see also Singmann & Kellen, 2019).

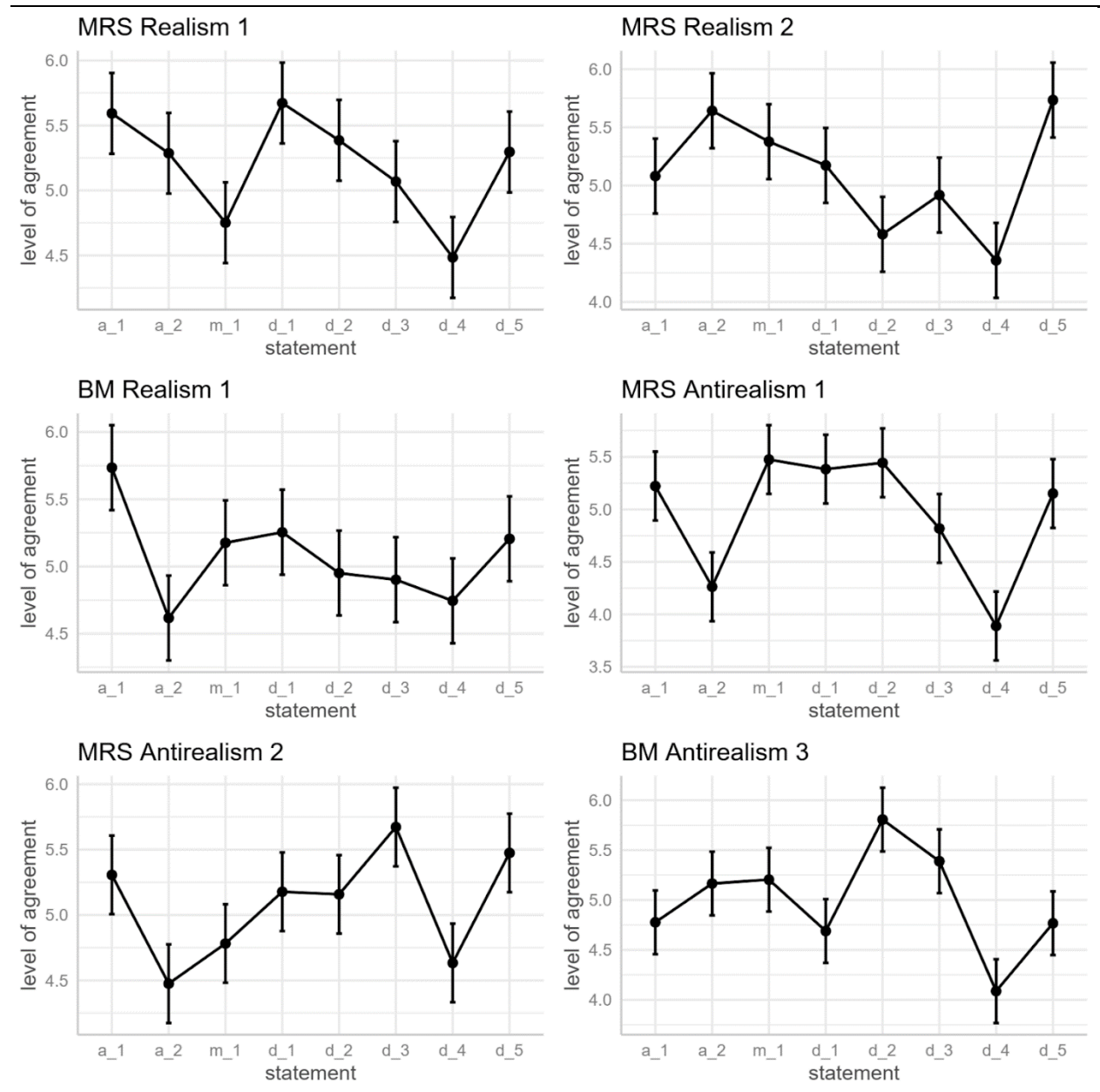
Although a linear mixed model would allow tests of whether agreement with the other target statements and the correct response were significantly greater than the incorrect responses, as I noted above, the goal of this analysis is primarily descriptive and exploratory (see above), and is primarily intended to provide evidence that participants do not overwhelmingly match target statements that reflect a particular metaethical position (e.g., realism) with responses that match their meaning (i.e., the other target statements and the correct response). As such, a significant difference between mean response levels for these statements and incorrect statements would not, by itself, be informative. Nevertheless, I conducted pairwise comparisons both without correction and with Bonferroni correction for all pairs of statements. Results are available in **Supplement 5**. Finally, since the same

general pattern occurred across all conditions, I report only the results for Condition 1 in the main text, and provide results for the remaining five conditions in **Supplement 5**.

Graphs displaying the estimated marginal means with 95% confidence intervals are featured in **Figure 5.1**. This figure displays the estimated marginal mean level of agreement for each of the eight statements across all six conditions along with 95% confidence intervals. Means above 4 indicate a tendency to agree that the statement matched the meaning of the target statement, while means below 4 indicate a tendency to disagree that the statement matched the meaning of the target statement. Moving from left to right for each condition, the first two items along the x-axis (a_1 and a_2) are the alternative target statements that are intended to reflect the same metaethical position (i.e., either realism or antirealism). The next item, m_1 , is the matched statement, which is intended to reflect the intended interpretation of the target statement. Finally, d_1 through d_5 are the decoy statements, which reflect unintended interpretations. Thus, higher mean scores for a_1 , a_2 , and m_1 indicate that participants correctly matched the target statement to its intended interpretations, while higher mean scores for d_1 through d_5 indicate that participants incorrectly matched the meaning of the target statement to unintended interpretations. At a glance, it is clear that participants showed no general tendency to overwhelmingly favor the intended interpretations (the alternative or matched statements) over the unintended interpretations (the decoy statements) for *any* of the target statements. The closest any condition came was for BM Realism 1. The estimated marginal mean for a_1 was 5.74, 95% CI [5.42, 6.05]. However, this overlaps with d_1 and d_5 . For every other condition, the alternative target statements and the matched statement did not overwhelmingly outperform other items in level of agreement. Indeed, the alternative and matched statements often underperformed compared to decoy statements.

Figure 5.1

Estimated marginal means with 95% CIs (by condition)



Note. Going from left to right, “a_1” and “a_2” refer to the two alternative target statements, “m_1” refers to the *matching* (or correct) statement, and “d_1” through “d_5” refer to the *decoy* (or incorrect) statements. As such, higher scores for the first three items are indicative of interpreting the target statement as intended, while higher scores for the five items on the right indicate a failure to interpret target statements as intended.

For MRS Realism 1, MRS Realism 2, MRS Antirealism 2, and BM Antirealism 3, the statement with the highest estimated marginal mean level of agreement was one of the decoy statements, while for MRS Antirealism 1 the difference between the matched statement and the first and second decoy statements (i.e., d_1 and d_2) was negligible. However, one of the most noteworthy patterns across all conditions is that participants *tended to agree that every statement matched the meaning of the target statement*. Across 48 opportunities to judge how well a statement matched the meaning of the target statement, only *one* item fell below the midpoint, d_4 in MRS Antirealism 1, and only just barely ($M = 3.89$). This may be attributable in part to acquiescence (Podsakoff et al., 2003). If so, it may be that a general tendency to agree obscures evidence that participants interpreted target statements as intended.

Participants did not perform well on the comprehension check. Only 69.2% ($n = 418$) participants chose the correct response, suggesting that many participants may not have performed the task as intended, and may have, e.g., been expressing their level of agreement with statements, rather than judging how well statements matched the meaning of the target statement. However, results did not substantially change when such participants were excluded from analysis.

Under ideal conditions, we would expect participants to agree that the alternative and matched statements match the meaning of the target statements, and to *disagree* that the decoy statements match the meaning of the target statement. Yet this is simply not what we find. A consistent tendency to agree that decoy statements match the meaning of the target statement threatens the validity of the target statements as measures of people's metaethical stance because it indicates that people do not interpret such statements as *only* expressing a metaethical claim. If an item on a metaethics scale is interpreted to express both a metaethical claim *and* one or more non-metaethical claims (e.g., descriptive or normative claims), then we'd be unable to know whether people's level of agreement with such items can be attributed to their metaethical views, or to the unintended non-metaethical content they attribute to the claim. In short, these findings provide further evidence that people do

not reliably interpret items intended to reflect metaethical claims as intended. Participants do not reliably interpret such claims as exclusively *metaethical* claims. Instead, they seem to judge that such items may also reflect a variety of unintended meanings.

Discussion

Once again, results indicate that participants do not appear to interpret items on metaethics scales as intended. In order for items on metaethics scales to serve as valid measures of realism and antirealism, participants should exhibit a reliable and overwhelming tendency to match the meaning of those items with statements that are intended to reflect the same metaethical concept, and a consistent tendency to recognize when a statement doesn't match the meaning of the item. Participants did not do this. Instead, they frequently judged that non-metaethical statements matched the meaning of the target statement. This provides further evidence that participants do not interpret items on metaethics scales as intended.

One limitation of these findings is that many participants failed the comprehension check. Although results did not substantially change when such participants were excluded from analysis, it's possible that many participants were expressing their level of agreement with the statements, rather than expressing how well those statements matched the meaning of the target statement. This could explain why participants exhibited a general tendency to agree with all statements. Although this may be due in part to acquiescence, it could be that the statements were, on average, the sorts of claims people tend to agree with. If so, it's worth noting that participants were evenly divided between realist and antirealist conditions, yet they exhibited a similar tendency to agree with items in both conditions. If participants tended to agree with statements in both realist and antirealist conditions, this would suggest a general tendency to endorse statements that reflect realism *and* to endorse statements that reflect antirealism, at least when they are presented in isolation. It may be the inconsistency between

statements reflecting realism and antirealism are more salient, or only become salient at all, when they are presented side-by-side.

Future research could improve on this design by making the nature of the task more explicit to participants. For example, this could be achieved by restating the purpose of the task for each item, e.g., participants could be asked to express their level of agreement that “[statement] means the same thing as [target statement]” or they could be asked for each statement how well it matches the meaning of the target statement, with a Likert scale ranging from 1 = Very poorly to 7 = Very well. Participants could also be given a training exercise on an unrelated topic at the start of the task to test their understanding. There may be still other ways to ensure participants interpret the task as intended.

5.2 Study 2: Indeterminacy in non-metaethical domains

In the previous section, I demonstrated that participants reliably fail to match statements that express realism, antirealism, and relativism to other statements that are intended to represent the same concepts. These findings suggest that people struggle to interpret folk metaethical stimuli as intended. However, this only serves to corroborate findings from the previous chapter: people don’t interpret questions about metaethics as intended. In this last study, I show that it’s possible to generate the superficial appearance of genuine patterns of folk philosophical stances even where it is unlikely any exist, and propose that such results are best explained as a result of *spontaneous theorizing*. Evidence of spontaneous theorizing in non-metaethical domains supports the possibility that spontaneous theorizing could also explain apparent patterns of realism and antirealism in folk metaethics research.

Anyone familiar with research on folk metaethics will observe that we find consistent, replicable patterns of results. If metaethical indeterminacy is true, this poses a challenge. If metaethics stimuli were completely unfathomable, we might expect an approximately *random* distribution of responses, or at least a pattern that could be readily explained by factors other than the construct of interest, e.g., acquiescence bias (Podsakoff et al., 2003).

Yet this isn't what we find. How can the hypothesis that ordinary people have no determinate metaethical stances or commitments account for this? The previous chapters provided evidence that most folk metaethics research relies on invalid measures. As a result, the pattern of results we observe is consistent with people having determinate metaethical stances or commitments, but systematically interpreting metaethical stimuli in unintended ways. In other words, existing findings are consistent with the possibility that people have determinate metaethical stances or commitments, but existing studies have simply failed to measure them.

I have yet to develop a method for ruling this explanation out. It may be that no simple set of studies could do so, and that it will take many converging lines of evidence to resolve the matter. Yet training paradigms (e.g., Wright, 2018; Pölzler & Wright, 2020a; 2020b; see **Chapter 3**) represent one method for breaking through this impasse. If we can ensure that participants interpret what we're asking as intended, then we can be sure that their answers genuinely reflect their metaethical stances or commitments because we've deliberately constructed studies that successfully *cause* the intended interpretation.

Unfortunately, this trades one dilemma for another: once we induce people to think about folk metaethics, their responses could reflect a stance or commitment they developed as a result of the training, rather than a stance or commitment they held prior to participating in the study. If so, then we could not make inferences about these participants to the population they were drawn from, since they would no longer be members of that population. Recall that the target population of folk metaethics is ordinary people. Ordinary people are, by definition, people without adequate philosophical training. How could we study such a population by recruiting people from that population, and then providing them with adequate philosophical training? That would make no more sense than attempting to study people who've never been to Japan by flying them to Japan for the study.

But suppose Pölzler and Wright (2020a; 2020b) convincingly argue that training paradigms can successfully induce participants to interpret questions about metaethics as intended, but don't require providing enough philosophical training that participants no longer qualify as ordinary people. That is, suppose we are dealing with a population of *ordinary people who understand questions about metaethics as intended*. When faced with such a situation, we can no longer dismiss results as the product of unintended interpretations. We'd have to accept that whatever answers people provide *do* reflect determinate metaethical stances or commitments.

Regrettably, even this would be insufficient to conclude that ordinary people have determinate metaethical stances or commitments. This is because training paradigms point to another possibility: participants may have held no determinate stance prior to participating in a study, but form a position in virtue of participating in the study. That is, they may engage in *spontaneous theorizing*. Spontaneous theorizing occurs *whenever a participant who held no determinate stance or commitment prior to participating in a study is induced to develop or express a stance or commitment due to the experience of participating in the study itself*.

This is not to say people genuinely develop a full-fledged philosophical position in the span of a few minutes that they will carry with them for the rest of their lives. Such theorizing may be transient: the context of a given task may prompt a person to favor one position over another and respond accordingly for the purposes of the task, yet this may represent only a fleeting inclination towards the view on superficial consideration, could quickly be buried in one's unconscious or forgotten altogether, and in any case may make no meaningful difference in how a person thinks, speaks, or acts after participating in the study. That is, spontaneous theorizing doesn't require that people become full-blown converts to a particular position; they need only think about it *enough* that they form enough of a position on the matter to meaningfully respond to stimuli, regardless of whether this has any long term psychological implications for their stances or commitments.

Spontaneous theorizing poses a serious threat to training paradigms. If we cannot tell whether responses reflect stances or commitments held prior to participating in a study, or were formed during the course of the study, training paradigms may lack external validity since we cannot be sure whether the pattern of responses obtained in a training paradigm tells us anything about the metaethical views of people outside the context of the study. Yet spontaneous theorizing also poses a threat to folk metaethics research in general. Such research typically relies on forced choice paradigms: participants are *required* to express a realist or antirealist stance. As a result, all possible responses are interpreted that participants hold a particular metaethical stance. When a participant has no recourse other than to select one of the available response options, even if they have no determinate position and even if they do not interpret the stimuli as intended, they may be disposed to respond in nonrandom ways. This may result in a systematic disposition to favor some response options over others, independent of whether those response options accurately reflect the putative psychological construct researchers are studying. Such inclinations may be too insignificant to reflect a determinate stance or commitment, e.g., superficial features of one response option may make it more appealing than another, such as one statement seeming more affirmative or positive (“Yes, we have free will”) than the alternative (“No, we don’t have free will”), even if the participant doesn’t understand what’s meant by “free will.” Yet it’s also possible that participants *do* interpretation questions as intended at least *some* of the time. When this occurs, conventional folk metaethics research cannot distinguish instances of spontaneous theorizing from instances in which a participant expressed a stance or commitment held prior to exposure to the study stimuli.

In the previous chapter, we saw that clear intended interpretation rates are not *literally zero*. Some people do appear to interpret questions about metaethics as intended. And while we cannot tell which people interpreted questions as intended rather than simply guessed correctly in the studies reported in this chapter, it’s possible that at least *some* participants interpreted statements about realism,

antirealism, and relativism as intended. While some of these cases may represent genuine instances of determinate folk metaethical views held prior to participation, some may also be the result of spontaneous theorizing. This would make sense of why many people don't interpret metaethical stimuli as intended, but *some* do, without requiring us to propose that *everyone* who appears to interpret questions about metaethics as intended has a determinate metaethical view.

Spontaneous theorizing can explain why at least some people appear to have determinate stances or commitments in the context of a study: because some people *do* have a determinate stance or commitment. Yet this explanation is consistent with folk indeterminacy. So long as the determinate stances or commitments result from participating in the study, they will have no necessary implications for how people think *outside* the context of the study. The goal of this final study is to illustrate that even in instances where, on *a priori* grounds, we have good reason to believe people don't genuinely understand certain concepts and distinctions, they may nevertheless develop a perspective on the matter via participation in the study.

This study also serves as a rebuttal to a recent critique of metaethical indeterminacy put forward by Pözlner and Wright (2020a).⁸⁸ Pözlner and Wright intend to reject indeterminacy with

⁸⁸ Their characterization of indeterminism strikes me as questionable, or at least incomplete:

"Indeterminism: Moral sentences and judgements are indeterminate with regard to the cognitivism/non-cognitivism distinction. That is, they are correctly analyzed by both theories (like men with a certain amount of hair may be correctly described both as being bald and as not being bald)" (p. 23).

My concern is with the notion that competing theories provide a *correct* analysis. This makes it seem like both accounts *win*, when it might be more apt to say the accounts are *not more explanatorily adequate than the other*. In his initial formulation, Gill (2009) characterized indeterminacy as the claim "that some parts of ordinary moral discourse give us *no reason to prefer* an analysis that involves one meta ethical commitment over an analysis that involves the commitment that has traditionally been taken to be its meta-ethical competitor" (p. 216). This does *not* sound like the claim that both analyses are literally correct. For comparison, suppose two scientific hypotheses were equally consistent with available evidence. Would we conclude that both hypotheses were *correct*? No. We'd simply conclude that both were equally adequate explanations. Gill (2009) is even more clear about this when he elaborates on indeterminacy later in the article:

"According to the Indeterminacy Thesis (which is the "I" of the IV Thesis), many parts of our moral thought and language provide no good answers to the questions that were central to much of 20th century meta-ethics, vindicating neither relativism nor absolutism, neither internalism nor externalism, etc. The Indeterminacy Thesis holds that the relationship between some instances of ordinary moral discourse and these meta-ethical debates is analogous to the relationship between ordinary arithmetic and debates in the philosophy of mathematics. There is no fact of the matter as to whether

respect to whether ordinary people are cognitivists or noncognitivists, rather than realists or antirealists but their critique applies equally well to the realism/antirealism dispute. In particular, they claim that “Many versions of indeterminism, hybrid expressivism, and incoherentism predict that ‘truth-apt’ and ‘not truth-apt’ classifications will be distributed fairly equally for each statement” (p. 24). Yet they argue, because their results showed that a significant majority of participants judged every statement to be truth-apt, the “distribution of responses was quite unequal” (p. 24). In particular, they found that about three quarters of their participants favored noncognitivist response options across conditions. Since this unequal distribution is inconsistent with what we’d expect if indeterminism were correct, their findings purportedly cast doubt on indeterminism.

There aren’t “many versions” of indeterminism or incoherentism, so it’s a bit puzzling for them to say that *many* versions of these accounts would make such predictions. Yet the more troubling problem feature of these claims is that it’s not at all clear that indeterminism does, in fact, predict that responses to questions about metaethics would be “fairly equally distributed” for any given set of statements. An equal distribution is what we might expect if people were completely indifferent between two accounts, and had no inclination whatsoever for one position over another. The result would be that nothing would bias participants towards selecting one response over another. However, we should only expect an equal distribution, in practice, if no superficial or extraneous features of a response option would systematically incline participants to favor one position over its alternative. Such conditions *might* be met in practice. Yet superficial or extraneous features of stimuli may, in practice, prompt a disproportionate number of participants to favor one position over another. Even

ordinary mathematic usage is better explained by a Platonist or anti-Platonist conception of number. The way people use numbers in everyday math simply does not contain answers to the questions that animate philosophy of mathematics.” (p. 218)

It is clear that Gill does not take indeterminacy to entail both accounts are correct, but that *neither* account is correct. Unfortunately, Pölzler and Wright have mischaracterized indeterminism.

if we would predict an equal distribution under *ideal* circumstances, where all features of the stimuli irrelevant to what is being measured are held constant across response options, *conditions are never actually ideal*.

Psychological characteristics of a given population may systematically incline a subset of a given sample (and, by extension, the population they're drawn from) towards one response over another merely due to features of the participants unrelated to the construct of interest. For instance, acquiescence bias is the well-documented “tendency to answer affirmatively to a question no matter what it’s content” (Knowles & Nathan, 1997). Even in instances where people hold no particular position towards the stimuli, they may still exhibit a general tendency to favor affirmative responses. This holds *even when the stimuli in question are meaningless, uninterpretable, or could not possibly represent any legitimate psychological construct*.

Maul provides one demonstration of this, showing that reliable patterns emerged from a set of scale items even when those items asked about nonsense concepts (e.g., *quintessence*), consisted of lorem ipsum, or were simply blank (that is, no question was even asked). Here, I provide another example. Nelson (2013) describes an amusing demonstration of such spurious patterns in a collaboration between Nelson and Meyvis. In the course of developing a manipulation, they presented participants with a set of 20 pictures of animals, and asked them to rate each in terms of its speed “goodness.” As Nelson observes, “The latter could be best construed as an evaluation of moral worth. *That is an absurd question*” (emphasis mine). Despite the absurdity of the question, participants answered. After all, as Nelson points out:

In surveys, most people answer most questions. That is true regardless of whether or not questions are coherently constructed and reasonably articulated. That means that absurd questions still receive answers, *and in part because humans are similar to one another, those answers can even look peculiarly consistent*. (Emphasis mine)

They found a number of amusing results. For instance, people judged tortoises to be the most morally virtuous. However, the most important result was that almost every animal was judged, on average, to be good. Their “scale had a meaningful midpoint, yet all but three animals are above it.” Apparently, hyenas, barracudas, and jellyfish were the most diabolical of the set of animals. Interestingly, they also recruited two experts (an ecologist and an evolutionary biologist) to respond to the scale as well. Critically, both judged that “any response would be random,” with one stating that “I would probably tie them all in ranking.” Of course, this isn’t what happened, and both experts likewise gave seemingly nonrandom responses that even correlated with one another ($r = 0.29$). In fact, both agreed with ordinary people that tortoises are highly virtuous.

Notably, both experts expected a response pattern, perhaps because the questions were nonsensical, when this clearly turned out not to be the case. I remain puzzled as to why researchers would think that people would respond randomly to questions, even when the questions are absurd. Simply because a question is nonsensical, it doesn’t follow that people will just throw their hands up and choose random responses or uniformly select the midpoint if one is available. While this might hold if stimuli are completely inscrutable, people are still generally motivated to engage with a study, and will often do their best to interpret stimuli in *some* way, even if this requires substituting what’s being asked for a more meaningful question. For instance, people may judge tortoises as more virtuous because they are more likable or less dangerous. Cultural associations may also influence people’s responses. Tortoises are associated in popular culture with patience (*The Tortoise and the Hare*), and, after all, patience is a virtue. Hyenas, in contrast, are portrayed as villains in iconic movies such as *The Lion King*, and are often depicted as malicious and unappealing.⁸⁹ Such associations may influence

⁸⁹ A report on the conservation status of hyenas notes that “they are viewed with contempt and fear” and observe that “Tourists do not rate hyenas very highly” (Hofer & Mills, 1998).

people's responses even if people don't literally believe any of these animals are actually more or less virtuous.

This example illustrates how meaningless or nonsensical questions can nevertheless result in meaningful response patterns. However, I wanted to construct a study that demonstrated a nonrandom response pattern that more closely resembled the stimuli used in folk metaethics research, and, in particular, the training paradigms employed by Pölzler and Wright. My goal was to show that if you employ measures that are intended to measure philosophical stances or commitments that couldn't plausibly figure into the way ordinary people speak or think, that you can still obtain results in which a disproportionate number of people favor one position over another.

I opted for competing interpretations of quantum mechanics. Quantum mechanics is a highly successful theory that describes the nature of atomic and subatomic phenomena. It is widely regarded as one of the most advanced, mathematically intense, and philosophically impenetrable fields of study, featuring mind-bendingly complicated equations and seemingly intractable puzzles about the fundamental nature of reality that have vexed many of the world's greatest thinkers over the past century. However precise and replicable its equations may prove, there is still the question of how to make sense of the various subatomic events captured by these equations. While some theorists wish simply to focus on the math, others are concerned about the best way to interpret quantum mechanics, with many taking the view that it ought to provide a "literal description of reality" (de Muynck, 2004, p. 92). This is not the place for an introduction to the topic, nor would I be competent to do so. For our purposes, it will suffice to say that interpretations of quantum mechanics revolve around extraordinarily complex issues in mathematics and philosophy that prompted Feynman to quip that "I can safely say that nobody understands quantum mechanics" (p. 123). In a popular comment on the remark, Mainwood (2020) points out that the quote was delivered during a lecture. According to Mainwood:

Feynman opens this lecture by musing on the fact that as science has progressed, it has moved further and further from our intuitive experience of everyday objects; and that while this is unsurprising given how limited our everyday experience is, it makes the work of understanding and explanation more and more difficult.

Mainwood's point is that Feynman is not claiming that literally nobody understands quantum mechanics, but rather that we cannot understand quantum mechanics "in terms of a simple, familiar model." This is because quantum mechanics represents a broader tendency for scientific discoveries to move further from everyday experience over time. In short, quantum mechanics has radically counterintuitive and bewildering implications.

Given its radically counterintuitive nature, it is incredibly implausible that an implicit commitment to any particular interpretation is implicit the way ordinary people speak or think. And it is nearly as implausible that any appreciable number of ordinary people have studied quantum mechanics well enough to have any particular stance on the matter. As such, it is an ideal candidate for folk indeterminism: there is simply no good reason to think most people have a determinate stance on the matter. In short, we should presume that most ordinary people have no determinate stance on how to interpret quantum mechanics. Nevertheless, you could present participants with a description of different interpretations of quantum mechanics, and a forced choice between two or more response options. If the pattern of responses is not equally distributed across different interpretations, this would indicate that responses could be unequally distributed across response options even if we have strong reasons to suspect ordinary people don't have a determinate stance or commitment on the topic.

This is exactly what I did. I decided to narrow my focus to the two most prominent interpretations of quantum mechanics: the Copenhagen and Many Worlds interpretation. Participants were given a brief explanation of what quantum mechanics was, and a description of each interpretation. Then they were asked to decide which interpretation they thought was correct. Since

the Many Worlds interpretation (MWI) suggests that the universe is constantly bifurcating into countless other universes, I expected ordinary people to find this far more bizarre and implausible than the Copenhagen interpretation, which treats outcomes as probabilistic. Although I am skeptical of the measures used in studies on free will, such studies do suggest that people tend to reject determinism (Nadelhoffer et al., 2014). Even in the absence of data, it strikes me as more plausible that, if confronted with the one of two possibilities: that some events in our universe have random, indeterminate outcomes that cannot be predicted in advance, or that the universe is *literally* in a perpetual state of spawning countless slightly different copies of itself, each of which is populated by everything from slightly different versions of themselves to worlds with honest-to-gosh mustache-twirling supervillain versions of themselves, I figured the former would strike most people as more plausible.

As such, I predicted that a significant proportion of people would favor the Copenhagen interpretation over the Many Worlds interpretation. In addition, I predicted that most people would be unfamiliar with quantum mechanics, would report having no prior position on the matter, would state that quantum mechanics rarely or never comes up in everyday conversations, and that many people would agree that the structure of the study compelled them to respond in ways that didn't reflect their actual position, suggesting that the forced choice paradigm prohibited them from conveying what they actually thought. I asked a handful of other questions, mostly for exploratory purposes, including whether most ordinary people have an opinion on how to interpret quantum mechanics, whether the way ordinary people speak presupposes a particular interpretation, and whether beliefs about quantum mechanics have any practical significance.

5.4.1 Study 2: Quantum mechanics

Methods

Participants. Participants consisted of 200 US residents on Amazon's Mechanical Turk (55 females, 144 males, 1 other, $M_{\text{age}} = 37.3$, $SD_{\text{age}} = 10.3$, age range = 20-72).

Procedure. All participants were presented with the same conditions and measures. Participants were presented instructions indicating that they would be presented with different perspectives on an issue, but that the issue may be unfamiliar to them and that no prior knowledge is expected or necessary. This was intended to make it clear that they were not expected to have any prior knowledge of quantum mechanics, which may have discouraged precisely those participants of interest from continuing with the study (i.e., people who aren't familiar with quantum mechanics). Next, participants were given five questions to assess their knowledge and familiarity with quantum mechanics and to assess whether they have an opinion on which interpretation of quantum mechanics is correct. Then participants were presented with a description of the Copenhagen and Many Worlds interpretations of quantum mechanics:

Quantum Mechanics

Quantum mechanics is a branch of physics that deals with atomic and subatomic phenomena. It is a highly successful theory that makes precise predictions and has inspired new technologies. Despite this success, scientists disagree about how to interpret it as a literal description of the world.

One disagreement concerns the nature of some subatomic events. These events appear to have more than one possible outcome, but when we measure them, we only observe one outcome. There are two competing explanations for what occurs when we observe one of these outcomes.

The Copenhagen Interpretation of Quantum Mechanics

According to the Copenhagen interpretation, we only observe one outcome because there is only one outcome. According to this view, such events have probabilistic outcomes. For example, a subatomic particle could have a 50% chance of moving to the left, and a 50% chance of moving to the right. We will not know which way it will move until we observe it. Once we observe it, there is an equal chance it ends up moving in either direction, like a coin flip. There is no way to know whether the particle would go to the left or the right in advance,

because the result is random. Thus, the Copenhagen interpretation holds that some subatomic events have random outcomes.

The Many Worlds Interpretation of Quantum Mechanics

According to the Many Worlds interpretation, we observe one outcome but there is actually more than one outcome. According to this view, events do not have probabilistic outcomes. Whenever a subatomic particle could move to either the left or the right, instead of it only moving in one or the other direction, the universe splits into two different universes. The particle moves to the left in one of the universes, and to the right in the other universe. Each of those universes contains an exact copy of us and everything else in the universe, with the only difference being the direction the particle moves. In one universe, we observe the particle move to the left, while in the other, we observe the particle move to the right. The reason the particle seems to have moved to either the left or the right is because we are in one of these universes, and cannot observe the other universe. But there is a copy of us in the other universe that saw the particle move in the opposite direction. As a result, the outcome of subatomic events is never random. Instead, all possible ways the particle could move are realized in different universes. Since many subatomic events occur every second, there are countless other universes, each different from the one you are in.

Participants were asked to select which of these interpretations was correct, a series of questions about the presence and practical relevance of different perspectives of quantum mechanics in everyday life. Finally, participants were given a set of six comprehension checks, were asked whether they felt forced to select one of the two interpretations despite it not representing their actual views, and demographic data was collected.

Measures. The main measure was a multiple choice question with two response options. Participants were asked “Which interpretation of quantum mechanics do you think is correct?” and were given the following response options:

The Copenhagen interpretation is correct

Whenever a subatomic event appears to have more than one possible outcome, only one of those outcomes randomly occurs.

The Many Worlds interpretation is correct

Whenever a subatomic event appears to have more than one possible outcome, all possible outcomes are realized. The universe splits into a different universe for each possible outcome, and each of those possibilities occurs in one of those universes.

The familiarity questions included (a) a checkbox featuring eight technical terms related to quantum mechanics (e.g., “quantum harmonic oscillator,” “finite potential well”). Participants were asked to select all terms they were familiar with (b) a question about their highest level of education in physics (ranging from “none” to “graduate degree in physics (MS, PhD, etc.),” (c) how well they understand quantum mechanics (1 = Very poorly, 7 = Very well), (d) their level of expertise in quantum mechanics (ranging from “no knowledge” to “expert”), and whether they have an opinion on which interpretation is correct (this question proceeded the description of the Copenhagen and Many Worlds interpretations and the main measures, which appeared on the next page).

After the main measure, participants were asked how often the topic of quantum mechanics comes up in everyday life (“never (It has never come up in conversation)” to “Frequently (several times a week)”) and whether most ordinary people have an opinion on how to interpret quantum mechanics (with three options (1) Yes, Copenhagen, (2) Yes, Many Worlds, or (3) No). Participants were also asked whether ordinary people presuppose the Copenhagen or Many Worlds interpretations when they speak (again with the option to say yes and choose one of the two interpretations, or no), and how much of an impact that opinions on quantum mechanics have on everyday life (from “no impact” to “enormous impact”).

Comprehension checks consisted of six true/false questions designed to assess whether participants understood the description of quantum mechanics and the Copenhagen and Many Worlds interpretations presented earlier in the study. All six questions are straightforward and simple questions that are based on the instructions. For instance, one item stated that, “According to the Copenhagen interpretation, some events have random outcomes.” Participants were then asked whether they felt forced to choose one of the two interpretations:

In this study, you were asked to choose between the Copenhagen interpretation and the Many Worlds interpretation. However, it is possible that you do not agree with either of these positions. Did you feel that you had to choose between one of these two answers, even though it did not accurately reflect what you thought?

- ☐ Yes
- ☐ No

Finally, I collected demographic information (age and gender).

Results

In order to assess the judgments of ordinary people, which in this case meant people without substantial training in physics, I excluded all participants who met one or more of the following characteristics from analyses:

- (1) Anyone with a bachelor's degree, graduate education, or a graduate degree in physics
- (2) Anyone who reported understanding quantum mechanics above the midpoint (i.e., anyone with a score of 5 or more on a 7-point Likert scale)
- (3) Anyone who reported having "proficient" or "expert" knowledge of quantum mechanics

13% ($n = 26$) of participants reported having a degree or graduate education in physics, 13.5% ($n = 27$) were above the midpoint for reported understanding of quantum mechanics, and 7% reported proficiency or expertise with respect to knowledge of quantum mechanics. Most of these responses overlapped, so this resulted in few exclusions. In total 15.5% ($n = 31$) chose at least one of these responses, and were excluded from subsequent analyses. However, all analyses were conducted without excluding these participants, which did not result in any meaningful changes. This resulted in a remaining pool of 161 participants.

A one sample proportion test without continuity correction was conducted to test whether the proportion of participants who selected each interpretation was significantly different than 0.5. The proportion of participants who selected the Copenhagen and Many Worlds interpretations was significantly different from 0.5, $\chi^2(1, N = 169) = 25.00, p < 0.001, 95\% \text{ CI } [0.69, 0.76]$. Overall, 69.2% ($n = 117$) of participants reported that the Copenhagen interpretation was correct, while 30.8% ($n =$

52) reported that the Many Worlds interpretation was correct. Participants who were not excluded due to their knowledge of physics overwhelmingly reported having no prior position, with 97% ($n = 164$) reporting that they had no opinion about which interpretation of quantum mechanics is correct.⁹⁰

53.3% ($n = 90$) of participants reported that quantum mechanics “never” comes up in everyday life, while another 34.9% ($n = 59$) reported that it “rarely” comes up. 11.2% ($n = 19$) reported that it sometimes comes up, 0.6% ($n = 1$) reported that it “often” comes up, and no participants reported that it “frequently” comes up. 72.2% ($n = 122$) reported that most ordinary people have no opinion about which interpretation of quantum mechanics is correct, while 21.3% ($n = 36$) reported that ordinary people favored the Copenhagen interpretation and 6.5% ($n = 11$) reported that ordinary people favored the Many Worlds interpretation. 69.2% ($n = 117$) reported that ordinary people don’t presuppose an interpretation of quantum mechanics when making claims about the past, present, or future, while 22.5% ($n = 38$) reported that reported that ordinary people presuppose the Copenhagen interpretation, and 8.3% ($n = 14$) reported that ordinary people presuppose the Many Worlds interpretation. 56.2% ($n = 95$) of participants reported that ordinary people’s opinions about how to interpret quantum mechanics have “No impact” on everyday life, 34.3% ($n = 58$) reported that opinions about quantum mechanics have “Very little impact,” 8.9% ($n = 15$) reported that opinions about quantum mechanics have “Some impact”, 0% ($n = 0$) reported a “moderate” impact, and 0.6% ($n = 1$) reported an “Enormous impact.” 34.3% ($n = 58$) of participants reported that they felt forced to choose one of the two interpretations, even though it did not accurately reflect what they thought.

Participants performed reasonably well for most comprehension checks with a majority of participants (71.6%, $n = 121$) getting five or six out of a total of six questions correct ($M = 4.99$, $SD = 1.13$). 43.2% ($n = 73$) got all six questions correct, 28.4% ($n = 48$) got five questions correct, 16.6%

⁹⁰ Obviously, this difference was significant $\chi^2(1, N = 169) = 149.59, p < 0.001$, 95% CI [0.93, 0.99]. It remained significant when including participants who were not excluded for knowledge or expertise, $\chi^2(1, N = 200) = 106.58, p < 0.001$, 95% CI [0.81, 0.91].

($n = 28$) got four correct, 8.9% ($n = 15$) got two correct, 2.4% ($n = 4$) got two correct, and 0.59% ($n = 1$) got 1 correct, and none got zero correct. 89.9% ($n = 152$) got the correct answer for Question 1, followed by 95.3% ($n = 161$) for Question 2, 76.3% ($n = 129$) for Question 3, 59.8% ($n = 101$) for Question 4, 85.2% ($n = 144$) for Question 5, and 92.9% ($n = 157$) for Question 6.⁹¹

Discussion

A substantial majority of ordinary people (nearly 70%) favored the Copenhagen over the Many Worlds interpretation, despite reporting no substantive training, understanding, or expertise in quantum mechanics and in spite overwhelmingly (97%) reporting that they had no opinion on how to interpret quantum mechanics prior to being asked to select which interpret they thought was correct. This provides a compelling illustration of how participants can be induced to express a stance on a position for which they held no position prior to participating in the study. It's not plausible that virtually all participants would report having no interpretation of quantum mechanics when in fact they did have an interpretation.

This may be a simple demonstration, but it has serious implications. This is because these findings present us with a dilemma: since participants were forced to express that either the Copenhagen or Many Worlds interpretation was correct, we have two options: (1) this does *not* genuinely reflect their views, in which case the forced choice nature of the paradigm has produced an invalid measure, or (2) it *does* represent their position on the matter, *but that position was formed seconds ago in the process of participating in the study*, i.e., it's a result of *spontaneous theorizing*. Each of these

⁹¹ Participants removed from analysis due to their self-reported knowledge or training in physics ($n = 31$) performed about the same on Question 1 (90.3%) and substantially worse on every other question. In order from Question 2 through 6, the proportion of correct responses was 71.0%, 48.4%, 38.7%, 45.2%, and 58.1%. I don't know what to make of this. It's possible that participants who reported expertise were overconfident, did not take the comprehension checks as seriously, or that participants who were more likely to fail comprehension checks were less engaged with the study, and the association between self-reported expertise or knowledge and poor performance on the comprehension checks is actually due to participants being inattentive or disengaged when responding to both sets of questions. If so, then at least some self-reported experts may have not actually been experts. Of course, 31 is also an incredibly small number so these proportions are all likely to be very noisy and not very informative.

possibilities could likewise reflect the results of folk metaethics research: responses may be invalid by forcing participants to choose responses that don't reflect their views *or* they could be engaged in spontaneous theorizing.

60% of participants reported that they did *not* feel forced to choose a response that did not accurately reflect their position, which was significantly different from 0.5, $\chi^2(1, N = 169) = 16.62, p < 0.001$, 95% CI [0.58, 0.72]. This result suggests that around two thirds of people may have engaged in spontaneous theorizing, with the remaining third forced to choose despite their response not accurately reflecting their views. Such self-reports should be interpreted with caution, since there may be considerable incentive due to e.g., self-presentation concerns, cognitive dissonance, or confabulation to report or mistakenly believe that one has formed a substantive stance on a topic even when one hasn't.

Nevertheless, these findings provide some evidence that when people are given the opportunity to read about a topic, they can readily form an opinion about it, even if they didn't previously hold an opinion. This seems like a sufficiently modest claim that it hardly requires evidence, yet this example still provides a clear illustration of the process occurring with respect to a complicated topic for which most people plausibly held no prior view, a topic adjacent in many ways to the kinds of paradigms employed in folk metaethics and folk philosophical research in general. If participants can feel compelled to answer questions even when response options don't represent their actual position with respect to quantum mechanics, the same could be true for metaethics as well. And if participants can spontaneously develop a position about how to interpret quantum mechanics after being introduced to the topic, they could potentially do so when responding to questions about metaethics as well, especially when they are participating in training paradigms that provide far more extensive instructions and explanations than what was provided in this study.

These results provide evidence that it is fairly easy to identify instances in which ordinary people are unlikely to have any particular position on a topic, but can be readily induced to adopt and express a position on that topic during the course of a study. This challenges Pölzler and Wright's (2020a) presumption that indeterminacy would predict an equal distribution across response options. To my knowledge, I'm the only full-fledged proponent of metaethical indeterminacy studying folk metaethics, and this is *not* what I predict.

These findings also illustrate how forced choice paradigms can coerce many people into expressing a position on a topic even if they don't feel their response accurately reflects their views. Even when over a third of participants responded in this way, a disproportionate number of people still favored the Copenhagen interpretation, as expected. If reliable patterns can emerge from data even when people feel compelled to express views that don't reflect what they think, the same could be true for many studies in folk philosophy, including folk metaethics. In fact, even if we focus only on participants who report that they felt forced to choose between the Copenhagen and Many Worlds interpretations even though neither accurately reflected their position, which represented 34.3% ($n = 58$) of the sample, 63.8% ($n = 37$) of participants still favored the Copenhagen interpretation, which was *still* significantly differed from 0.5, $\chi^2(1, N = 58) = 4.41, p < 0.036$, 95% CI [0.51, 0.75]. In other words, if we look only at the responses of people who claim that neither response option accurately reflected their views, they *still* exhibited an unequal response pattern.⁹² Given the wide variety of biases, errors in design, and factors that may favor a disproportionate number of people reliably favoring one position over another for a variety of philosophical issues, there is simply little reason to suppose that folk indeterminacy with respect to any given philosophical issues should result in responses that are eventually distributed across response options.

⁹² Results were about the same for participants who reported that they did not feel forced, with 72.1% ($n = 80$) favoring the Copenhagen response, which remained significantly different from 0.5, $\chi^2(1, N = 121) = 4.41, p < 0.001$, 95% CI [0.63, 0.80].

General discussion

In studies 1A-1E, I demonstrated that when people are tasked with matching the meaning of items used in prominent folk metaethics paradigms, items adapted from such paradigms, or novel items designed with my collaborators to exhibit greater face validity, with response options that share the meaning of those items, participants reliably fail to do so. Across studies 1A-1D, less than half of participants were able to correctly match items with their intended meaning. The only exception was a single item that, once I removed participants who failed a comprehension check, was close enough to the 50% mark that I could no longer reject the null hypothesis.

In fact, these findings may underestimate how difficult it is for participants to interpret questions about metaethics as intended. Matching one statement to another may be some evidence that people interpreted the target statement as intended. But it is not decisive evidence. In this case, about 42% of people who passed a comprehension check correctly judged the first of these statements matches the meaning of the second better than any of available alternatives::

“There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.”

“Moral truth does not depend on cultural standards or personal values”

However, understanding that two statements mean the same thing as one another does not demonstrate that one understands either of those statements in anything more than a superficial respect, much less that one understands the relevant items well enough to meaningfully express a stance on the matter. It could simply be that people have sufficient general linguistic competence to pair appropriately similar statements when prompted to do so. At least some of the consistency in people’s response to multiple scale items may be attributable to semantic overlap between the items, rather than successful measures of underlying psychological constructs. As Maul (2017) observes, responses to typical survey items “may exhibit statistical associations explainable purely by overlap in

the semantics of the items, rather than relationships between the underlying target attributes” (p. 7; Arnulf et al., 2014). Given this possibility, it could be that some (if not many) correct responses result from people recognizing semantic associations between the target statement and the correct response, without genuinely understanding the meaning of either.

It would be another thing entirely for people to demonstrate that they understand what the target statement means without any explicit prompting and without the ability to select from among a list of candidate options. These studies are, effectively, comprehension checks. Yet there is an asymmetry with such checks: widespread failure to answer them appropriately may be stronger evidence that people didn’t understand the concepts in question than passing such checks indicates that they do. This is because comprehension is best seen as a continuum, with people understanding certain concepts to a greater or lesser extent. Comprehension checks typically affix the threshold somewhere on the low end of this continuum, such that to pass a comprehension check is to demonstrate a fairly shallow degree of understanding. Around half of participants failing to meet even these minimal standards is thus, if anything, still an alarmingly poor performance, and in no way indicates that folk metaethics research is vindicated. If anything, if my best efforts can only reach this point, my meager ability to drag performance from 10-25% up to about double that is *more* conclusive than the failure of earlier studies: at least those items have the excuse of lacking face validity. If even face valid items perform this poorly, efforts to devise appropriate measures of folk metaethics may be hopeless.⁹³

⁹³ Of course, the fact that I was able to drag those numbers up at all points to the possibility that future efforts to devise stimuli may succeed where I failed. Note, however, that my other effort failed badly, performing worse than average across the previous item, with only 9.9% choosing the correct response. The ~40% rate for the realism item may be a fluke or an indication that some superficial feature of the item is prompting the higher success rate. Whatever the cause, I have certainly not demonstrated that I can *reliably* produce valid items for various forms of realism and antirealism. But perhaps I could. We cannot rule such a possibility out. Perhaps researchers could use a slightly more elaborate description of realism or antirealism that doesn’t prompt spontaneous theorizing but still manages to be sufficiently clear so as to prompt a high rate of intended interpretations. Future studies can address this possibility.

Setting this issue aside, the rest of the findings provide further evidence that people do not interpret questions about metaethics as intended. Critically, they do so without relying on open response questions or qualitative analysis, but still yield surprisingly similar proportions of “correct” responses when compared to the proportion of clear intended interpretations from the previous chapter. Given the superior performance of one of David and I’s items, which had greater face validity, the most plausible explanation for the extremely low rate of clear intended interpretations found in the previous chapter is a *combination* of both indeterminacy and invalidity: a considerable portion of participants respond in unintended ways due to researcher error. Items are so poorly designed that the ambiguities, conflation, and underspecificity of these items reliably prompts substantial numbers of participants to interpret these items in unintended ways. At the same time, people have no determinate position on realism or antirealism, so even when they manage to successfully navigate the many hazards of flawed measures, most are still left without any way to express a meaningful position.

This brings us to our last study, which illustrates that, even under such circumstances, people can still exhibit a nonrandom tendency to favor one position over another. As such, researchers cannot point to the fact that a disproportionate number of participants favor a particular metaethical position as evidence that people really do have determinate positions, since there is no good reason to think that such patterns wouldn’t persist even if folk metaethical determinacy were correct. This result challenges Pölzler and Wright (2020a). There is little evidence that many ordinary people have an explicit position on how to interpret quantum mechanics, nor evidence that the way people speak or think implicitly commits them to any particular interpretation. In short, there is no “folk quantum mechanics.” Yet I did not expect that responses to questions about quantum mechanics would be equally distributed between the Copenhagen and Many Worlds interpretations. On the contrary, I expected an unequal distribution. And that’s exactly what I found. Endorsing indeterminacy simply does *not* entail the prediction that people would equally favor competing positions, especially when

those positions are explained to the participant. This is *exactly* what Pölzler and Wright (2020a; 2020b) did. Participants were given extensive instructions and engaged in exercises to train them in the relevant concepts and distinctions. If they didn't have a metaethical position before doing so, such training would virtually guarantee that they developed one in the course of the study.

Yet the outcome of a study similar to the one I conducted on quantum mechanics could reveal quite different results if it were adapted to folk metaethics, *even if* folk metaethical indeterminacy is correct. Here's why. When it comes to quantum mechanics, it's plausible that ordinary people recognize that the topic is one very distant from everyday experience, that it requires extensive technical knowledge (and involves a bunch of scary math), and that it isn't the sort of thing they're *expected* to have any stance towards. For comparison, I don't know anything about veterinary medicine. If my wife showed me a blood sample and asked me whether the best diagnosis was Disease A or Disease B, I'd have no idea which disease was more likely. More importantly, I'd *know* that I have no idea. Just so for quantum mechanics. Is the same true for metaethics? I'm not sure. Moral considerations are far more embedded in everyday life, and metaethical concepts plausibly do arise in everyday discussions often enough that people may have the impression that they understand the relevant concepts and distinctions. If they were asked, ordinary people may report that they don't have any position on realism or antirealism. Yet they could just as plausibly report that they do, even if they don't actually understand these positions.

Granted, realism and antirealism are not nearly as technical and difficult to comprehend the Schrödinger equation, but I suspect they're technical *enough*. Indeed, while referencing the study of folk metaethics, Nichols (2014) states that "Giving a precise characterization of objectivity is itself a major philosophical endeavor" (p. 734). While Nichols goes on to characterize *one* conception of objectivism as, "roughly speaking," a type of attitude-independence, others point out that there are antirealist conceptions of objectivity as well. For instance, Hopster (2017) argues that certain forms of

constructivism can carve a middle path between realist accounts of objectivism and antirealist accounts which reject objectivism, resulting in a kind of antirealist objectivism. And Cohen (2021) draws a distinction between two other forms of objectivism: mind independence and mind nongroundedness. As I observe in **Supplement 3**, characterizing these concepts and distinguishing them from one another is a technical matter, e.g., according to Cohen:

Grounding theorists generally agree that grounding is an asymmetric, irreflexive, and transitive relation. It is further generally accepted that a ground Γ of some fact p has to be such that each member of Γ plays some role in noncausally making p the case. Facts that play no role whatsoever in noncausally making p the case cannot be part of what grounds p . (p. 183)

I don't even understand most of this, and these remarks from an article adjacent to my own area of specialization. Yet what is clear from articles like these is that Nichols is correct: it would be a considerable philosophical undertaking to develop a satisfactory account of just what it is we're trying to say when we speak of morality being "objective" or attitude-, stance- or mind-independent, with such distinctions breaking down into a variety of distinct concepts, each with a panoply of distinct qualities and implications. Sure, quantum mechanics may be far more complicated. But anyone who thinks metaethics is by any reasonable standard a simple topic characterized by straightforward, categorical distinctions that ordinary people could readily understand is simply unfamiliar with the field.

Yet for some reason philosophers have traditionally seemed convinced that a commitment to realism or antirealism is implicit in the way ordinary people speak and think, and even have beliefs about which metaethical positions are correct. Folk metaethics research isn't about discovering how ordinary people *would* think if they engaged in philosophy, it's about *how they already think*. This presumption enjoys the illusion of plausibility in part due to a lack of familiarity with just how abstract, sophisticated, and distant from everyday considerations various accounts of realism and antirealism actually are. These are highly sophisticated technical theories. It's not plausible ordinary people

endorse any of these views without first learning about them. Yet it is far more difficult to show that a commitment to realism or antirealism is absent from everyday speech. Such a possibility could fly under the radar of conventional survey methods. People need not know the syntactical rules of their language to competently comply with them; likewise, they could be oblivious to their commitment to realism or antirealism if all this amounts to is the best account of their linguistic outputs, without claiming to capture our beliefs. I see little reason to presume such commitments are built into ordinary moral thought. But even if they were, it's unclear whether conventional folk metaethics research would be appropriate for identifying such commitments.

This is because, if our goal is to uncover people's linguistic commitments, survey methods provide only indirect data about our *metalinguistic* judgments. That is, folk metaethics studies, such as Pölzler and Wright's (2020a; 2020b) training paradigms, at best only prompt people to express their theories about how they *think* they (or other people) use words, rather than directly assessing how they actually use those words in practice. Martí (2009) raised this objection with respect to early folk philosophical research on theories of reference conducted by Machery et al. (2004):

[...] it is important to distinguish carefully between observations that will reveal how people do things (in this case, use names) and observations designed to reveal how they think they do them. The latter will only provide grounds to determine how they are disposed to theorize about their practices, i.e., predict which theories about what they do they are disposed to favour. If we want to test, for instance, whether people use *modus tollens* when they reason, it may not be the best strategy to ask them, 'John knows that not B, and he knows that if A then B, should John conclude that not A?' That question prompts people to reflect on the principles they regard as correct - their answers will tell us something about their theory of reasoning; but they won't tell us how they really reason. (pp. 44-45)

If Pölzler and Wright's studies are designed to uncover people's linguistic commitments, then it's unclear whether they succeed, since this isn't what they (directly) test for, and we cannot be sure whether people's metalinguistic intuitions about metaethics comport with actual linguistic practice. If, on the other hand, they're designed to capture people's *stances*, i.e., their actual explicit metaethical

beliefs, then these studies would fail for the reasons outlined throughout this dissertation. *Even if* metalinguistic practices *were* an accurate proxy for actual linguistic practice, we'd have no way to confirm this unless *we had some way of directly measuring the metaethical commitments in our linguistic practices in the first place*, which is precisely what we *don't have*. If we did, and the purpose of folk metaethics research were to study metaethical commitments, we'd have no need to do so, because we'd already have the answer! Either way, existing research on folk metaethics provides little in the way of conclusive evidence that most ordinary people are realists or antirealists, or indeed that they have any determinate metaethical stances or commitments at all. In short, just as we have good reason to believe ordinary people have no determinate stances or commitments with respect to quantum mechanics, we likewise have little reason to suppose people have determinate metaethical stances or commitments. At the very least, those who believe people do have yet to furnish compelling evidence to the contrary.

Limitations

Although the findings in this chapter support the case against the validity of existing research and lend at least some support for metaethical indeterminacy, they are far from conclusive, and remain subject to a number of objections and limitations. With respect to Studies 1A-1E, one objection is that these studies don't provide direct evidence that participants don't interpret metaethics stimuli as intended. All they show is that participants fail to match the meaning of a given stimulus to some response option intended to reflect the same meaning as that item. There are several ways participants may have chosen "incorrect" responses that are indicative of flaws in the designs of these studies, rather than evidence that the target statements aren't a valid representation of the intended metaethical position.

First, participants were asked to select the statement or statements which are the best "interpretation" of the meaning of the statement. While they were also asked to choose the option which "most closely matches the meaning" of the target statement, *interpretation* can be understood to consist not merely of the task of judging which sentence means the same thing, but to reflect the

implications of a given statement, which are often carried by pragmatic considerations. Suppose participants were asked to interpret the sentence “human life begins at conception” in the following exchange:

Alex: “Do you think abortion is morally wrong?”

Sam: “Well, I think human life begins at conception.”

If asked to interpret Sam, it would be reasonable to infer not only that Sam thinks that “human life begins at conception,” but that Sam probably endorses a pro-life stance, and thinks abortion is morally wrong. This illustrates that “interpreting” a statement can go beyond merely identifying its semantic equivalent; it also involves making inferences. People are sensitive to pragmatic implicature; much of what people mean when they say things is carried by implication, and most people are sensitive to e.g., subtext, innuendo, and sarcasm, and recognize that certain claims or turns of phrase can hint at a person’s beliefs, values, or motivations. To illustrate, do you think Sam is more likely to be a Christian or an atheist? I suspect most readers would presume Sam is more likely to be a Christian, despite nothing about the meaning of “I think human life begins at conception,” necessarily entailing that Sam thinks that e.g., God exists, or that Jesus rose from the dead.

Participants were not directly asked about the implications of a given statement, and instructions did ask them to choose responses which *matched* the meaning of the response option, but instructions may still have misled some participants, and participants may still have felt that it was appropriate to select response options whose content was implied rather than directly conveyed by the target statement. Without knowing how participants interpreted the task, it’s hard to know how much of an impact this possibility had on responses. Yet if this problem were present, it’s unclear why the effect would be so pronounced and ubiquitous that it would so overwhelm intended interpretations and efforts to match the meaning of statements that the proportion of participants who chose the intended interpretation was often marginally better than chance.

Furthermore, response options were designed to either restate the item or to convey implications or entailments, but were mostly designed to convey what I took to be plausible interpretations of the meaning of the original statement. The response options were not pulled out of a hat, but were crafted to reflect the kinds of conflation I expected people to make, both due to armchair consideration of the kinds of ambiguities that arise in metaethics stimuli, and due to the kinds of responses people given when they were explicitly asked to explain how they interpreted metaethics stimuli. Even so, I cannot rule out the possibility that people didn't interpret the meaning-matching paradigms as intended. It remains an open, and amusingly ironic possibility that I have failed to determine whether people interpreted stimuli as intended because they didn't interpret the stimuli in my studies as intended. Since my assurances that I consider this unlikely don't count for much, I'll leave it to future research to refute or corroborate my findings.

Another limitation with these studies is that they were all conducted on MTurk, among participants who had little incentive to take the task seriously. Pölzler (2021) has shown that insufficient effort responding (IER) can pose a significant threat to the validity of studies in experimental philosophy, and it is worth noting that Pölzler is one of the only people whose research focuses primarily on folk metaethics.⁹⁴ Since these studies were incredibly short, this buffers them against IER, but the difficulty of the task may have exacerbated IER. Since around 13-20% of participants failed the comprehension checks in these studies, it's plausible that IER threatened the outcome of these studies, though all results remained significant when excluding participants who failed comprehension checks. However, little research has directly addressed the role comprehension

⁹⁴ Pölzler estimates IER to account for 10% of responses in self-report surveys in general, yet as low as 5% IER is sufficient to threaten study results. It's less clear what the overall rate of IER is in experimental philosophy, but Pölzler points to a variety of measures that are roughly consistent with around 10% IER, e.g., 10-18% of participants failed various comprehension checks and attention checks across a variety of studies. Around 13-20% of participants failed the comprehension checks in studies 1A-1D, which is roughly comparable.

checks play in assessing IER, so it's unclear how helpful these results are in estimating the presence of IER.

One potential concern with Study 2 is the possibility that many people really do have determinate stances towards the Copenhagen and Many Worlds interpretations. It's possible that quantum mechanics has become a sufficiently popular topic that it has permeated popular culture. Indeed, the study was launched in June of 2022, just one month after the release of *Doctor Strange in the Multiverse of Madness*. Shows like *Star Trek* and *Rick and Morty* have likewise familiarized people with the notion of multiple universes. In fact, the first episode of Season 2 of *Rick and Morty*, "A Rickle in Time," explicitly involves a state of quantum uncertainty resolving by branching into multiple universes. While it's possible enough ordinary people are familiar with different interpretations to have a determinate stance, this isn't a plausible explanation for the results. First, after excluding participants who reported understanding quantum mechanics or having formal training in physics, 22.4% ($n = 61$) of participants reported that they were familiar with the Many Worlds interpretation, while only 8.5% ($n = 23$) reported familiarity with the Copenhagen interpretation. These are hardly reassuring numbers. Furthermore, mere familiarity with the *terms* is hardly sufficient to have a view on the matter. Most importantly, almost all participants explicitly stated that they had no position. Yet even if people did report having a position, we should interpret such claims with caution. Pop cultural references to quantum mechanics are hardly adequate to furnish people with a sufficiently substantive or accurate understanding. In fact, the prevalence of cultural references to quantum mechanics overlaps in some ways with references to metaethics that occasionally trickle outside of academia, as when politicians or religious leaders speak of the disastrous consequences of "moral relativism." In both cases, popular understanding of these concepts is likely minimal and riddled with misconceptions. Such misconceptions and associations may even play a significant role in explaining *why* responses to

questions for both topics are nonrandom: people *think* they have some sense of what various technical terms mean, even if they don't.

Finally, while I do not develop on these points in much detail, there remains the question of whether or not, and to what extent, these findings support metaethical indeterminacy. I will address this in greater detail in the conclusion, and will focus here on how these findings support indeterminacy. Studies 1A-1E illustrate that, even when participants are presented with a variety of metaethics stimuli, they reliably fail to match the meaning of these items with the intended metaethical meaning.

Future directions

IER may have been a significant factor in the poor performance for Studies 1A-1E. Future studies should address the same or similar questions using methods that increase effortful responding, e.g., incentives for getting the correct response. Such methods could also be expanded to include additional metaethics paradigms that have yet to be tested. It's also possible that people perform poorly with these tasks in general, which would suggest flaws with the method rather than flaws with the stimuli. There are two ways to assess this. First, we could provide the same stimuli to professional philosophers, and see how well they perform. If they perform much better, this would indicate that training does allow people to identify the sentences which match the meaning of the target statement. We could also employ similar matching exercises for topics we expect ordinary people to be competent with, as a control. Once again, if the response rates are higher in these cases, this would suggest that there are distinctive difficulties with interpreting metaethics items as intended.

This method could also be adapted for assessing interpretation in other nonmoral domains, e.g., judgments about taste preferences, aesthetics, epistemic norms, social conventions, and so on. It could even be adapted to other philosophical questions, or to psychological stimuli more broadly. Note, however, that what I've done essentially amounts to glorified comprehension checks. I would

not suggest so much that researchers begin employing the meaning-matching paradigm as a standalone method. Rather, I would suggest instead that researchers follow at least three general guidelines:

(1) *Be more careful about designing stimuli.* This is so general as to be nearly worthless, but the general point stands. Researchers are often inattentive to potential interpretative difficulties with their stimuli. As *G.I. Joe* has taught us, knowing is half the battle. We can't begin to address interpretative difficulties if we're not aware of the problem, and make an active effort to think about the role pragmatics and ambiguity can prompt unintended interpretations. Researchers should also be more wary of the curse of knowledge (Birch et al., 2017) which is difficult to suppress, and the methods employed are often inadequate to mitigate its influence. For instance, Collier-Spruel et al. (2017) had a panel of *experts* evaluate items intended to represent relativism. While these experts may have been good at judging how well these items represented relativism, they did not judge how *nonexperts* would interpret these items, nor would they be an appropriate pool of judges to do so. The only way to know *that* would be to conduct the proper empirical tests. This brings me to the second piece of advice

(2) *Pretest items to assess interpretability.* Researchers should employ methods like the ones reported here *prior* to conducting more extensive research. It's absurd to rely on a set of measurement tools without finding out if those tools actually work. Relying on previously validated measures and scales will not be adequate, since a scale validated in one population may not be valid in another, and purported evidence of validity is often inadequate to ensure that the measures in question actually exhibit good validity (Hussey & Hughes, 2020; cf. Wetzel & Roberts, 2020). Furthermore, researchers often modify or adapt items to a new population without retesting validity (Flake, Pek, & Hehman, 2017).

At the risk of being impertinent, I'll just come out and say what I actually think: researchers have historically been *far too careless* about employing decent measures. They've largely coasted on identifying some purportedly validated measure (or not, as they often just make up *ad hoc* measures on the fly), plopping it into a study, citing some paper that allegedly validated the measure that they

probably didn't read (and probably didn't provide good evidence of validity anyway, since traditional validation procedures are inadequate; see Maul, 2017), and declaring victory. Mindlessly copying-and-pasting measures or conjuring a measure on the fly is barely scientific. There are probably astrologers that put more work into making predictions. In any case, when the topic in question is subtle or complicated, as is often the case with folk philosophy, researchers have an especially strong incentive to pretest items to assess how well people interpret them, and to modify them accordingly. There are already methods for doing this (Hunt, Sparkman, & Wilcox, 1982; Reynolds, Diamantopoulos, & Schlegelmilch, 1993). However, these methods may be insufficient for research on folk philosophy. We may need to develop more rigorous and distinctive methods for pretesting questions related to philosophy, and be open to the possibility that in some cases there aren't psychological constructs that function as analogs to a philosophical account. Finally, pretesting may fall short of ensuring that everyone interprets philosophical questions as intended. This brings me to the final suggestion.

(3) *Include better comprehension checks.* In addition to pretesting, researchers should include comprehension checks in their studies. Yet it's not enough to employ simple multiple choice questions. The comprehension checks that appear in studies often demand little more than minimal attention to the instructions. For instance, the comprehension checks I employed in the quantum mechanics studies simply require recalling the instructions presented earlier in the survey. People can recall such information even if they have an incredibly superficial understanding of the stimuli. In other stances, researchers may collect open response data, only to pass analysis off to people who lack the training, expertise, or appropriate level of skepticism to analyze responses (Bush & Moss, 2020).

Future studies should also be directed at developing on the concept of spontaneous theorizing and devising novel methods for assessing if, when, and how it occurs. I'm not sure what form these studies would take, but providing more robust evidence that spontaneous theorizing can occur, and that it can explain the results of studies in folk philosophy or more generally could go some way

towards recognizing a novel methodological problem that could impede progress in psychology. At a minimum, researchers could begin by more carefully assessing whether participants themselves report having held no position on a topic prior to the study, but developing one over the course of it. Indications of this occurring could perhaps be achieved by e.g., cognitive interviewing (Beatty & Willis, 2007; Drennan, 2003; Madans et al., 2011; Willis, 2004), which incidentally may also be valuable for designing folk philosophical studies in the first place.

5.3 Conclusion

I had three goals in this chapter: to provide further evidence for the invalidity of existing measures of folk metaethics, to provide an illustration of spontaneous theorizing, and to provide evidence that nonrandom response patterns can emerge from surveys even when we have strong theoretical grounds for presuming indeterminacy. All three goals were achieved. These findings bolster my argument that existing folk metaethics research relies on invalid measures. In addition, I have shown that studies can yield nonrandom response patterns even if people held no prior stance on an issue. This illustrates the possibility that studies can yield replicable patterns of results that give the appearance of capturing the intended psychological phenomena, even if they do not. I have provided two explanations for this outcome. First, some participants may engage in spontaneous theorizing, developing a position in the course of a study that does not reflect a position they held prior to the study. Second, people may systematically favor one response over another for reasons unrelated to it, accurately representing their position on the matter as a result of the forced choice design typical of most research in folk metaethics. Neither of these explanations are mutually exclusive, but may work in tandem to produce artifactual findings.

While these findings do not directly demonstrate metaethical indeterminacy, they show how existing studies could create the *illusion* of a determinate folk metaethics even if none existed. Given the absence of strong theoretical foundations for supposing that there is a determinate folk metaethics,

the onus should be on those who maintain that ordinary people are realists or antirealists to provide adequate evidence for this claim. Such evidence would, at a minimum, require valid measures that don't prompt spontaneous theorizing. At present, no studies have met these conditions.

CHAPTER 6: Third Person Paradigm & Normative Entanglement

6.0 Introduction

One of the most common ways ordinary people interpret questions about metaethics in unintended ways is by interpreting them as questions about their *first-order*, or *normative* moral standards (these terms may be used interchangeably). In other words, when asked a question intended to determine whether the participant endorses moral realism with respect to a particular moral issue, such as lying, participants instead interpret the question about whether they think lying is morally right or wrong. Pölzler (2018b) drew attention to this methodological concern in a critique of research on folk metaethics that focused primarily on the disagreement paradigm. According to Pölzler:

Another problem with many studies on folk moral realism is that they have not sufficiently accounted for moral realism and anti-realism's moral neutrality. In particular, researchers have attempted to measure subjects' intuitions about these views by using scenarios, questions and answer choices that may also have prompted strong first-order moral intuitions. Subjects' responses in these studies may consequently partly be explained by these intuitions, rather than by their views about moral realism and anti-realism alone. (p. 657)

Early versions of the disagreement paradigm are especially susceptible to this concern. For instance, after asking participants about their own first-order moral stance towards a moral issue, then telling them that another person disagreed with them, Goodwin and Darley (2008) asked participants to select one of the following response options:

- (1) *The other person is surely mistaken.*
- (2) *It is possible that neither you nor the other person is mistaken.*
- (3) *It could be that you are mistaken, and the other person is correct.*
- (4) *Other.*

Participants may have interpreted this as a question about their moral stance towards the issue in question, *not* their metaethical stance. Suppose, for instance, you are a moral relativist, and your moral

claims reflect your subjective moral standards. When asked whether someone who holds contrary moral standards is *mistaken*, you could say *yes* as a way to express your normative moral stance that you reject contrary moral standards, *not* to convey the view that people with contrary moral views fail to endorse the stance-independently correct set of moral facts. I refer to these unintended interpretations *normative conflations*. Normative conflations occur whenever someone interprets a question or statement about metanormative (second-order) concepts to instead be a question about or statement about normative (first-order) concepts. In this case, there is no need to speculate. In **Supplement 4**, thematic analysis of Goodwin and Darley's data revealed that normative responses (rather than metaethical responses) were the most common theme for unintended interpretations in both of the studies for which their data was available (19% of responses in Study 1 and 17.8% in Study 2). In other words, I have already provided empirical evidence that normative conflations are (in at least some cases) very common.

One indirect line of evidence supporting the ubiquity of normative conflations is the consistent correlation between measures of the strength of people's first-order moral positions and their metaethical positions towards the same moral issue (Pölzler, 2018b). In other words, the more strongly people reported that a given issue was morally right or wrong, the more likely they were to also endorse the realist response option towards that same issue (e.g., Beebe & Sackris, 2016; Beebe et al., 2015, Goodwin & Darley, 2008). Overall, a majority of studies have found a correlation between attitude strength and realism (though Wright et al., 2013 found no significant correlation between attitude strength and realism). Such findings provide some circumstantial support for the claim that normative conflations are common. It's possible that greater attitude strength predicts greater propensity for normative/metaethical conflation. For instance, one possibility is that attitude strength is associated with a greater affective response, which could result in an emotional bias that influences how participants respond to what are intended to be questions about their metaethical positions.

Consider, for instance, two questions about a person's metaethical stance: whether they're a realist about the moral status of (a) speeding on the highway and (b) kidnapping children to harvest their organs. A person may have a more sober reaction to (a) than to (b), regarding the latter as far more evocative and disturbing. As a result, when asked about their metaethical stance towards these issues, if they're presented with an opportunity to express that someone who disagrees with them, and thinks speeding on the highway and kidnapping children to harvest their organs are "not wrong," they may be more inclined to judge that this person is mistaken in the latter case, not because they're a realist about the issue, but because they're more motivated to express their *normative* opposition to (b).

While this explanation would characterize normative conflations as a form of bias, normative conflations could also be a byproduct of a rational tendency to prioritize managing one's reputation rather than responding to questionnaires as researchers intend. If we think about our own experiences, we can probably recall times where people made an explicit point, or go out of their way, to express their condemnation for a particular action, or to disavow some person or deed. Such instances are often ones in which a person's reputation is on the line, and they deem it important to distance themselves from the act in question. People may internalize a disposition to manage their reputation in this way and, as a result, be more inclined to opt to express a normative stance towards highly egregious moral transgressions most people would oppose, and less disposed to do so when the moral issue in question is controversial or not severe.⁹⁵

Far from indicating a bias or error, such responses would instead indicate an emphasis on ensuring one does not say things that could harm their reputation over a literal but socially flat-footed style of response where we ignore the potential reputational harms that could come to us by answering questions in particular ways. This sociofunctional explanation may account for normative conflations

⁹⁵ Drawing on the same example, expressing opposition to kidnapping and harvesting organs may be more important for managing your reputation than condemning speeding, since giving the impression that you don't find the former repugnant would be far more counternormative and reputationally costly than a failure to express opposition to speeding.

not by attributing them to a performance error, but to a practical approach to responding to questions that prioritizes maintaining a positive reputation over responding in the most literal, direct, or philosophically sophisticated way possible, regardless of the consequences.⁹⁶

One line of evidence consistent with this hypothesis is a reliable and strong correlation between realism and perceived consensus (Goodwin & Darley, 2012). In fact, when Goodwin and Darley experimentally manipulated the alleged consensus on a moral issue, this caused an increase in realist responses. One explanation for these findings is that people use consensus as *evidence* of realism, i.e., if most people agree about a particular moral issue, people may take this to be evidence that there is a stance-independent fact about that issue, or at least that there is a single correct answer (Ayars & Nichols, 2020). However, it is also possible that perceived consensus tracks perceived social stakes: the more there is a general consensus about the moral status of the action, the greater the cost of expressing antirealism towards that stance, since doing so could fail to convey a strong and intolerant stance towards people with contrary moral views.

To illustrate, consider how people might react to a person who expresses relativism towards a low consensus moral issue, such as euthanasia. A person who expresses relativism about this issue may be perceived as having a tolerant and permissive attitude towards those with contrary views. Whatever we think of such a permissive attitude towards euthanasia, consider instead a person who expresses relativism towards genocide. If this is perceived as indicating a tolerant or permissive attitude towards genocide, others may recoil at such an attitude, and respond with disgust and outrage. At least, they may do so to a comparatively greater extent than they would towards someone who holds a contrary view towards euthanasia. This illustrates that character judgments about a person expressing

⁹⁶ Such responses need not even be motivated by any awareness of such a prioritization. One's interpretation of such stimuli could in principle occur in a way that is introspectively inaccessible and precedes conscious reflection over one's interpretation of the question.

a particular metaethical position towards a given moral issue may vary in accordance with perceived consensus towards that issue (and, for that matter, perceived seriousness).

Regardless of *why* normative conflations may occur, a correlation between attitude strength and metaethical stance only alludes to a potential conflation, but does not directly establish it. My findings, on the other hand, *do* strongly indicate that normative conflations are common, at least for certain metaethics paradigms. While **Chapter 4** focuses on intended and unintended interpretation rates, the accompanying **Supplement 4** employs thematic analysis to categorize responses in accordance with a variety of ways in which participants interpreted questions about metaethics. The *normative* theme captures those instances in which participants appeared to express a normative moral stance rather than a metaethical position. The *normative* theme was one of the most prominent and recurring themes across a majority of studies, appearing in the top five most common themes in at least one condition in studies 1, 2, 4, 5, and 6 (and only failing to appear in the top five in studies 3 and 7). I'll provide a handful of examples to illustrate instances in which participants offer responses that appear to express a normative rather than metaethical rationale for their answers. In study 4B, participants were asked:

In your own words, what does it mean to say that the truth of the moral claim "murder is morally wrong" is objective?

Here are a handful of participant responses classified as *normative*:

To say that denying someone else's equally valid claim to life as that of the murderer would be morally wrong.

that murder is a bad thing

Murder IS morally wrong, it's taking the life of someone who has more life to live. It's taking their right to live their full life.

When directly asked what it means for morality to be “objective,” these participants expressed what appear to be normative claims about the moral status of murder, *not* claims about the stance-independent of murder. A similar response pattern emerged in most other studies that appeared in

Chapter 4, with the *normative* theme often being the first or second most common theme, including (as previously noted) Study 1A (19% of responses) and 1B (17.8% of responses), as well as 4B (21.1% in the realism condition, 25.6% in the relativism condition), 5A (11.5%), and 5B (15.4%). People consistently expressed a normative moral stance in response to a variety of questions, including what they thought the source of disagreement was between them and a previous participant (Studies 1A and 1B), asking the participant what someone who expressed a realist or antirealist stance meant (Studies 2A and 2B), directly asking participants what they think it means to say that “abortion is morally wrong” or “murder is morally wrong” is objective or relative (Study 4B), and asking participants to explain their level of agreement with items that appear on metaethics scales (Studies 5A, 5B, 6A, and 6B).

While the normative theme did not always emerge as one of the most common themes, it frequently appeared in a variety of paradigms. This illustrates both the ubiquity and generality of the tendency for people to appeal to normative moral considerations in studies intended to assess their metaethical stances. Such findings provide considerable support for the conclusion that normative conflations are quite common. Indeed, in subsequent work, Pölzler (2018a) and Pölzler and Wright (2020b) repeated their concerns with the potential for studies to prompt conflations between normative and metaethics, concerns which culminated in efforts to partially circumvent the problem through the use of training paradigms (i.e., paradigms that provide extensive instructions or exercises intended to familiarize participants with the relevant metaethical terms and concepts). In spite of optimism about the prospects for more rigorous methods of evaluating folk metaethical stances and commitments, Pölzler (2018b) concedes that “Avoiding first-order moral intuitions in studies on folk moral realism altogether may be methodologically infeasible” (p. 658).

I am sympathetic to Pölzler’s concern. It may not be possible to completely separate metaethical and normative considerations when evaluating how ordinary people think about

metaethical questions.⁹⁷ However, we can address normative conflations and the potential associations between folk conceptions of metaethics and normative ethics in at least two ways:

(1) to the extent that such conflations occur, we can attempt to understand *why*, and (2) we can attempt to mitigate the influence of normative conflations by devising alternative paradigms that are less likely to prompt normative conflations.

One notable feature of all previous metaethics paradigms, is that they attempt to directly evaluate the *participant's* metaethical stance, and often do so while evaluating their normative stances as well, often simultaneously or in near proximity to one another. For instance, participants in a typical study employing the disagreement paradigm will be asked to express their level of agreement or disagreement with a moral claim. Then, they will be asked whether someone who disagrees with them is mistaken or is also correct. The use of concrete moral issues may exacerbate the degree to which people are disposed towards normative conflations. In addition, by putting people in situations in which they are asked to consider *normative* moral disagreements with others, people may retain a desire to express their opposition to people with conflicting normative moral standards. Consider an ordinary person asked to consider a dispute about the moral status of abortion. When confronted with a person who holds a contrary position, the importance of signaling one's opposition to a contrary position on abortion may be more important than expressing a position on the metaethical status of moral claims about abortion. In other words, such scenarios are especially prone to amplifying the social incentives participants have to signal their normative moral stances. For ordinary people, everyday moral judgment isn't merely a matter of expressing the correct or most logically consistent

⁹⁷ It is also unclear whether normative conflations are an artifact of limitations in experimental design, or reflect substantive features of the way ordinary people think. In other words, it could be that ordinary people can and do distinguish metaethical and normative considerations, but the way studies are conducted fail to properly disentangle and independently elicit people's metaethical and normative judgments. Alternatively, it could be that ordinary moral psychology is structured in such a way that the two are inextricably linked (for at least some people, in some populations). If so, normative conflations may be a feature of ordinary moral psychology, *not* a methodological artifact. While philosophers may draw a conceptual distinction between metaethics and normative ethics, this is no guarantee that ordinary people distinguish them from one another, at least not without training.

judgment, it is a matter of maintaining one's status both within their society at large and within one's ingroup. In the former case, this is achieved by expressing condemnation of antisocial or universally condemned behaviors, and in the latter it is achieved by demonstrating convincing display of loyalty and commitment to the group's, and a rejection of rival moral standards.

Even when participants are asked in abstract terms about their metaethical standards, and they are not also asked about their normative moral standards, as in e.g., Collier Spruel et al.'s (2019) moral relativism scale, which does not include questions about specific moral issues, normative connotations may *still* be present. This is because metaethical claims, even when they are expressed in abstract terms, can prompt normative concerns, insofar as a particular metaethical stance may be taken to signal or imply a normative moral stance, or to more generally imply aspects of a person's character that are not strictly entailed by a metaethical stance. In particular, the *expression*, if not a genuine commitment to or endorsement of, a particular metaethical position may serve one or more sociofunctional purposes, e.g., expressing relativism may signal one's tolerance for others, while expressing realism may signal one's commitment to their values. Conversely, expressing a relativist could signal a lackadaisical attitude towards morality that could make one seem noncommittal and wishy-washy, while expressing a realist stance could lead others to infer that you are dogmatic, rigid, or inflexible.

Previous research is consistent with these suggestions. Collier-Spruel et al. (2019) found a strong significant correlation between responses to their metaethical relativism and tolerance scales ($r = .56$). This indicates that the stronger people endorse relativism, the more they express tolerance for people with contrary moral views. A slate of other studies likewise reveal that greater endorsement of relativism is associated with greater tolerance for people with opposing moral standards (Wright, Cullum, & Schwab, 2008; Wright, McWhite, & Grandjean, 2014; Wright & Pölzler, 2022) and openness to experience (Feltz & Cokely, 2008). Moreover, Goodwin and Darley (2012) found that realist responses were associated with reduced willingness to reconsider one's moral beliefs, greater

discomfort with people with different moral values, and a tendency to regard people with conflicting moral standards as more immoral. Taken together, these findings suggest that, at the very least, relativist responses are associated with tolerance, while realist responses are associated with discomfort, negative moral evaluation, and close-mindedness.⁹⁸

These findings suggest that there may be a genuine tendency for relativists to be more tolerant than realists. Of course, they don't directly establish that people perceive others who express relativism or realism towards a moral issue to be more tolerant or intolerant. Yet the fact that such an association exists points to the possibility that people anticipate how others would perceive their response to questions about metaethics, and that this in turn influences what metaethical position they express or adopt. In other words, it could be that, to the extent that people anticipate, consciously or otherwise, how others might perceive the expression of a particular metaethical stance, they may strategically eschew or employ such language in accordance with their social goals. If so, "realist" and "relativist" responses could be caused by one's normative moral attitudes, including their tolerance for opposing moral views, but it could also be the result of the strategic adoption of locally appropriate language for the purposes of managing one's reputation or persuading others (Mercier & Sperber, 2011). In sum, there may be a variety of factors prompting normative conflation: ordinary people may interpret questions about metaethics as questions about their normative standards, they may wish to signal a genuine commitment to particular normative moral standards, or they may be motivated to express

⁹⁸ Of course, previous chapters have raised substantial doubts about the validity of the measures used to identify people as realists or relativists. Such concerns are less of an issue here because my primary concern is with whether people *perceive* an association between metaethical views and moral attitudes. It could be that people who are more disposed towards tolerance are more likely to favor relativist-sounding responses, while those less disposed towards tolerance favor realist-sounding responses, not because relativists are more tolerant than realists, but because more tolerant people favor more relativist-sounding responses, *even if* they don't interpret them as intended. That is one advantage of the present set of studies: at the very least, they can establish whether people's *responses* to questions about metaethics are associated with their normative moral standards, even if we remain skeptical about the validity of those measures. Indeed, to the extent that responses to questions ostensibly about metaethics are associated with a variety of normative moral attitudes, character traits, and so on, this could be leveraged to corroborate doubts about the validity of metaethics measures. But even if such concerns are set aside these results would still be informative.

normative moral standards rather than their metaethical views to achieve various practical social goals, such as managing their reputation or persuading others.

Whatever factors drive normative conflations, it may be possible to minimize the risk of these conflations by asking participants about what they believe *other* people's metaethical standards are. This approach could mitigate normative conflations by providing participants with the opportunity to assess what metaethical standards they believe typify the way ordinary people speak or think without entangling these judgments with their personal moral standards and commitments. This could reduce both the risk of unintended interpretations and the risk that participants would believe their responses to questions about metaethics could signal additional information about their normative moral standards, character, or other traits.

The first goal of this chapter is to develop just such a paradigm. I implement a *third person paradigm* that the participant's judgments about the metaethical standards of a typical person in their society rather than their own metaethical views. The third person paradigm also exhibits a number of other advantages. Part of the goal of traditional philosophical approaches to descriptive metaethics is to assess the public meaning of moral claims. Philosophers are not generally interested in each individual's proprietary ways of thinking and speaking, but in the best externally adequate account of ordinary moral discourse (see e.g., Sinnott-Armstrong, 2009). A third person paradigm may provide more direct evidence of such an account than other paradigms (e.g., the disagreement paradigm), thereby yielding results with greater relevance to philosophers. Such an outcome could be of more immediate interest to psychologists as well, since it's possible this approach could minimize the risk of spontaneous theorizing (i.e., when participating in a study causes a participant to express or develop a view they did not hold prior to participation). Participants encouraged to reflect on their own views may be prompted to reflect in ways that result in novel or spontaneous judgments that don't genuinely reflect how they or others tend to think outside the experimental context in which the study is

conducted. In contrast, judgments about how a typical person thinks or speaks about metaethics may be a better reflection of typical usage. If so, third person paradigms may have greater external validity than standard first-person paradigms.

It's also possible that the third person paradigm could find a similar pattern of endorsement for realism and antirealism as existing measures. If so, this could reveal that many of the concerns expressed here are unfounded, which could partially vindicate existing paradigms. If, on the other hand, the overall pattern differs, this would suggest either a shortcoming with the third person paradigm or further substantiate concerns with the validity of existing measures. I test for this by presenting participants with first-person versions of the same questions, which allows for a direct comparison of first-person and third-person responses. Studies 1 and 2 address these questions in the moral domain by assessing what metaethical inferences people make about a typical person in their society, either in the absence of any information about that person at all, or given that the person has made a moral assertion that an unspecified action is morally right or wrong (Study 1), or that a particular concrete moral issue is morally wrong (Study 2).

Finally, previous studies employing the disagreement paradigm have also used questions in nonmoral domains, including taste/aesthetics, social conventional, and factual claims about e.g., science and history (Goodwin & Darley, 2008; 2012; Wright, Grandjean, & McWhite). I employ a third person paradigm for these same nonmoral domains in Study 3. Since normative confluations are less plausible in these domains, such findings may also reveal a similar pattern of results in these domains when compared to first person responses but a different pattern of results when comparing third and first person moral judgments, a disparity that could raise questions about the degree to which certain methodological problems may be distinctive to moral questions, further bolstering the case for the role normative confluations may play in threatening the validity of first-person metaethics paradigms.

If nothing else, this approach provides a novel way to assess how ordinary people think about metaethics that could yield insights inaccessible to existing paradigms. This is especially important, since attempts to assess realism and antirealism in nonmoral domains have received comparatively little attention. The attention that been directed at them has also typically been perfunctory and superficial, since these domains are treated as control groups or as ancillary to questions about metaethics, and often rely on crude adaptations of measures intended for metaethics, rather than measures carefully designed specifically to assess realism and antirealism in that particular domain. Of all the nonmoral domains, folk realism and antirealism about aesthetics has received the most attention (e.g., Bonard, Cova, & Humbert-Droz Cova & Pain, 2012; Cova, Garcia, & Liao, 2015; Cova et al., 2019; Rabb et al., 2020). Unfortunately, as Moss and I argue in Moss and Bush (2020), the measures that appear in Rabb et al. (2020) suffer especially poor validity, in that their instructions fail to adequately distinguish realism and antirealism. Other problems are also common, including misleading or biasing instructions and the outright misoperationalization of the relevant constructs, both of which threaten the validity of at least some studies on folk metaaesthetics in ways that are even more serious than the threats to folk metaethics. This further highlights the need for novel paradigms that do not suffer from as many methodological deficiencies.

Given the poor validity of existing measures, and the need to devise new measures in the absence of substantive and reliable evidence about folk metaethics, my goals are largely exploratory. Nevertheless, consistent with the most rigorous and recent studies (in particular, Pölzler & Wright, 2020a; 2020b), I expect people to judge a typical person to lean towards relativism, and to be less disposed to endorse realism and universalism.⁹⁹ I have no meaningful expectations about

⁹⁹ This expectation is at least partially based on more recent and more comprehensive studies on the psychological of metaethics, which tend to find that a majority of participants favor antirealist positions (Davis, 2021; Pölzler, Zijlstra, & Dijkstra, 2022; Pölzler & Wright, 2020a; 2020b). In addition, my studies draw on a participant pool that is disproportionately likely to be nonreligious (Levy, Freese, & Druckman, 2016), a trait negatively associated with endorsing

noncognitivism, though it's worth noting that when Beebe (2015) introduced a noncognitivist response option to the disagreement paradigm it was frequently selected, while Davis (2021) found noncognitivism to be the modal response using a modified version of the disagreement paradigm (cf. Pölzler & Wright, 2020a).

I am skeptical of such results, however, since judging that moral claims can be neither right nor wrong could serve as an epistemic hedge that avoids committing one to a *normative* stance on a given issue; as such, I suspect people may not be interpreting this response option as intended, or may not be selecting it as a genuine reflection of their metaethical stance. In addition, I introduce distinct measures of both realism and universalism. Some researchers have noted that realism and universalism are distinct positions, and have opted to measure the latter instead of the former (Ayars & Nichols, 2020; Rose & Nichols, 2019). My goal in assessing both is largely exploratory. Findings may indicate that people make little or no distinction between the two, or people's responses to realism and universalism could turn out to vary.. I also introduce measures of character to assess how participants perceive the character of people who make particular normative moral claims.

Although studies 1-3 assess whether participants think a typical person in their society is a realist or antirealist given a normative claim (or at least a first-order claim, in the case of the factual domain), Study 4 departs from the third person paradigm in a critical way: participants are asked to judge the character of a person who expresses a metaethical stance, effectively reversing the approach taken in studies 1-3. In other words, the third person paradigm asks people to infer a person's metaethical stance or commitment (either in the absence of any information, or given the target of their judgment's normative moral stance), while Study 4 examines the participant's normative moral evaluation of a person given that person's metaethical stance.

moral realism (Goodwin & Darley, 2008) and positively associated with endorsing moral relativism (Collier Spruel et al., 2019).

Study 4 has a different aim. It tests a specific hypothesis about why some people may be disinclined to express antirealism towards some moral issues: namely, that doing so may pragmatically imply that they are less opposed to serious moral transgression and lack an appropriate emotional attitude towards them (i.e., disgust or outrage). If participants perceive a person who expresses an antirealist stance to have similar or equal moral character to a person who expresses a realist response only when they *cancel* whatever implicature or pragmatic inference may be associated with expressing such a stance, but not when they don't, this would support the possibility that participants perceive expressing antirealism as potentially signaling undesirable character traits, a factor that could motivate participants to avoid expressing antirealism due to its anticipated reputational consequences.

6.1 Study 1: Third person paradigm (pilot)

Study 1 introduces the third person paradigm. In this initial version of the paradigm, I devised three conditions: a condition in which participants were asked to judge the metaethical beliefs of a typical person in their society (the *no statement* condition), and two conditions in which they were asked to judge the metaethical beliefs of a person who either asserted that an unspecified action was morally right (*abstract right* condition) or morally wrong (*abstract wrong* condition).

Although my primary goals were largely descriptive and exploratory, I predicted that judgments about whether another person was a realist or antirealist would vary significantly across conditions and, in particular, that people would be least likely to judge another person to be a realist in the no statement condition, most likely to judge others to be realists in the abstract wrong condition, and that the abstract right condition would be intermediate between the two. This is because a person who has made no statements at all provides little or no information about how they think about morality, so participants must infer what such a person is like based on their prior knowledge about what people tend to be like. However, the moment a person makes a normative moral assertion that a given action is right or wrong, this provides *some* information that they take a stance on moral issues,

which may increase or render salient the notion that they also think there are objective or universal facts about moral issues. Counterfactually, one might reason that a person who wouldn't state that an action is morally right or wrong would be less likely to think an action is *objectively* or *universally* right or wrong, given that they've provided no indication that they think anything is morally right or wrong at all (objectively, universally, or otherwise). Someone who asserts that an action is morally wrong may also be seen as especially judgmental, since someone who asserts that a given action is right could, in comparison, appear more morally permissive.

Aside from these predictions, another goal was to assess how various metaethics measures introduced in this study performed in the context of the paradigm, since these same measures were initially devised for a conventional self-report scale. I also wanted to see whether responses to the third person and first person versions of these measures would differ, though I didn't have any specific predictions. For instance, I wanted to assess whether mean scores on the various subscales (realism, universalism, relativism, and noncognitivism) roughly approximate how people responded when asked to report their own metaethical standards, or whether there would instead be meaningful differences between third person and first person measures. Finally, I introduced both realism and universalism as separate measures, in order to assess whether or not, and to what extent, people might distinguish the two.

Methods

Participants. I aimed to recruit 400 participants. 402 participants began the study.¹⁰⁰ However, 12 participants did not complete the study and were excluded from further analysis. This left a total of 390 participants. Participants consisted of 390 adult US residents on Amazon's Mechanical Turk (185 females, 203 males, 1 other, 1 unreported, $M_{\text{age}} = 39.8$, $SD_{\text{age}} = 11.7$, age range = 21-73).

¹⁰⁰ The collection of participants beyond the intended sample size is a quirk of the recruitment process on Amazon's Mechanical Turk, which were recruited through CloudResearch.

Procedure. All participants were given the same initial set of instructions, which asked them to imagine a typical person in their society. Then they were randomly assigned to one of three conditions: (1) a *no statement* condition, (2) an *abstract “wrong”* condition, or (3) an *abstract “right”* condition. In the no statement condition, participants were given the following additional instructions:

Think about a typical person in your society. Please rate how likely or unlikely it is that they hold the following beliefs.

In the *abstract “wrong”* and *abstract “right”* conditions participants were instead asked to consider a scenario in which a person was having a discussion about moral issues and, during the course of this discussion stated, “That is morally wrong,” or “That is morally right,” respectively. After reading this scenario, participants were asked to:

Think about the person in this scenario. Please rate how likely or unlikely it is that they hold the following beliefs.

After this, participants were asked a set of twelve questions about the metaethical standards they believe this person would have. After completing these measures, participants were then told that they’d be asked what they thought about the statements they were just asked about. They were then asked to express their level of agreement with each of the same twelve items and to answer demographic questions.

Measures. All participants judged the typical person in their society’s metaethical views by expressing how likely it was that the person endorsed realism, universalism, relativism, and noncognitivism 1 (Very unlikely) to 7 (Very likely).

Realism reflects the view that there are stance-independent moral facts about what is right or wrong. This means that moral facts are not made true by people’s attitudes or stances. For example, one realism item states, “They believe moral truth is independent of cultural standards and personal beliefs.”

Universalism is the view that there is a single correct moral standard (Ayars & Nichols, 2018).¹⁰¹ Universalism is conceptually distinct from realism. At the risk of oversimplification, universalism concerns *how many* correct sets of moral standards there are (just one), while realism concerns whether moral truths are stance-independent, i.e. *what makes them true* (not people's stances).¹⁰² Items representing universalism were intended to capture this distinction, e.g.: "They believe there is a single set of moral standards for all people and cultures."

Relativism holds that moral claims are only true or false relative to the standards of different individuals or groups. One such item stated that "They believe that things are only morally right or wrong according to different points of view."

Finally, cognitivism is the view that moral claims *propositions* (sentences that can be true or false), while noncognitivism holds that they are neither true nor false because they aren't propositions. Instead, noncognitivists maintain that moral claims express nonpropositional content, such as emotions or commands (e.g., "don't murder!"). This distinction was represented by three items, one reflecting noncognitivism and two that were reverse coded to reflect cognitivism. One such item is, "They believe that judgments about whether an action was morally right or wrong can be correct or incorrect."

All participants were then asked to express how much they agreed or disagreed with each of these items 1 (Strongly disagree) to 7 (Strongly agree). All items were reworded to remove the third-person attribution at the start of the sentence (e.g., "They believed that..."), resulting in otherwise

¹⁰¹ The term "universalism" could be used in other ways as well. For instance, it could refer to *who moral standards apply to*, i.e., the *scope* of moral concerns (Goodwin & Darley, 2008). What matters is the content of the construct (in this case, the notion that there is a single moral standard), and not the term we use to refer to it.

¹⁰² Some theories allow for both stance-dependence and universalism, including ideal observer theory, which holds that moral facts are those facts that an ideally rational and fully informed person would endorse (Joyce, 2021). There may be a single universal moral standard, but it would still be made true by a stance, even if it is the stance of a hypothetical agent. More generally, one could in principle hold that all moral stances depend on a single agent's stance. This would make this account a form of moral antirealism (since moral facts are stance-dependent) but universal (since there is only one correct moral standard).

identical items, such as “Judgments about whether an action was morally right or wrong can be correct or incorrect.” All items are available in **Appendix D**.

Results

Cronbach’s alpha was calculated for the *third person* measure for each of the subscales. Following convention, Cronbach’s alpha was poor for the 3-item realism subscale ($\alpha = .49$) and good for the 3-item universalism subscale ($\alpha = .86$) and the 3-item relativism subscale ($\alpha = .81$). However, Cronbach’s alpha for the 3-item noncognitivism scale was exceptionally low ($\alpha = .19$). This pattern changed somewhat when assessing participant’s self-reported metaethical positions (i.e., *first person* judgments). Cronbach’s alpha was adequate for the realism subscale ($\alpha = .78$), very high (and possibly somewhat redundant; see Tavakol & Dennick, 2011) for the universalism subscale ($\alpha = .92$), good for the relativism subscale ($\alpha = .82$), and again extremely poor for the noncognitivism condition ($\alpha = .42$). There are considerable limitations with relying on Cronbach’s alpha as a measure of internal consistency or unidimensionality (Sijtsma, 2009).¹⁰³ As such, the quality of each subscale was also assessed by conducting exploratory factor analysis for both the third person and first person scales. Results can be seen in **Supplement 6**, section **S6.1.1**. Put briefly, these findings suggest that while items on the relativism subscale loaded reasonably well onto their own factor in both the self and other conditions, realism and universalism items all loaded onto the same factor in both conditions, suggesting that participants may make little to no distinction between them. Critically, in both the self and other conditions, none of the three items in the noncognitivism subscale loaded onto the same

¹⁰³ Sijtsma puts it bluntly:

“Alpha is not a measure of internal consistency. Neither is it a measure of the degree of unidimensionality [...] Alpha has been shown to correlate with many other statistics and much as these results are interesting, they are also confusing in the sense that without additional information, both very low and very high alpha values can go either with unidimensionality or multidimensionality of the data. But given that one needs the additional information to know what alpha stands for, alpha itself cannot be interpreted as a measure of internal consistency. (2009, p. 119)

factor, and most loadings did not exceed 0.3 on any factor. As a result, the cognitivism subscale was dropped from all subsequent analyses.

For the third person measures of metaethical beliefs, the mean for the realism subscale was above the midpoint ($M = 5.26$, $SD = 0.97$), as were scores for the universalism subscale ($M = 5.46$, $SD = 1.15$), while the mean score for relativism subscale was below the midpoint ($M = 3.43$, $SD = 1.43$). Overall, these findings suggest that people on Amazon's Mechanical Turk in the US tend to think, at least in abstract cases where no concrete moral issue is specified, that a typical person in their society tends to endorse realism and universalism, and to reject relativism. For first person measures of metaethical beliefs, the mean score for the realism subscale was marginally above the midpoint ($M = 4.19$, $SD = 1.51$), and the mean score for universalism was, surprisingly *exactly* at the midpoint ($M = 4.00$, $SD = 1.80$), while the mean score for relativism was slightly above the midpoint ($M = 4.46$, $SD = 1.54$). These findings indicate that participants exhibit no strong tendency towards realism or antirealism, and may slightly favor relativism on average. Descriptive statistics for all measures appear in **Table 6.1**.

Table 6.1

Descriptive statistics for Study 1: Third person paradigm (pilot)

condition		n	mean	SD
realism	- abstract	130	4.96	0.990
	- right	132	5.31	0.928
	- wrong	128	5.52	0.902
universalism	- abstract	130	5.01	1.264
	- right	132	5.54	1.064
	- wrong	128	5.83	0.946
relativism	- abstract	130	3.86	1.385
	- right	132	3.27	1.364
	- wrong	128	3.16	1.449

Averaging scores within each subscale, there was a small positive correlation between third person and first person responses for realism, $r(388) = .13, p = 0.01$, universalism $r(388) = .11, p = 0.03$, and relativism $r(388) = .16, p = 0.002$.

A one-way ANOVA and pairwise comparisons across conditions (*no statement*, *abstract right*, and *abstract wrong*) were conducted for each of the three subscales, in both the third person and first person measures.¹⁰⁴ A one-way ANOVA was performed to compare the effect of condition on third person realism subscale. Consistent with predictions, results showed that there was a significant main effect of condition on realism scores $F(2, 258) = 11.54, p < .001$. Games-Howell post-hoc tests were conducted for all pairwise comparisons. As predicted, the mean realism score for the no statement condition ($M = 4.96, SD = .99$) was significantly different than the mean realism score for the abstract right condition ($M = 5.31, SD = .93$), $t(258) = -2.94, p = .01$, and it was also significantly different than the mean realism score for the abstract wrong condition ($M = 5.52, SD = .90$), $t(254) = -4.79, p < .001$. Contrary to predictions, there was no significant difference between the abstract right and wrong conditions, $t(258) = -1.90, p = .142$.

Consistent with predictions, a one-way ANOVA revealed a significant main effect of condition on third person universalism scores $F(2, 255) = 17.30, p < .001$. Pairwise comparisons indicate that the mean universalism score for the no statement condition ($M = 5.01, SD = 1.26$) was significantly different than the abstract right condition ($M = 5.54, SD = 1.06$), $t(251) = -3.67, p < .001$, and the abstract wrong condition ($M = 5.83, SD = .95$), $t(239) = -5.89, p < .001$. The mean universalism score for the abstract right condition ($M = 5.54, SD = 1.06$). Contrary to predictions,

¹⁰⁴ I employed Welch's ANOVAs throughout. A Shapiro-Wilk test of normality revealed that none of the six conditions appeared to be normally distributed. More generally, Welch's tests are likely to perform better under a variety of conditions common to social psychological research, and are less likely to inflate Type I error rates compared to classical (Fisher's) ANOVA since the latter relies on assumptions about the data that many datasets don't meet (Delacre et al., 2019).

there was no significant difference between the abstract right and wrong conditions ($M = 5.83$, $SD = .95$), $t(256) = -2.31$, $p = .057$.

Consistent with predictions, a one-way ANOVA revealed a significant main effect of condition on third person relativism scores $F(2, 258) = 9.22$, $p < .001$. Pairwise comparisons indicate that the mean relativism score for the no statement condition ($M = 3.86$, $SD = 1.39$) was significantly different than the abstract right condition ($M = 3.27$, $SD = 1.36$), $t(260) = 3.47$ $p = .002$, and the abstract wrong condition ($M = 3.16$, $SD = 1.45$), $t(255) = 3.92$, $p < .001$. Contrary to predictions, there was no significant difference between abstract right and wrong conditions $t(256) = 0.59$, $p = .824$.

Although there was no specific prediction, one-way ANOVAs revealed no main effect of condition for first person realism, universalism, or relativism scores.¹⁰⁵ Summary tables and graphs for the third person tests are available in **Supplement 6**, section **S6.1.2**.

Discussion

Overall, findings were consistent with predictions with respect to the difference between the no statement and abstract right and wrong statements, but not with respect to the difference between the abstract right and wrong statements. Consistent with expectations, participants perceived other people to endorse realism and universalism and to reject relativism the most when those people stated that an action was wrong, and the least when they provided no statement. Findings were less consistent with demonstrating a consistently intermediate pattern for those who stated an action was right. While there was a significant difference between the no statement and right conditions for all three subscales, the difference between the abstract right and wrong statement conditions was only significant for universalism. Although it may have approached significance for the realism condition, it fell very far from significance when comparing relativism scores between the two conditions. This indicates that

¹⁰⁵ A one-way ANOVA was conducted for each of the three subscales for the first person condition. A one-way ANOVA found no significant effect for realism $F(2, 258) = 0.66$, $p = .518$, universalism $F(2, 258) = 1.27$, $p = .282$, or relativism, $F(2, 258) = 2.42$, $p = .091$.

while the difference between the no statement and abstract statement conditions may be robust, there may be little or no difference in perception of metaethical beliefs between those who state an action is right and those who state that it is wrong, or, at the very least, this may not be true for perception of relativism in particular.

While these results were the only predictions that were made explicit, they are of secondary interest to what I take to be the more interesting implications of these findings. Among other considerations, the results of Study 1 suggest that MTurk participants in the United States tend to view typical people in their society as exhibiting a slight tendency towards realism and universalism, and a slight tendency against relativism. This was especially true in the abstract right and wrong statement conditions, compared to the no statement condition, which was closer to the midpoint. In the absence of any additional information, people from this population may be hesitant to draw strong conclusions about other people's metaethical standards.

There was also a slight tendency towards judging a typical person to be especially likely to endorse realism and universalism, and reject relativism, when that person stated that an action was morally wrong rather than morally right. However, this tendency was not significant. This may indicate a very slight tendency to judge people who have a more restrictive moral attitude to be realists/universalists, and to reject relativism, an effect which could perhaps be detected in a sample with greater power. However, even if there were such an effect, it would appear to be weak. This is surprising, since realism (though it's less clear the same holds for universalism) is associated with a more intolerant and close-minded attitude towards people with contrary moral beliefs. It seems plausible in light of this that judging an action to be morally wrong would be a stronger cue towards a less permissive and more closed perspective. It could simply be that the difference between conditions is too insignificant, given how the conditions are, overall, extraordinarily impoverished in

terms of the information given. It may be that in real-world circumstances, people are sensitive to a variety of cues to infer other people's metaethical views that were absent from these conditions.

Yet the most important finding is that people's judgments about a typical person's metaethical standards were only very weakly correlated with their own first person metaethical judgments. This conflicts with traditional philosophical presumptions. Philosophers have historically presumed that all competent members of a particular linguistic community share a uniform and determinate set of metaethical presuppositions implicit in the way they speak and think (Gill, 2009). Yet participants do not appear to assume the typical person shares their own metaethical presuppositions. Of course, philosophers could maintain both that people do share the same metaethical presuppositions, but that ordinary people's judgments about how other people's metaethical stances and commitments may reflect a kind of second-order incompetence or performance error (Martí, 2009).

This could be true, but if so, why the disparity between first and third person judgments? Are we to consider people accurate judges of their own metaethical standards, but incompetent with respect to judging a typical person's standards? And that people have made the further mistake of presuming other people have different metaethical standards than they do, even though *ex hypothesi* this is false (or even impossible) since all competent members of a given community must share the same uniform metaethical presuppositions (i.e., *everyone* must either be a realist about *all* moral claims or *none*)? Perhaps so, but we're not entitled to presume this is the case. Philosophers are in no position to dismiss empirical data merely on the basis of it conflicting with their idealistic conceptions of people *ought* to speak and think. The matter calls for empirical resolution.

Taken at face value, these findings suggest that ordinary people have little trouble supposing that other people *don't* share their metaethical presuppositions. And this raises important methodological questions with respect to existing metaethics paradigms. First, suppose that participants are a representative sample of the population about whom they are judging in the third

person measures. If so, it is possible that participants have an inaccurate picture of how other people think: people mistakenly think other people are more inclined towards realism and universalism, and less inclined towards relativism, than themselves. This would be an important discovery all on its own, and one well worth exploring, given that one's perception of other people's metaethical beliefs may have important psychological and behavioral implications. It's also possible people's third person judgments provide a more accurate picture of folk metaethics. People may be more subject to normative conflation when asked about their own metaethical standards. If so, the disparity between third person and first person judgments may be due in part to the minimization of normative conflation, though present findings cannot resolve the matter.

These findings suggest another possibility, however. It may be that participants *don't* represent the population they have in mind when judging third person items. If participants are judging a typical person in their society, i.e., someone in the United States, but participants are skewed towards particular demographics dimensions relevant to their own metaethical standards, then differences between first and third person judgments could both be accurate. If so, there may be another explanation for these results: they may reflect an accurate assessment of the typical American, relative to participants. Although previous studies have found no association between political conservatism and realism when judging concrete moral issues (Goodwin & Darley, 2008, Pölzler & Wright, 2020b), they *have* found an association between conservatism and realism for *abstract* measures (Pölzler & Wright, 2020b). Collier-Spruel et al. (2019) also found that scores on their Moral Relativism Scale (MRS) were negatively associated with a conservative political orientation, right wing authoritarianism, but positively associated with progressive values.¹⁰⁶ In addition, participants who choose realist responses have consistently reported greater religiosity while low religiosity has been consistently

¹⁰⁶ Progressive values were calculated by subtracting the three foundations emphasized more by American conservatives—purity, ingroup loyalty, and respect for authority—from their average score on the remaining two foundations, harm and fairness, which tend to be emphasized more by American liberals (Collier-Spruel et al., 2019, p. 12; Graham et al., 2009).

linked to antirealist responses (Collier-Spruel et al., 2019; Goodwin & Darley, 2008, Yilmaz & Bahçekapili).

If participants are less religious and conservative than whoever they have in mind when responding to third person items, this could account for the difference between third and first person judgments. Participants in the present sample were on average more inclined towards liberalism than conservatism ($M = 3.58$ $SD = 1.80$) and were extremely low in religiosity ($M = 2.99$, $SD = 2.16$), with 43.8% of participants selecting 1 out of 7 for “not religious at all.” In line with research assessing demographic differences between MTurk participants and the general population, these results show that the participants in this sample are both less conservative than the general population and *much* less religious. As one recent study found, “a strikingly substantial percentage of the MTurk sample—over 40%—identifies as agnostic or atheist,” while a more representative sample of the general population finds only 10% of the population identifies as agnostic or atheist (Levay, Freese, & Druckman, 2016, p. 4).

This suggests a potentially major advantage of third person paradigms over first person paradigms: first person paradigms may be more susceptible to poor generalizability, insofar as the samples researchers tend to draw on are skewed towards particular demographic variables that render them less representative of the samples they are drawn from in demonstrably relevant ways. Third person paradigms may serve as a buffer against unrepresentative sampling, and provide a more accurate picture of the metaethical views of the populations they are drawn from. If so, present findings would suggest that researchers may be underestimating the degree to which people in the United States endorse realism, insofar as present sampling methods are skewed towards less religious and less conservative participants.

There are a number of important observations about these findings that are relevant both to subsequent studies reported here and to future research on folk metaethics. One intriguing takeaway

is the lack of distinction between realism and universalism. Psychologists interested in mapping latent variables in a bottom-up way, without the theoretical baggage imported from philosophy, may shrug and suggest that these subscales should be collapsed into a single realism-universalism dimension. Philosophers may balk at such a notion. They may point out that realism and universalism *are not conceptually identical*, and that at least under ideal circumstances a reflective person ought not to conflate them, any more than one ought to treat a square and a rectangle as the same shape. Whatever affinities the two may have, there is a difference, and it would be quite a concession to make to accept that ordinary people do not conceptually distinguish realism and universalism.

I cannot adjudicate this dispute here, but it is worth noting that both sides have points in their favor; on the one hand, philosophers are not entitled to insist that folk psychology must comport with the conceptual distinctions that animate contemporary analytic philosophy. On the other hand, psychologists may be too quick to treat as distinct or indistinct, as the case may be, features of ordinary thought, rather than consider the possibility that the patterns observed in their data are a methodological artifact: it could be that the lack of distinction between realism and universalism is due to a pattern of unintended interpretations that yield a converging response pattern but does not reflect a genuine psychological construct (i.e., some shared folk notion that combines realism and universalism). Psychologists should be cautious in identifying new psychological constructs when results could be due to the poor validity of their measures, especially when there is some theoretical rationale for suspecting such constructs would entail attributing conceptually muddled notions to ordinary human thought. For the purposes of these studies, realism and universalism will be treated as distinct. Assessment of the degree to which people distinguish or fail to distinguish realism and universalism will be left to future research.

It's also unclear why Cronbach's alpha was so low for the realism subscale in the third person condition, given that it was much higher in the first person condition. It may be that people are less

certain about the metaethical beliefs of a typical person than themselves, especially when they are given little or no information about people in their society. However, this doesn't explain why Cronbach's alpha was high in the third person versions of the universalism and relativism subscales.

Examination of the frequency of participants who selected each of the possible responses on the Likert scale suggests participants are fairly evenly distributed across all response options, with roughly half above and below the midpoint (a finding which holds true even without collapsing the no statement, right, and wrong conditions). At face value, this suggests people have highly variable views towards moral realism, though the even distribution of responses could suggest that participants had trouble interpreting the items and chose randomly. One might wonder why acquiescence didn't favor overall average agreement, though this could have been offset by the participants in the sample exhibiting a tendency towards lower realism and universalism and greater relativism. The lack of main effects in the self conditions may also be due in part to the greater standard deviations for responses to the first person versions of the subscales, compared to the third person versions of those scales. First person scales may require larger sample sizes to detect significant differences. While this may be a cost to employing first person measures, it may be necessary insofar as it captures genuinely greater individual differences in self-reported metaethical views, compared to third person judgments.

Previous studies have shown differences in the proportion of realist and antirealist responses when considering morality in the abstract versus considering concrete moral issues (Pözlner & Wright, 2020a; 2020b). Most importantly they've shown high levels of variation between different concrete moral issues, and in line with previous research, results often vary in accordance with concrete moral issues (e.g., Beebe, 2015; Beebe & Sackris, 2016; Goodwin & Darley, 2008; 2012; Pözlner & Wright 2020b, Wright, Grandjean, & McWhite, 2013). In the next study, I sought to expand on the present findings by exploring third personal judgments in response to a set of concrete moral issues.

6.2 Study 2: Third person paradigm with concrete moral issues

Study one introduced the third person paradigm. This involved assessing how participants evaluated other people's metaethical beliefs either when asked about a typical person in their society, or when presented with a person who made a moral assertion that an unspecified action was morally right or wrong.

This study expands on Study 1 by introducing four concrete moral statements in which the participant is told that someone asserted that a particular action was morally wrong. This study also employs a within-subjects design, with all participants responding to both the no statement condition in Study 1 and each of the four concrete moral issues. I also introduced a handful of items that measure how participants evaluate the character of the person who made the statement. These measures were introduced primarily for exploratory purposes, so there were no predictions related to them. Finally, the sample size was increased. These changes were intended to both assess the results of the third person paradigm for concrete moral issues and to do so under conditions with substantially greater power. Once again, since these findings represent the introduction of a novel paradigm for measuring folk metaethical stances and commitments, my goals were largely exploratory. Due to the exploratory nature of these findings, I have relegated detailed description of the methods and results of Study 2 to **Supplement 6**, section **S6.2**.

6.3 Study 3: Third person paradigm (nonmoral domains)

The goal of Study 3 is to expand the third person paradigm to nonmoral domains. One of the most fascinating and important elements of moral judgments is the way they differ from nonmoral judgments, such as matters of taste and claims about science and history. Although people may turn out to have more diverse and nuanced views than researchers imagine, one might generally suppose that people lean towards realism about factual matters in math, science, history, and so on, but lean more towards antirealist views such as subjectivism when thinking about matters of personal taste.

Matters of social convention may likewise be thought of not as stance-independent facts, but as culturally constructed norms and institutions that exist by social consensus. Previous research on folk metaethics has occasionally sought to compare judgments in each of these domains: morality, taste (and aesthetics), social convention, and facts (e.g., Beebe, 2015; Beebe et al., 2015, Goodwin & Darley, 2008, Wright, Grandjean, & McWhite, 2013). These findings generally comport with what researchers might expect: in descending order, realist ratings are highest for the factual domain, followed by morality, social conventions, and taste.

This general pattern is complicated by extremely high domain variation *within* each domain, meaning that many participants choose realist responses for some issues and antirealist responses for others within each domain. For Instance, Beebe (2015) found substantial variation in responses towards factual issues (such as scientific claims), with a consensus in favor of realism towards some and antirealism towards others, while Goodwin and Darley (2008) found that their participants were almost evenly divided about matters of social convention (Goodwin & Darley, 2008), prompting researchers to conclude that people may be *pluralists* about these topics (Pözlner & Wright, 2020b; Wright, Grandjean, & McWhite, 2013).

The reasons *why* people are (or at least appear to be) pluralists are complex and have yet to be fully explicated. At least one reason people may only appear to be pluralists about moral issues could be due to normative conflations, though, as I have argued at length in preceding chapters, a variety of other factors likely influence variable response patterns as well. Yet the factors contributing to apparent pluralism in the moral domain may differ in relative importance for nonmoral domains, or may differ altogether. For instance, epistemic conflations may play an important role when considering claims in non-normative domains, e.g., descriptive statements about science, history, or math. Beebe (2015) found that when participants were asked to judge disagreements about uncontroversial factual issues, realist response rates were very high, a finding consistent with previous research. However,

when they were asked to judge disagreements about factual disputes issues for which there is greater uncertainty, or no practical way to know what the correct answer is, antirealist response rates shot up dramatically, approaching and in some cases exceeding 50%. For instance, about 45% of participants judged that if two people disagreed about whether Julius Caesar drank wine on his 21st birthday, they could both be correct. We *could* interpret this as evidence that around half of Beebe’s participants believe there is no objective fact about certain historical events, but it is far more plausible that these participants did not interpret this question as intended. Given that the greater the *uncertainty* about the factual dispute in question, the greater the “antirealist” response rate, the best candidate for these findings are epistemic conflation.

Critically, Beebe’s findings also suggest that participants may interpret analogous stimuli in different domains in different ways, even when virtually all other aspects of the design are held constant. For instance, if participants are asked to judge an otherwise identically-phrased disagreement between two people, disagreements about moral issues and disagreements about scientific or historical disputes may be interpreted differently. This possibility threatens inter-domain comparisons, since it suggests that there may not only be substantial methodological shortcomings with questions about realism and antirealism in general, but that there may also be an additional layer of methodological shortcomings whenever one seeks to compare judgments about realism and antirealism in different domains: people may interpret the “same” question about whether, e.g., if two people disagree, both can be correct, differently based entirely on the domain of disagreement.

If so, responses to what is intended to be the same question across domains would effectively function as responses to different questions, undermining our ability to make inter-domain comparisons. In short, it is important to devise paradigms with cross-domain interpretative consistency...yet, notably, most research making inter-domain comparisons so far has relied

exclusively on the disagreement paradigm with all its attendant methodological shortcomings.¹⁰⁷ This makes the rationale for studying these domains using a novel paradigm especially worthwhile, but it comes at the cost of abandoning the primary rationale for the third person paradigm in the first place: to minimize normative conflations. Such conflations seem far less likely in nonmoral domains. Even so, this is offset by the potential advantages of introducing a novel method for assessing realism, universalism, and relativism in nonmoral domains.

It would also be helpful to establish just what these nonmoral domains are supposed to reflect. The *taste* domain encompasses aesthetic judgments about art, music, and other matters of creative expression, as well as personal taste in e.g., food and drink. For example, one item from the realist subscale states:

They believe there are facts about which foods taste good or bad that are true even if some people or societies think otherwise.

Social conventions refer to nonmoral norms regarding appropriate social conduct, such as rules about how to dress, how to refer to others, how to greet other people (bowing, shaking hands, and so on), use of body language, how much personal space to accord others, and so on. While many Western populations may reliably distinguish paradigmatic moral norms from social conventions, the boundary between these domains often blurs, and people vary both within and between populations in the extent to which they regard a given issue as a matter of morality, social convention, or neither (see **Appendix E**). Nevertheless, researchers may still draw *a priori* distinctions when designing measures such that, if participants do not themselves conceive of moral and social conventions as distinct, this may or may not be reflected in their pattern of responses, and need not jeopardize the validity of the measures.

Finally, the factual domain refers to *descriptive* (rather than normative) claims that purport to tell us what the world is like. This domain encompasses claims about science, history, math, and other

¹⁰⁷ I address *why* this has occurred in **Supplement 6**, section **S6.3.1**.

related claims. The following is perhaps the most illustrative example of an item in the factual domain (from the realist subscale):

They believe that claims that describe what the world is like, such as the claims made in geography, history, and physics, are correct regardless of the preferences or values of different individuals or cultures.

This is arguably the least conceptually well-defined category. Roughly, the factual domain refers to claims ostensibly intended to reflect what the world around us is like, and more generally issues about which one might imagine people are generally inclined to think are discovered, rather than created or a matter of individual preference or social consensus. This is an admittedly somewhat unsatisfying way of defining the domain since, after all, the whole point of distinguishing these domains is to determine whether people endorse realism, universalism, and relativism about the domain in question. It seems a bit strange, given this, to define the domain as presupposing e.g., realism by definition. This is why it may be more helpful to distinguish this domain from others by noting that the claims in question are not normative, but instead consist of *descriptive* claims, claims intended to describe some state of affairs. Perhaps it would be more helpful to describe the *descriptive* domain.¹⁰⁸

My general expectation is that we would find roughly the same pattern for the third person versions of these measures: participants would generally favor realism and universalism but reject relativism in the factual domain, and this pattern would decline and perhaps reverse as one moved in descending order to the moral domain, to social conventions, and finally to matters of taste. Consistent with the results of Study 1 and Study 2, I expected participants to continue to judge a typical person to lean towards realism and universalism and to reject relativism. I expected this pattern to reverse for matters of taste, with third person judgments indicating that participants thought a typical person rejected realism and universalism about such matters, and was more inclined towards relativism. It is less clear what to expect in the case of social conventions. They tend to fall somewhere in the middle, with

¹⁰⁸ See **Supplement 6** section **S6.3.2** for additional commentary.

Goodwin and Darley (2008)'s findings placing them closer to facts and morality, but Wright et al.'s (2013) findings suggesting they perform more similarly to matters of taste. Nevertheless, I expected social conventions to perform closer to matters of taste.

Methods

Participants. I aimed to recruit 800 participants. 801 participants began the study. However, one participant did not complete the study and was excluded from further analysis. This left a total of 800 participants. Participants consisted of 800 adult US residents on Amazon's Mechanical Turk (430 females, 362 males, 8 other, $M_{\text{age}} = 41.9$, $SD_{\text{age}} = 12.3$, age range = 19-78).

Procedure. The procedure employed in Study 3 was almost identical to Study 1. All participants were randomly assigned to one of four conditions: moral, aesthetic, social convention, or facts. All conditions were identical to the no statement condition in Study 1, in that participants were asked to consider the beliefs of a typical person in their society, without being provided any additional information. The only difference between conditions were the measures. The moral condition used the same measures as Study 1, with the exclusion of the three-item noncognitivism subscale, resulting in a total of nine items: a three-item realism subscale, a three-item universalism subscale, and a three-item relativism subscale. Order of all nine items was randomized within condition. The measures in each of the nonmoral conditions also consisted of three-item subscales for realism, universalism, and relativism, each adapted to reflect the respective realist, universalism, and relativist notions of truth in the relevant domain and, where possible, to serve as close analogs to their counterparts in the morality subscales.¹⁰⁹ After answering the third-person versions of each question, participants were presented with first person versions of the same items, and asked to judge how much they agreed or disagreed with them. Finally, participants were asked to report their age and gender.

¹⁰⁹ I could describe these measures as metanormative positions for the moral, aesthetic, and social conventional domains, but it is less clear whether the measures used in the factual condition could be appropriately described as normative. Nevertheless items in the factual domain still reflect realism, universalism, and relativism about the relevant claims.

Measures. Measures used in the third and first person moral condition were identical to the measures used for the realism, universalism, and relativism subscales in Study 1. Three-item subscales for realism, universalism, and relativism were introduced for each of the nonmoral domains (taste, social conventions, and facts), and were likewise adapted for use as both third and first person measures. All scale items are available in **Appendix D**.

Results

The mean, standard deviation, and Cronbach's alpha for the realism, universalism, and relativism subscales for each of the four domains are available on **Table 6.2**.

My primary goal was to evaluate differences in the degree to which participants would judge a typical person in their society to endorse realism, universalism, and relativism with respect to each of the four domains: morality, taste, convention, and facts. First, I conducted a Welch's one-way ANOVA to compare the effect of domain on realism. There was a significant main effect of condition on realism $F(3, 438) = 50.2, p < .001$. All pairwise comparisons are available on **Table 6.2**, and means with 95% CIs are featured in **Figure 6.1**.

Table 6.2*Descriptive statistics for Study 3: Moral and nonmoral domains*

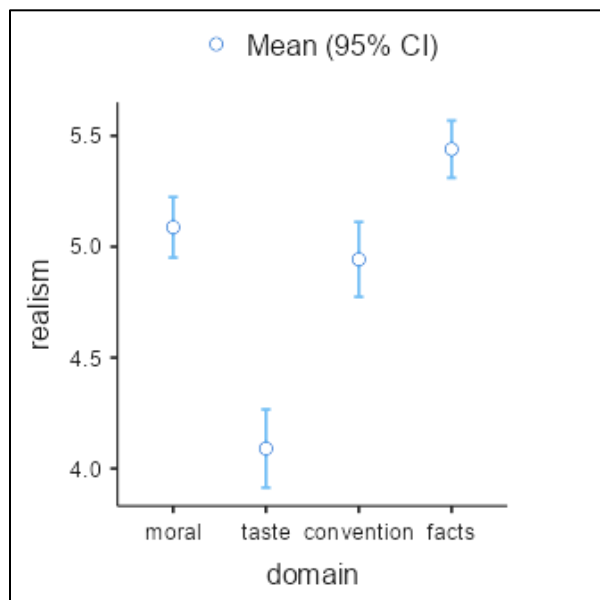
3 rd /1 st person	metaethical stance	condition	mean	SD	α
third person	realism	moral	5.09	0.98	0.576
		taste	4.09	1.25	0.686
		convention	4.94	1.20	0.787
		facts	5.44	0.93	0.635
	universalism	moral	4.88	1.42	0.913
		taste	3.53	1.35	0.776
		convention	4.37	1.48	0.838
		facts	5.03	1.12	0.635
	relativism	moral	3.45	1.31	0.801
		taste	5.06	1.20	0.808
		convention	4.53	1.32	0.854
		facts	3.72	1.14	0.636
first person	realism	moral	4.36	1.42	0.748
		taste ¹¹⁰	3.98*	1.72*	0.629*
		convention	3.37	1.53	0.822
		facts	5.33	1.20	0.641
	universalism	moral	4.01	1.86	0.93
		taste	2.48	1.39	0.941
		convention	2.54	1.58	0.89
		facts	5.06	1.35	0.646
	relativism	moral	4.10	1.57	0.822
		taste	5.78	1.05	0.987
		convention	5.62	1.22	0.793
		facts	3.85	1.33	0.658

*Note. The mean, standard deviation, and alpha reported for the first person realism taste subscale are based on items #2 and items #3. Item #1 was dropped from analysis due to an error.

¹¹⁰ Item #1 in the first person realism taste subscale was dropped from analysis due to an error. Specifically, I accidentally included item #3 from the first person taste relativism subscale in place of the correct item. Unfortunately, this means that only two items are included in the first person realism taste subscale. In addition, item #1 on the first person realism convention subscale and item #1 on the first person realism facts subscale both mistakenly include the word “that.” In particular, item #1 for the first person realism convention subscale states, “When two people disagree about cultural practices, such as appropriate work attire, that [sic] at least one of them must be incorrect.” Item #1 for the first person realism facts subscale states, “When two people disagree about matters of science or history that [sic] at least one of them must be incorrect.” Corrected versions of all three items appear in **Appendix D**.

Table 6.3*Post Hoc Comparisons - realism*

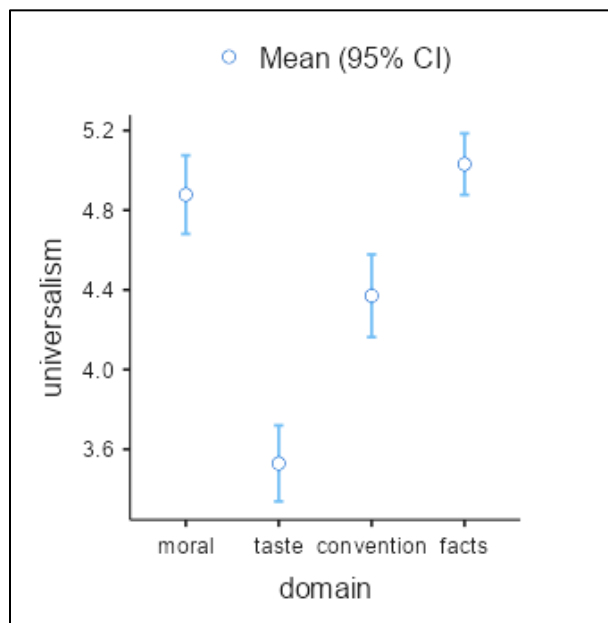
comparison		mean difference	df	t	p-value
condition	condition				
moral	- taste	0.997	370	8.82	< .001
	- convention	0.145	380	1.31	0.554
	- facts	-0.352	402	-3.70	0.001
taste	convention	-0.853	391	-6.89	< .001
	facts	-1.349	360	-12.19	< .001
convention	facts	-0.497	371	-4.61	< .001

Figure 6.1*Study 3: Comparison of means (with 95% CIs) across domains for third person | realism condition*

The next metaethical domain I assessed was universalism. There was a significant main effect of domain on universalism $F(3, 438) = 54.1, p < .001$. All pairwise comparisons are available in **Table 6.3**, and means with 95% CIs are featured in **Figure 6.2**.

Table 6.4*Post Hoc Comparisons – universalism*

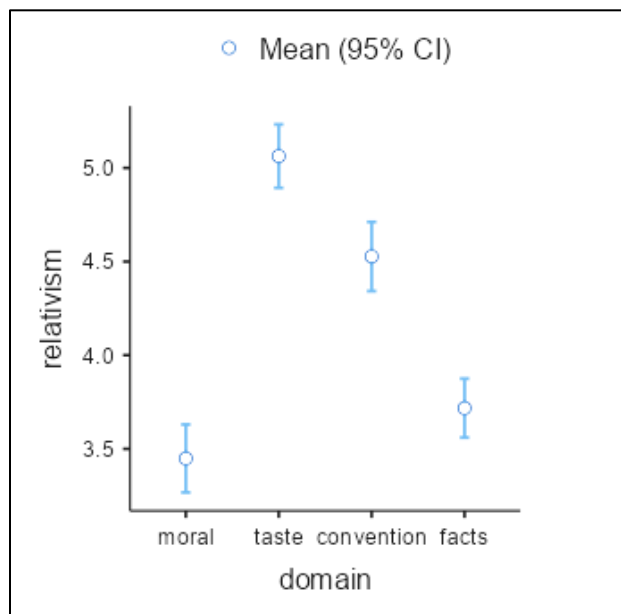
comparison		mean difference	df	t	p-value
condition	condition				
moral	- taste	1.350	396	9.70	< .001
	- convention	0.508	397	3.50	0.003
	- facts	-0.153	381	-1.21	0.623
taste	convention	-0.841	390	-5.89	< .001
	facts	-1.502	378	-12.09	< .001
convention	facts	-0.661	366	-5.04	< .001

Figure 6.2*Study 3: Comparison of means (with 95% CIs) across domains for third person | universalism condition*

Finally, I assessed domain differences with respect to relativism. There was a significant main effect of domain on relativism $F(3, 438) = 54.1, p < .001$. All pairwise comparisons are available in **Table 6.4**, and means with 95% CIs are featured in **Figure 6.3**.

Table 6.5*Post Hoc Comparisons - relativism*

comparison		mean difference	df	t	p-value
condition	condition				
moral	- taste	-1.610	395	-12.80	< .001
	- convention	-1.078	398	-8.21	< .001
	- facts	-0.269	395	-2.21	0.123
taste	convention	0.536	389	4.22	< .001
	facts	1.345	394	11.49	< .001
convention	facts	0.809	387	6.58	< .001

Figure 6.3*Study 3: Comparison of means (with 95% CIs) for third person | relativism condition*

Almost all results were consistent with expectations. I expected participants to judge that a typical person in their society exhibited comparatively higher levels of realism and universalism for facts, followed by morality, social convention, and finally taste. Two of these differences were not

statistically significant. First, with respect to realism, the difference between morality ($M = 5.09$, $SD = 0.98$), and convention ($M = 4.94$, $SD = 1.20$) was not significant, $t(380) = 1.31$, $p = 0.554$. Second, with respect to universalism, the difference between morality ($M = 4.88$, $SD = 1.42$), and facts ($M = 5.03$, $SD = 1.12$) was not significant, $t(381) = -1.21$, $p = .623$. However, all differences were in the expected order, with point estimates falling exactly in line with expectations for both universalism and relativism. This suggests that the expected pattern may generally hold, albeit in some cases the size of the difference may be small, and, while it is possible that the order of one or more domains may flip in a handful of cases, such deviations from the expected pattern would likely be very small and merely reflect an approximate lack of difference that still mostly maintains the expected ordering of realism and universalism judgments across domains.

With respect to relativism, the general pattern was also mostly in line with expectations. Relativism scores were highest for the taste domain, followed by social conventions. However, contrary to expectations, there was no significant difference in relativism scores between morality ($M = 3.45$, $SD = 1.41$) and facts ($M = 3.72$, $SD = 1.14$), $t(395) = -2.21$, $p = .123$. Critically, this difference was *not* in the expected direction, with relativism scores somewhat lower for morality than for facts. Of course, this difference wasn't statistically significant, but it still reflects a departure from expectations about the overall ordering of relativism judgments across domains. It is possible that this difference is due to the particular stimuli I used. The items used for the factual domain were first introduced in this study, and their general psychometric properties have yet to be fully explored. As such, future studies could vindicate the hypothesized pattern. Of course, it is also possible that people genuinely do tend to judge a typical person in their society to be approximately similar with respect to relativism for both moral and factual issues, or to even regard morality as less relativistic than factual claims. Given serious concerns about adapting paradigms designed to assess moral disagreements to factual disagreements (see Beebe, 2015), it may be that the factual domain is subject to distinctive

methodological concerns. In particular, participants may conflate epistemic considerations with metaphysical ones, resulting in implausibly high rates of apparent “relativism” for factual disputes. As such, I’m skeptical of this inconsistency, though of course such skepticism should be qualified by the clear motivation to not want to believe results contrary to my expectations.

One result that did unambiguously fly in the face of expectations was the overall performance of social conventions. I expected social conventions to either be significantly less than or not significantly different from the midpoint for realism and universalism, while I expected social conventions to be significantly above the midpoint for relativism, or at least not significantly different from it. In other words, I expected people to tend to regard social conventions as relative, but not stance-independently true or universal. To test for these hypotheses, I conducted right-tailed one-sample t-tests to compare the mean score for the convention domain against the midpoint (i.e., $H_a \mu > 4$) for realism, universalism, and relativism. Contrary to predictions, the mean score for realism ($M = 4.94$, $SD = 1.20$) was significantly *greater* than the midpoint $t(197) = 11.02$, $p < .001$, $d = 0.783$. Also inconsistent with predictions, the mean score for universalism ($M = 4.37$, $SD = 1.48$) significantly greater than the midpoint, $t(197) = 3.52$, $p < .001$, $d = 0.250$. However, consistent with predictions, the mean relativism score for conventions ($M = 4.53$, $SD = 1.32$) was significantly greater than the midpoint, $t(197) = 5.63$ $p < .001$, $d = 0.400$.

Discussion

Overall, the results of Study 3 mostly corroborated expectations. Consistent with what I expected, participants tended to regard a typical person in their society as strongly inclined towards realism and universalism for factual disputes, and disinclined towards relativism. This same pattern held, in descending order of support for realism and universalism and rejection of relativism, for their judgments about morality, social conventions, and taste.

However, I expected the factual and moral domains to remain above the midpoint for realism and universalism, and below the midpoint for relativism, but for this pattern to flip for both social conventions and matters of taste. Instead I found that while social conventions appeared in the expected order (i.e., between morality and taste) for all three metaethical dimensions, participants' judgments of realism and universalism were both above the midpoint, which unexpectedly indicates that participants thought of a typical person as more inclined on average towards realism and universalism. In other words, if participants interpreted items in the conventional domain as intended, this would mean that they believe a typical person in their society is more inclined to think of social conventions such as how to dress and how to greet one another as stance-independent normative facts that are universally applicable to everyone, including people *outside* that person's culture.

Perhaps participants do think this way, but if so, this would be a surprising result. To add to the puzzling nature of these results, if participants were generally inclined towards unintended interpretations of the conventional domain in a way that inclined them towards thinking of it in realist and universalist terms, one might also expect participants' first person judgments to also be significantly above the midpoint. This is not what I found. Left-tailed one-sample t-tests revealed, just as I would expect, that the mean realism ($M = 3.37$, $SD = 1.53$) score for first person judgments was significantly *below* the midpoint $t(197) = -5.82$ $p < .001$, $d = -0.414$, and the mean score for universalism ($M = 2.54$, $SD = 1.58$) was likewise significantly below the midpoint, $t(197) = -13.00$ $p < .001$, $d = -0.924$, while a right-tailed test found, again consistent with expectations, that the mean relativism score ($M = 5.62$, $SD = 1.22$) was above the midpoint, $t(197) = 18.78$, $p < .001$, $d = 1.335$.

This is exactly the pattern I would expect. The disparity between first and third person is thus at least somewhat more likely to reflect a genuine difference in people's conception of a typical person relative to themselves, and counts somewhat against dismissing such findings as a methodological fluke. I am not sure what to make of these particular findings. While I expect they may still reflect a

non-obvious pattern of unintended interpretations, evaluating the validity of these measures and assessing how participants are interpreting items within each of these domains is a task for future research.

6.4 Study 4: Normative entanglement

Study 4 diverges from the third person paradigm in order to assess a proposed phenomenon that could play a role in accounting for why people are on occasion disposed to express moral realism, even if they don't actually endorse it, or endorse it based on mistaken inferences about its implications: *normative entanglement*. Normative entanglement occurs whenever a metaethical claim is embedded in a more complicated assertion that also includes normative content, such that whoever asserts or endorses the metaethical element of the claim pragmatically implies information about their normative moral stance and/or their attitude towards a given normative moral issue, *even though such implications are not logically entailed by the expression of a metaethical stance*. The best way to illustrate how normative entanglement arises is to point to a hypothetical exchange between a moral realist and a moral antirealist:

Realist: *"Do you think it's objectively wrong to torture babies for fun?"*

Antirealist: *"No, I don't think it's objectively wrong to torture babies for fun."*

Realist: *"Wow! So you think it's totally okay to torture babies for fun? You're a disgusting monster!"*

One problem with the realist's response is a mistake with respect to the negation of a normative claim. To deny that an action is wrong does not entail that one thinks the action is "totally okay," or, to put it in less colloquial terms, it does not entail that the action in question is permissible (or, for that matter, recommended, required, supererogatory, and so on). All denying that an action is wrong would

entail is the denial that it is “wrong”; it doesn’t *necessarily* entail any positive normative stance towards the issue in question.¹¹¹

However, this concern is moot, because the antirealist’s response does not logically entail *any normative position at all*. Denying that an action is objectively wrong *is not* a denial that it is wrong. Merely because someone does not think that an action is *objectively* wrong, it does not follow that they don’t think it’s wrong in some non-objective way. For comparison, you could think that your favorite flavor of ice cream is good, but this does not mean that you must believe it is *objectively* good, and that anyone who prefers another flavor of ice cream is mistaken (Loeb, 2003). Having a normative moral stance doesn’t *require* objectivity. *Some* moral antirealist positions may also entail normative implications. A thoroughgoing nihilist could reject the notion there is any respect at all in which anything could be morally right or wrong, and may also deny not only that anything is intrinsically valuable, but that anything can have meaning or value in any respect at all. On account of this, they may reject all notions of compassion or concern for the welfare of others. Such an account *is not* a logical entailment of denying moral realism. Contrary to how the rejection of realism is sometimes framed, our options aren’t restricted to either realism or a deeply nihilistic existential malaise that rejects any and all meaningful conceptions of goodness and value. A variety of moral antirealist positions are consistent with having substantive first-order moral stances about the rightness and wrongness of moral actions, including cultural relativism, individual subjectivism, constructivism, and ideal observer theory. Some expressivist accounts likewise allow antirealists to sincerely assert that actions are morally right or wrong without error or contradiction (Blackburn, 1991; Joyce, 2021). In short, antirealism only involves the rejection of stance-independent moral facts. It does not entail the rejection of normative moral standards, i.e., beliefs about what is morally right or wrong.¹¹²

¹¹¹ See **Supplement 6**, section **6.3.3** for additional commentary.

¹¹² See **Supplement 6**, section **6.3.4** for additional commentary.

Rejecting the notion that a given action is objectively wrong likewise does not logically entail anything about one's attitudes or emotional responses towards the action in question. Someone who denies that it's objectively wrong to torture babies can still oppose baby torture. Antirealism does not require moral indifference or a muted emotional response to moral transgressions. Denying that something is objectively wrong is completely consistent with finding the action in question repugnant, viewing those who commit such actions with contempt, supporting laws against the action, wanting to see perpetrators punished, and so on.

For comparison, denying that chocolate cake is *objectively* tasty doesn't entail that you believe chocolate cake doesn't taste good, or that you find it disgusting, or that you are as equally willing to eat chocolate cake as a bucket of dirt (Loeb, 2003). A person who denies realism about gastronomic facts could have equally strong food preferences, and be just as disposed to judge foods as "good" or "bad" as a realist. The heart of the problem, then, is an equivocation between *good* and *objectively good*, where one mistakenly infers that to deny that something is objectively good is to deny that it is good, full stop. This is a mistake because our normative and evaluative judgments about whether things are right or wrong, and good or bad, respectively, need not *require* realism. Metaethical and normative considerations are conceptually distinct.

Of course, such rhetorical ploys likely occur primarily in academic disputes, since ordinary people are less disposed to engage in philosophical arguments about moral realism and antirealism. Explicit metaethical language may appear in everyday discourse on occasion (though efforts could be made to evaluate how often and under what circumstances). Yet the specific example given here merely reflects an explicit form of the phenomenon. Normative entanglement could appear in far subtler ways in everyday discourse. More importantly, it reflects a distinct way in which pragmatic considerations, such conversational implicature, can influence the inferences people make about what others say (Grice, 1975; Levinson, 2000). Unfortunately, both philosophers and psychologists often

ignore, or at least fail to adequately consider, the potential role pragmatics could play in influencing how participants respond to stimuli. In some cases, pragmatics may fully drive participant responses in ways that create the illusion of substantive psychological phenomenon (Adams & Steadman, 2004a; 2004b; Lindauer and Southwood, 2021; cf. Sytsma, Bishop, & Schwenkler, 2022).

This shortcoming highlights yet another methodological vulnerability to research on metaethics: participants may make inferences about the implications of selecting a realist or antirealist response that cause them to select responses intended to express a particular metaethical position for reasons unrelated to endorsing the actual content of the metaethical position itself. In practice, this may involve participants inferring that a response option intended to measure antirealism carries additional implications, and these implications are what drive them to not choose the antirealist response, *not* the fact that the statement expresses antirealism. More generally, if those inferences are *good*, if, e.g., participants perceive endorsing an antirealist stance to imply a desirable normative moral stance, or to have socially desirable implications, such as signaling tolerance or open-mindedness, participants may favor those responses, either because they mistakenly believe the position entails those implications, or because, even if they do not, they judge (quite possibly *accurately*) that selecting that response would signal positive character traits, even though the response does not accurately reflect their metaethical stance.

Conversely, if selecting a particular metaethical response would signal undesirable attributes, such as a normative moral stance the participant rejects, or an attitude that could harm their reputation, they may shy away from such responses, again, not because such responses accurately reflect their metaethical stances, but because they want to avoid signaling undesirable information about themselves.

Here, I will focus exclusively on the latter possibility. For instance, expressing an antirealist stance towards a serious moral transgression could signal inadequate opposition or insufficient

repugnance, not merely in arcane academic disputes, but in real-world social contexts. Imagine Alex is a politician who has just launched her first political campaign. The public doesn't know who she is. She has studied metaethics, and privately endorses moral antirealism. Catching wind of this, her political rival Sam directly asks her in front of a crowd:

"Alex, do you think it would be objectively good to prevent school shootings? Yes or no?"

Technically speaking, the honest answer would be "no." And yet it seems obvious to me at least that to respond with a "no" would instantly end Alex's political career. To say "no," would not lead her audience to express curiosity about the subtleties of her antirealist position. They would not pause to consider the possibility that she could be equally or more opposed to school shootings than Sam, is just as horrified by shootings as Sam, and wants just as strongly to prevent them. Instead, they would respond with immediate outrage, disgust, and horror. At best, people would infer that Alex isn't especially concerned with school shootings. At worst, they may conclude that Alex is a raving psychopath who is utterly unconcerned with the lives of children.

This example highlights a far more important reason for studying normative entanglement than its methodological implications: if normative entanglement occurs in lab settings, it would not only illustrate people's sensitivity to pragmatics in an experimental context, it would point to a potential *social* and *reputational* role metaethical discourse could have in everyday thought and discourse. And it highlights the possibility that pragmatic concerns like these could generalize to moral psychology as a whole, and indeed, *all* of human psychology.

Researchers should be doubly concerned, and doubly attentive to two concerns: first, inattention to the role pragmatics can play in how participants interpret and think about experimental stimuli could reflect a ubiquitous, subtle, and serious threat to the validity of existing research. Second, much of how people think and speak in everyday interactions is plausibly driven not by a blind adherence to strictly stating what seems most logically consistent and philosophically astute, but to

comporting one's thought, language, and action in accordance with one's *goals*, many of which are deeply embedded in our social interests, e.g., maintaining a positive reputation, ensuring our success and the success of our friends and family, outcompeting our rivals, improving our status, and identifying enemies and allies. Given that ordinary people's interests are far more aligned with practical rather than philosophical goals we should, if anything, expect the ordinary person to be *more* sensitive to normative entanglement and related concerns than philosophers. Given the frequency of such ploys even in academic disputes, the failure of researchers to grapple with the reputational and social costs of endorsing or rejecting a metaethical position, with all the attendant implications (real or imagined) associated with doing so, is a serious shortcoming with research on metaethics.

Of course, all this pulpiteering is irrelevant if it turns out participants aren't sensitive to normative entanglement. Study 4 assesses this question by testing whether comparing perceptions of three different responses to a direct question about whether a particular action is morally wrong. Participants were presented with a hypothetical exchange where one person, Alex, asks another person, Sam, whether an action is objectively morally wrong. In the realist condition Sam responds by stating that yes, they do think the action is objectively wrong. Other participants were assigned to one of two antirealist conditions: antirealism without cancellation, and antirealism with cancellation. In the condition without cancellation, Sam responds by simply stating that no, they do not think the action in question is objectively wrong. In the cancellation condition, Sam likewise states that no, they do not think the action is objectively wrong, but they then attempt to negate any inferences others might make about what this means by adding that they do think the action is wrong, they just don't think anything is *objectively* wrong. They then add that they are opposed to the action and find it terrible. These additional remarks attempt to *cancel* any implication that the antirealist lacks an appropriate normative moral stance, and fails to exhibit an appropriately negative attitude towards the action in question.

I predicted that participants will successfully perceive Sam to endorse realism (and universalism) and to reject relativism in the realism condition, and to likewise successfully perceive Sam to reject realism and universalism and endorse realism in both antirealist conditions. My hypothesis relies on this outcome, since it is especially critical that participants continue to perceive Sam as endorsing antirealism in the cancellation condition. In addition, participants were asked to judge Sam's moral character, trustworthiness, and empathy, how judgmental Sam is, how committed Sam is to their values, how desirable Sam would be as a social partner, and to assess how religious and politically liberal or conservative Sam is. I predicted participants would perceive Sam in the realist condition to be more morally good, more trustworthy, more empathic, more judgmental, more committed to their values, more desirable as a social partner, more politically conservative, and more religious compared to the antirealism without cancellation condition, but for the antirealist condition with cancellation to mitigate or completely negate several of these differences. In particular, I expected antirealism with cancellation to prompt participants to perceive Sam as having similar moral character, trustworthiness, and empathy, and to be similarly desirable as a social partner as Sam in the realist condition, but to have comparatively less of an effect on perceptions of how judgmental Sam is, how committed they are to their values, and how conservative and religious they are. In other words, I expected cancellation to buffer antirealists against the perception that they have worse moral character, are untrustworthy, and lack empathy, but not necessarily to influence judgments about other features of their beliefs and attitudes that are more plausibly associated with antirealism as a metaethical position. However, this does *not* mean that I expected cancellation to fully eliminate any negative character judgments, such that participants would perceive Sam's character identically in the realism and cancellation conditions. While this is possible, I predicted *either* that there will be no difference between these conditions, or, at the very least, that the cancellation condition will be intermediate between the realist and no cancellation conditions.

I am less confident about the mechanism that would drive these effects (if they are real). Much of the preceding account relies on the presumption that pragmatics, and in particular conversational implicature, could play a central role in how participants think about and respond to such questions. While these factors could be driving responses, there is considerable room for the results to emerge even in the absence of pragmatics per se. It could be that, at least in the experimental context of this particular study, that nonphilosophers who lack adequate knowledge and training in certain technical distinctions are being asked to assess the meaning of an exchange when they simply lack adequate background knowledge and context to competently do so. If so, results could be driven by task-specific performance error.

Even so, pragmatics could very well be driving this or similar events in everyday interactions. Here's why. Suppose one person asks another if an action is objectively wrong. Ordinarily, when one's normative moral standards are made salient, we might expect it to be important for a person to confirm what those standards actually are. In other words, when asked whether e.g., murder is objectively wrong, we might expect that, whatever else that person says, they should at least state that they think murder is wrong. Failure to do so could conversationally imply that they don't think murder is wrong. Since it would be strange to avoid an implicature that would have serious reputational costs, people may expect someone who is asked about their metaethical stance and their normative stance simultaneously to respond to both, *even if* the question is framed in such a way that the question appears to be only explicitly solicit a response about the respondent's metaethical position. Suppose Alex asks Sam:

Alex: *"Do you think your spouse is one of the fifty most attractive people in the world?"*¹¹³.

We might raise an eyebrow if Sam were to simply respond, "No, I do not." Such a response may be a direct response to a literal interpretation of the question, but it could readily be interpreted to imply

¹¹³ David Moss proposed an example of this kind to illustrate the point. I have adapted it for present purposes.

that Sam doesn't find their spouse to be especially attractive (even though they could think their spouse is the 51st most attractive person in the world, or at least very attractive). We expect Sam to make some remark about how attractive their spouse is, either by stating that, yes, they do think their spouse is in the top fifty (or is the most attractive of all), perhaps in a lighthearted tone that conveys that they are not interpreting the question literally nor responding literally, or they may not directly answer the question at all and merely state that their spouse is attractive, or they may reject the presumption that you can rank people in this way and maintain that beauty is in the eye of the beholder, perhaps affirming that they consider their spouse the most attractive person in the world (though they should be careful in doing so not to insinuate that their appeal to subjectivism isn't a dodge to avoid implying that their partner is unattractive), or, finally, they may say "No," but qualify this by affirming their spouse's attractiveness. As these many possibilities illustrate, the way we respond to such questions often involves a sensitivity to social context, and people are motivated to navigate such questions with a motivation not simply to respond in the most flat-footed and literal way possible, but with a sensitivity to what would be implied by what they say and, often just as importantly, by what they *don't* say.

I suspect a similar implication by omission could drive negative perceptions of antirealist responses, discouraging participants from selecting responses intended to convey antirealism in studies, and discouraging both philosophers and ordinary people from endorsing antirealist positions in academic and everyday contexts. Critically, research on the psychology of metaethics rarely presents participants with the ability to qualify their endorsement of antirealism by allowing them to simultaneously affirm their normative moral stance or their attitude towards the action in question¹¹⁴.

¹¹⁴ Even when participants are asked both about their normative stance and their metaethical stance towards an issue, these questions are asked separately and independently of one another. This means their response to a metaethical question is still straightforward and without qualification, absent contextualization with the rest of their responses. As such, even these studies are still vulnerable to lack of qualification.

As such, selecting an antirealist response option typically leaves participants with no choice but to risk signaling undesirable character traits. If my predictions are born out, this would raise new questions about the validity of existing paradigms.

Methods

Participants. I aimed to recruit 900 participants. 901 participants began the study. However, one participant reported being 3 years old and was excluded from further analysis. This left a total of 900 participants. Participants consisted of 900 adult US residents on Amazon's Mechanical Turk (405 females, 490 males, 4 other, 1 unreported, $M_{\text{age}} = 41.8$, $SD_{\text{age}} = 12.4$, age range = 19-85, 1 unreported age).

Procedure. All participants were randomly assigned to one of three response conditions: (1) realism, (2) antirealism without cancellation, or (3) antirealism with cancellation. In all conditions, participants were asked to consider a conversation between two people named Alex and Sam. In all conditions, Alex asks Sam whether an action is objectively morally right or wrong:

Alex: "Sam, do you believe it is objectively wrong [to torture and kill a baby just for fun]?"

In each condition, Sam's response was displayed in bold immediately following the question. In the realism condition, Sam said "Yes," and affirmed that the action is "objectively wrong." In the antirealism condition Sam said "No," and denied the action is objectively wrong. Finally, in the antirealism with cancellation condition, Sam said "No" and denied the action is objectively wrong but then followed this with additional remarks stating that, although Sam doesn't find the action objectively wrong, Sam does think the action is morally wrong, but doesn't believe in objective morality. Sam then added that they consider the action terrible and are opposed to it. The exact wording for Sam's response in each condition is as follows:

Realism

Sam: "Yes, I believe it is objectively wrong [statement]."

Antirealism without cancellation

Sam: “No, I don’t believe it is objectively wrong [statement].”

Antirealism with cancellation

Sam: “No, I don’t believe it is objectively wrong [statement]. But I do think it is morally wrong. I just don't believe morality is objective. I still find it to be terrible and I am deeply opposed to it.”

In addition to being randomly assigned to one of the three conditions, participants within each condition were randomly assigned to one of three distinct moral issues. Two issues were identical to those used in Study 2:

It is morally wrong for a person to go to a funeral to mock the deceased person in front of their family.

It is morally wrong for a woman who knows she is pregnant to drink alcohol.

Normative entanglement is most likely to occur when the moral transgression in question is perceived as serious. Of the four items, these appeared to be the most serious transgressions. However, I devised an especially serious transgression specifically for this study:

It is morally wrong to torture and kill a baby just for fun.

Next, participants were told to “Think about what Sam said in this scenario,” and were then asked to judge what Sam’s metaethical beliefs are using the same measures used in Study 1 (except for the noncognitivism items) and to assess Sam’s character, attitudes, beliefs, and desirability as a social partner. Finally, participants were asked to provide their age and gender.

Measures. Participants assessed Sam’s metaethical beliefs using the same three-item subscales used in Study 1 for realism, universalism, and relativism (but not noncognitivism). They were also asked to assess Sam’s moral character (1 = Very morally bad, 7 = Very morally good), trustworthiness (1 = Very untrustworthy, 7 = Very trustworthy), empathy (1 = Not empathy at all, 7 = Very empathic), how judgmental Sam is (1 = Not judgmental at all, 7 = Very judgmental), how committed Sam is to their values (1 = Not committed at all, 7 = Very committed), how liberal or conservative Sam is (1 =

Very liberal, 7 = Very conservative), how religious Sam is (1 = Not religious at all, 7 = Very religious), and how desirable Sam is as a social partner (1 = Not at all, 7 = Very good).

Results

Although I employed three distinct moral issues, most of the differences in the dependent variables were not statistically significant. I collapsed all measures across the three moral issues and analyzed results only across the realism, antirealism without cancellation, and antirealism with cancellation conditions.

First, I assessed whether participants attributed metaethical stances in each condition consistent with what the speaker stated. If the manipulations were successful, this would mean that participants would judge the speaker in the realism condition to endorse realism and reject relativism (I also expected participants to attribute universalism to them, given its consistent, close association with realism). Conversely, participants should judge the speaker in the antirealism without cancellation condition to endorse relativism, but to deny realism (and universalism). Critically, participants should also judge the speaker in the antirealism with cancellation condition to endorse relativism and to deny realism (and universalism). Cancellation is only intended to cancel the normative implications of rejecting realism, i.e., that the person in question may be indifferent to or fail to exhibit appropriate opposition to the moral transgression. It is not intended to override the perception that the person in question nevertheless endorses an antirealist metaethical stance, and rejects a realist stance. If it did, this would mean that the manipulation failed, since the cancellation condition failed to disentangle attributions of metaethical and normative stances.

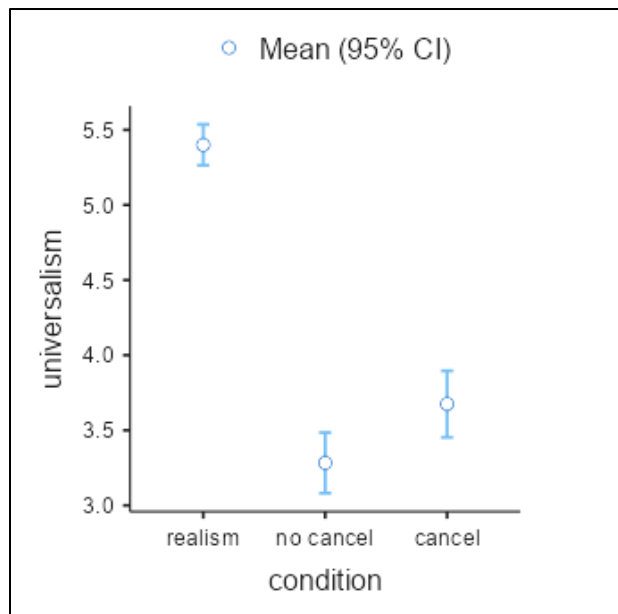
I conducted a one-way ANOVA to compare the effect of condition on realism. Consistent with predictions, there was a significant main effect of condition on attribution of realism $F(2, 576) = 75, p < .001$. As predicted, the mean realism score for the realist condition ($M = 5.38, SD = 1.10$) was significantly different than the mean realism score for both the antirealism without cancellation

condition ($M = 4.39$, $SD = 1.51$), $t(542) = 9.18$, $p < .001$ and the antirealism with cancellation conditions ($M = 4.19$, $SD = 1.59$), $t(523) = 10.70$, $p < .001$. Also consistent with expectations, there was *no* significant difference in realism attributions when comparing the antirealist conditions, $t(590) = 1.59$, $p = .251$.

Next I conducted a one-way ANOVA to compare the effect of condition on universalism. Consistent with predictions, there was a significant main effect of condition on attribution of universalism $F(2, 565) = 182$, $p < .001$. As predicted, the mean universalism score for the universalism condition ($M = 5.40$, $SD = 1.21$) was significantly different than the mean universalism score for both the antirealism without cancellation condition ($M = 3.28$, $SD = 1.77$), $t(524) = 17.1$, $p < .001$ and the antirealism with cancellation conditions ($M = 3.67$, $SD = 1.94$), $t(492) = 13.06$, $p < .001$. Contrary to predictions, there was also a significant difference between antirealist conditions $t(586) = -2.57$, $p = .028$. However, consistent with expectations, the difference in means was fairly small. For the no cancellation condition, the 95% CI centered on the mean of 3.28 was [3.08, 3.48] while for the cancellation condition the 95% CI was [3.45, 3.90]. The upper bound of the no cancellation condition just barely passes the lower bound of the cancellation condition. Thus, while the difference between conditions may have been statistically significant, and may indicate that participants do differ in the degree to which they attribute universalism to the antirealist in the cancellation and no cancellation conditions, this difference is likely *much* smaller than the difference between these conditions and the realist condition, and still suggests that participants perceive the antirealist in both conditions far more similarly to one another than to the realist. This is apparent by examining **Figure 6.4**. Examination of the differences between the conditions merely suggests cancellation was not *perfect*, and that's not something strictly required (or even expected) of my hypothesis.

Figure 6.4

Study 4: Comparison of means (with 95% CIs) for universalism



For the final test to assess proper metaethical attribution I conducted a one-way ANOVA to compare the effect of condition on relativism. Consistent with predictions, there was a significant main effect of condition on attribution of relativism $F(2, 597) = 85.1, p < .001$. As predicted, the mean relativism score for the realist condition ($M = 3.24, SD = 1.45$) was significantly different than the mean relativism score for both the antirealism without cancellation condition ($M = 4.53, SD = 1.47$), $t(601) = -10.90, p < .001$ and the antirealism with cancellation conditions ($M = 4.65, SD = 1.54$), $t(595) = -11.53, p < .001$. Also consistent with expectations, there was *no* significant difference in relativism attributions when comparing the antirealist conditions, $t(590) = -0.96, p = .604$.

These results demonstrate a consistent tendency for participants to attribute metaethical stances across conditions as expected. Next, I assessed how participants judged Sam's character and non-metaethical beliefs by condition. First I conducted a one-way ANOVA to compare the effect of condition on judgments of morally good or bad Sam is. Consistent with predictions, there was a significant main effect of condition on judgment of moral character $F(2, 582) = 236, p < .001$. As

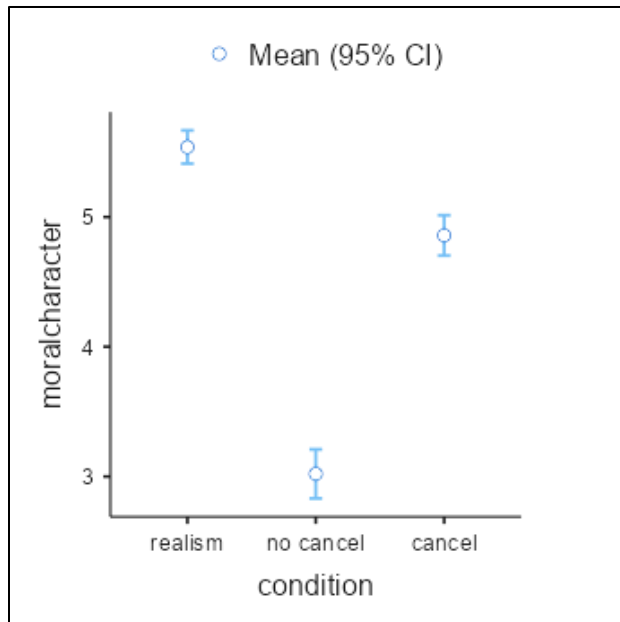
predicted, the mean score for judgment of moral character was significantly different in the realist condition ($M = 5.54$, $SD = 1.14$) than the mean score for judgment of moral character in the antirealism without cancellation condition ($M = 3.02$, $SD = 1.66$), $t(524) = 21.7$, $p < .001$. Although I made no decisive prediction either way, the mean moral character score was also significantly different than the antirealism with cancellation condition ($M = 4.86$, $SD = 1.36$), $t(575) = 6.66$, $p < .001$. Also consistent with expectations, there was a significant difference in mean judgments of moral character when comparing the cancellation and no cancellation conditions, $t(571) = -14.79$, $p < .001$. Notably, all of these significant differences were in the expected direction: as expected, participants judged the realist to be the most morally good, followed by the antirealist who canceled their rejection of moral realism, and then finally the antirealist who did not cancel their rejection of moral realism. While I was uncertain whether cancellation would *completely* cancel comparatively worse character judgments, it came very close. As can be seen in **Figure 6.5**, cancellation appears to have substantially mitigated negative character evaluation, even if it did not completely eliminate it.

Judgments about trustworthiness exhibited a similar pattern. Once again I conducted a one-way ANOVA to assess the effect of condition on judgments of trustworthiness. Consistent with predictions, there was a significant main effect of condition on judgment of trustworthiness $F(2, 575) = 151$, $p < .001$. As predicted, the mean score for trustworthiness was significantly different in the realist condition ($M = 5.43$, $SD = 1.16$) than the mean score for judgment of moral character in the no cancellation condition ($M = 3.32$, $SD = 1.77$), $t(510) = 17.4$, $p < .001$. Although I made no decisive prediction either way, the mean trustworthiness score was also significantly different than the cancellation condition ($M = 4.86$, $SD = 1.51$), $t(553) = 5.18$, $p < .001$. Also consistent with expectations, there was a significant difference in judgments of trustworthiness between the cancellation and no cancellation conditions $t(579) = -11.47$, $p < .001$. Once again, all differences were in the expected direction, with participants judging Sam to be most trustworthy in the realist condition,

somewhat less trustworthy in the cancellation condition, and much less trustworthy in the no cancellation condition.

Figure 6.5

Study 4: Comparison of means (with 95% CIs) for moral character judgments



The same pattern also held for perceived empathy. Consistent with predictions, there was a significant main effect of condition on perception of empathy $F(2, 588) = 173, p < .001$. As predicted, the mean empathy score was significantly different in the realist condition ($M = 5.12, SD = 1.33$) than the no cancellation condition ($M = 2.78, SD = 1.76$), $t(552) = 18.4, p < .001$. Once again, although I made no decisive prediction, the mean empathy score was also significantly different than in the cancellation condition ($M = 4.60, SD = 1.50$), $t(586) = 4.49, p < .001$. Also consistent with expectations, there was a significant difference in judgments of empathy between the cancellation and no cancellation conditions $t(579) = -13.53, p < .001$. As before, all differences were in the expected direction, with participants judging Sam to be most empathic in the realist condition, somewhat less empathic in the cancellation condition, and much less empathic in the no cancellation condition.

Next, I assessed the degree to which participants perceived Sam to be judgmental across conditions. Consistent with predictions, there was a significant main effect of condition on perceived judgmentalness $F(2, 588) = 23.2, p < .001$. As predicted, the mean judgmentalness score was significantly different in the realist condition ($M = 4.42, SD = 1.37$) than the no cancellation condition ($M = 3.66, SD = 1.78$), $t(558) = 5.89, p < .001$, and the cancellation condition ($M = 3.76, SD = 1.61$), $t(579) = 5.46, p < .001$. I expected cancellation to *not* fully eliminate differences between the realism and cancellation conditions. This is because I anticipated that stating that an action is objectively wrong would plausibly be seen as judgmental, but stating that it isn't objectively wrong, and that in fact one does not think *anything* is objectively wrong, but that one still opposes it would be perceived as *less* judgmental. This was born out in the results. However, it was less clear whether there would be any significant difference between the cancellation condition and the no cancellation condition. As such, I expected the cancellation condition to at least be intermediate between the realist and no cancellation conditions. Thus, my expectations were consistent both with the possibility that the cancellation and no cancellation difference were significantly different or not significantly different. So long as both appeared to have lower mean scores than the realist condition, whether they differed from one another is incidental to my main hypotheses. It turned out that there was no significant difference between the cancellation and no cancellation conditions $t(587) = -0.69, p = .770$. This suggests that cancellation (or at least the precise cancellation wording used in this study, for the particular moral issues it was used for) has little or no impact on how judgmental a person is perceived when they qualify an expression of antirealism with additional remarks about their opposition to the issue in question.

I next performed a one-way ANOVA to assess perceived commitment to their moral values across conditions. Consistent with expectations, there was a significant main effect of condition on perceived commitment to moral values $F(2, 578) = 68.7, p < .001$. As predicted, the mean perceived

commitment score was significantly different in the realist condition ($M = 5.67$, $SD = 1.11$) than the no cancellation condition ($M = 4.33$, $SD = 1.67$), $t(516) = 11.6$, $p < .001$, and the cancellation condition ($M = 5.10$, $SD = 1.40$), $t(562) = 5.56$, $p < .001$. Once again, I expected the cancellation condition to be at least intermediate between the realist and no cancellation conditions for perceived commitment (i.e., significantly less than realism, and significantly more than no cancellation), but I did not predict whether it would fail to significantly differ from the no cancellation condition. Consistent with expectations, the cancellation condition did significantly differ from the no cancellation condition in the expected direction, $t(576) = -0.606$, $p < .001$. Once again, the mean score for the cancellation condition was intermediate between the realist and no cancellation condition, in line with expectations.

Next, I assessed whether participants judged Sam to be a good person to have as a social partner, such as a coworker or a friend. Consistent with predictions, there was a significant main effect of condition on partner preference $F(2, 577) = 148$, $p < .001$. As predicted, partner preference was significantly different in the realist condition ($M = 5.14$, $SD = 1.20$) than both the no cancellation condition ($M = 3.01$, $SD = 1.79$), $t(516) = 17.2$, $p < .001$, and the cancellation condition ($M = 4.56$, $SD = 1.56$), $t(553) = 5.14$, $p < .001$. I expected the cancellation condition to be at least intermediate between the realist and no cancellation conditions for partner preference, such that Sam would have the highest mean partner preference score in the realist condition, and the lowest in the no cancellation condition, with the cancellation condition either being intermediate or not significantly different than the realist condition. Consistent with this expectation, partner preference was significantly different in the cancellation and no cancellation conditions in the expected direction, $t(582) = -11.29$, $p < .001$. Once again, the mean score for the cancellation condition was intermediate between the realist and no cancellation condition.

After this, I assessed differences in judgments about whether Sam was more politically liberal or conservative across conditions. Consistent with expectations, there was a significant main effect of

condition on perceived political orientation $F(2, 587) = 22.3, p < .001$. As predicted, Sam's perceived political orientation was significantly different in the realist condition ($M = 4.22, SD = 1.26$) than both the no cancellation condition ($M = 3.47, SD = 1.61$), $t(563) = 6.29, p < .001$, and the cancellation condition ($M = 3.70, SD = 1.57$), $t(566) = 4.47, p < .001$. Consistent with expectations, there was no significant difference in perceived political orientation for the cancellation and no cancellation conditions $t(592) = -1.71, p = .203$. Overall, people perceived Sam in the realist condition to be more inclined towards political conservatism than in either antirealist condition, while the antirealist conditions did not significantly differ from one another. This suggests that cancellation did not substantially change how participants perceived Sam's political orientation and that, as expected, those who endorse moral realism (at least for the moral issues used in this study) are perceived as more disposed towards political conservatism.

Lastly, I assessed differences in perceived religiosity across conditions. Consistent with expectations, there was a significant main effect of condition on perceived religiosity $F(2, 585) = 107, p < .001$. As predicted, perceived religiosity was significantly different in the realist condition ($M = 4.31, SD = 1.28$) than both the no cancellation condition ($M = 2.56, SD = 1.65$), $t(559) = 14.50, p < .001$, and the cancellation condition ($M = 3.47, SD = 1.63$), $t(559) = 7.04, p < .001$. Contrary to expectations, there was a significant difference in perceived religiosity between the cancellation and no cancellation conditions $t(592) = -6.74, p < .001$. However, consistent with expectations, perceived religiosity for the cancellation condition was intermediate between the realism and no cancellation conditions. Overall, I didn't expect cancellation to substantially increase perceived religiosity, though it appears that it did. Nevertheless, perceived religiosity fell midway between the realism and no cancellation conditions, suggesting that while the additional cancellation remarks may have increased perceived religiosity, they did not so as much as expressing realism does. What would have been genuinely shocking and contrary to my expectations would be if the cancellation condition was not

significantly different from the realism condition, or was significantly larger. This did not occur, so while the overall findings are not precisely what was anticipated, they are still in line with the generally expected pattern, with the cancellation being intermediate between realism and no cancellation.

Discussion

Overall, the results of this study were highly consistent with the proposed phenomenon of *normative entanglement*. When asked whether a serious moral transgression is objectively wrong, people who express a realist stance, compared to those who bluntly express antirealism without qualification, are perceived as more morally good, more trustworthy, more empathic, more committed to their values, more judgmental, better social partners, and as more politically conservative and religious. Yet precisely those effects that were expected to be mitigated or eliminated when the antirealist attempts to cancel their rejection of moral realism with an expression of their opposition to the transgression were, attenuated or eliminated, including moral character, trustworthiness, and empathy, as well as quality as a social partner, yet a person who expressed antirealism with cancellation was still perceived as endorsing relativism and rejecting realism and universalism, and was still regarded as less committed to their moral values than the realist. These findings suggest that cancellation can function to attenuate the negative effects of endorsing antirealism without qualification, while leaving other judgments about the person intact.

Discovering *why* cancellation has an ameliorative effect on perceptions of the moral character of antirealists is a task for future research. While I suspect this effect may be driven by conversational implicature, it could also be driven by simple performance errors or unintended interpretations of the stimuli. If so, it could be an artifact of experimental design. However, insofar as similar remarks are made in ordinary circumstances, such errors may likewise occur in the real world. As such, even this latter possibility doesn't rule out a potentially important discovery.

Finally, these results should be qualified by another observation. I have been careful to frame judgments about the antirealist who does not cancel their rejection of realism as inferences that are not *logically entailed* by an endorsement of antirealism. This isn't merely because they aren't entailed, but could be *mistakenly* perceived to be implied by expressions of antirealism. Rather, to reject moral realism in response to a question that involves normative entanglement technically provides *some* evidence, all else being equal, that the person in question *really does* endorse both the metaethical question (whether the act in question is *objectively* wrong) *and* the normative claim (that the act in question is *wrong* simpliciter). This is one of the most insidious and unfortunate aspects of normative entanglement. Normative entanglement amounts to a kind of complex question, where a person is explicitly asked a question about their metaethical stance towards a moral act, but, technically, they are *also* asked about their *normative* stance...*conditional* on whether they are a realist about the moral issue in question. If asked whether torturing babies just for fun is objectively wrong, if Sam says "Yes," Sam is in fact expressing *both* a metaethical position *and* a normative moral position:

(1) Torturing babies just for fun is morally wrong (normative)

(2) This normative fact is objectively true (metaethical)

Yet if Sam says "No," this is technically only a denial of (2), not (1). There is, therefore, an asymmetry between saying "Yes" and "No": saying yes *explicitly* addresses both questions, while saying "No" only explicitly addresses one of the two questions. This asymmetry creates a kind of pragmatic vacuum that could be filled by implicatures about the speaker's normative stance. Yet setting aside what may be conversationally implied by a "No," observe that if someone says "No," this provides, in the absence of cancellation, some genuine evidence about their normative moral stance: relative to a person who says yes, and thereby explicitly expresses a normative moral stance (namely, that the action is wrong), they are more likely not to endorse a contrary moral stance (at the very least, that the action is *not* wrong). For comparison, imagine Alex asks Sam:

Do you like pepperoni pizza?

If Sam says “Yes,” this indicates both that Sam likes pepperoni pizza *and that Sam likes pizza*. But suppose Sam says “No.” While this would indicate that Sam does not like pepperoni pizza, it does not follow that Sam dislikes pizza *without* pepperoni. Consider the epistemic asymmetry between the two conditions: We know that if Sam says “Yes,” that Sam likes Pizza, but if Sam says “No,” we still don’t know whether Sam likes or dislikes pizza (without pepperoni). Yet we do know one thing: if Sam says “No,” Sam is *less likely* to like pizza than if Sam says “Yes.” If we think like a good Bayesian, to say “No” to this question should cause us to reduce our confidence that Sam likes pizza. A similar epistemic asymmetry is present in cases of normative entanglement. To oversimplify for demonstrative purposes, there are four positions a person could have with respect to the metaethical status, and the normative status, of an act:

	Morally wrong	Not morally wrong
Realism	The action is objectively wrong	The action is objectively not wrong
Antirealism	The action is wrong, but is not objectively wrong	The action is not wrong and is not objectively wrong

A person who responds with a “Yes” puts themselves in the top left category, while a person who says “No” could endorse *any* of the other three positions. However, only the two positions in the right column indicate that one does not hold the normative moral stance that the action is wrong. Thus, even if a person endorses the bottom left position, saying “No,” does not clearly indicate this in the way saying “Yes” indicates the top left position. Since a “Yes” allows us to exclude endorsement of either position on the right, but saying “No,” doesn’t, to say “No,” provides at least *some* evidence, relative to the counterfactual in which a person says “Yes,” that the person does *not* hold the normative

moral stance that the action in question is wrong. Whether or not, and to what extent ordinary people are responsive to such evidence is a matter of future empirical inquiry.

6.5 General discussion

Given the somewhat disparate nature of the three main studies reported here, most of the relevant discussion appears in the discussion sections for each of those studies. Nevertheless, there are a few central findings worth emphasizing. First, Study 1 demonstrated that judgments of realism, universalism, and relativism are not the same when participants are judging the metaethical stances of other people compared to when they report their own metaethical stance. There is at best only a weak correlation between third and first person judgments, suggesting that people's judgments about other people's metaethical standards are not simply a projection of their own metaethical stances. The disparity between first and third person judgments is an important finding, since it vindicates the importance of comparing first and third person judgments, and raises questions about relying exclusively on first person judgments.

While I favor normative confluences as a proposed mechanism behind at least some of this disparity, I did not directly test for the effects of normative confluences. However, this wasn't the goal of the present set of studies. Instead, the goal was to devise a set of measures that could first point to a disparity in need of explanation. Having established such a disparity, future efforts will be directed at explaining its cause. The pattern of results I found is consistent with normative conflation accounting for at least some of the disparity between third and first person judgments. Study 1 revealed that participants judged other people to exhibit greater support for realism and universalism, and less support for relativism, than themselves. This *could be* driven, in part, by normative confluences. Participants may be disinclined to endorse realism and universalism, and more inclined to endorse relativism, because doing so could signal an undesirable normative moral stance, or undesirable character traits, such as rigidity, intolerance, and close-mindedness.

However, I suspect an additional factor is in play. The participants in these studies were less politically conservative, and much less religious, than the population of the United States as a whole. Since each of these factors is linked with greater support for realism and rejection of relativism, it seems possible that the participants that were sampled were not representative of the population they were drawn from (i.e., adults in the United States). If so, participants may have both accurately reported their own metaethical stances *and* accurately judged a typical person in their society to be more inclined towards realism and universalism, and less inclined towards relativism. After all, the typical person is more politically conservative, and much more religious, than participants themselves. In imagining what other people think, participants may be responsive to these differences, and correctly recognize that their own metaethical stance may differ from the general population.

One way these findings corroborate previous research is that the same pattern of realism and relativism emerged across moral and nonmoral domains as prior research using first person measures (e.g., Goodwin & Darley, 2008; Wright, Grandjean, & McWhite, 2013). Participants judged others to be most inclined to support realism and reject relativism for facts, followed by morals, social conventions, and matters of taste. While this corroborates the *order* of judgments about realism and relativism with respect to these domains, it remains an open question whether or not, and to what extent, any given population is more inclined towards realism or antirealism about morality in particular. The factual and taste domains serve largely to anchor the assessment of different domains, given that most studies find the former to exhibit high levels of realism and the latter to exhibit low levels of realism. The findings reported here are consistent with that pattern. Social conventions stand out as an anomaly, however. Previous research has likewise failed to pin down how people think about the metanormative status of social conventions. Social conventions have historically straddled a middle ground between morality and taste.

The current findings are consistent with this pattern, yet it was surprising that I also found that people judged others in their society to be more inclined towards realism and universalism, and less inclined towards relativism for social conventions than themselves. I am unsure why this is the case. On the one hand, the unsatisfyingly equivocal results of previous research, coupled with these findings, point to the potential for a rich and interesting possibility for research on metanormative psychology outside the moral domain. On the other hand, I suspect these findings may be due to systemic patterns of unintended interpretations when questions about metanormativity are directed at matters of social convention.

Finally, Study 4 points to a potentially significant phenomenon: normative entanglement. Current findings suggest that when a person responds to a question about whether a serious moral transgression is “objectively wrong,” by stating that it is, people have a fairly positive appraisal of such a person, especially compared to a person who denies that the action in question is objectively wrong. Yet current findings indicate that, at least among the participants that were sampled, when someone who initially responded by simply denying that the act in question was objectively wrong, but then qualified that remark by also expressing their opposition and negative attitude towards the act in question, most of these negative character judgments were nearly eliminated, while participants continued to regard the person who made such a remark as a moral antirealist. These findings suggest that people who bluntly express moral antirealism may be perceived extremely negatively, at least in contexts where questions about their metaethical standards are entangled with normative questions about serious moral transgressions.

When asked about their own metaethical stances, participants are rarely given the opportunity to qualify those remarks in this way. As a result, participants may be disinclined to endorse antirealist responses when they judge that doing so could carry reputational costs or when they judge those responses to also carry implications about their normative moral stance, and their attitudes and

emotional responsiveness to a given moral transgression. If so, studies that rely on concrete measures may underestimate support for antirealism, or may provide misleading evidence of “pluralism” (i.e., participants endorsing realism for some moral issues but not others) that is actually driven in part by variation in the degree to which participants judge an “antirealist” response to carry undesirable implications. More generally, while the observed effect may be limited to the stimuli and sample reported here, these findings point to the more general possibility that pragmatic factors may influence how participants respond to stimuli, potentially threatening the validity of other measures and providing the impetus for new research on the role pragmatics may play in how participants interpret psychological stimuli.

However, a more intriguing possibility, given these findings, is that the consequences of entangling metaethical and normative/social implications may also emerge outside the lab. Reasoning has traditionally been thought of as a tool for acquiring accurate beliefs about what the world is like. Yet Mercier and Sperber (2011; 2017) marshal an impressive array of arguments and evidence that reasoning may have evolved to construct and evaluate arguments and, more generally, to *persuade*. Indeed, it is also possible that, as Haidt (2001) argues, many moral judgments are not generated so much by reasoning as such, but are instead the output of automatic and intuitive processes, with reasoning largely to provide post hoc justification for one’s moral conclusions.

If these findings are on the right track, they may explain why normative entanglement occasionally emerges in academic contexts and disputes among those without formal training in philosophy: such examples may illustrate a more *general* tendency for people to exploit the rhetorical impact of normative entanglement in interactions outside academic contexts. For instance, people may exploit realist-sounding language in a moral dispute in an attempt to smear a rival by implying that their rival fails to adequately oppose atrocities. If so, there is little reason to think such rhetoric would be limited specifically to normative entanglement. Rather, it could be that people strategically

express realist-sounding and anti-realist sounding language, or attribute realism or antirealism to others, in ways that serve their social and argumentative goals. Such rhetorical ploys could be highly effective even if the speaker has little or no understanding of the relevant metaethical positions as they are understood by academic philosophers. In short, to the extent that metaethical terms and concepts, or at least language that superficially resembles their use within academic philosophy, appears in everyday discourse, it may appear there for reasons unrelated to the genuine adoption of metaethical stances among nonphilosophers, or the robust comprehension of distinct metaethical positions.

If so, this could partially account for the poor validity of existing research on folk metaethics. It may even go some way in explaining why philosophers and psychologists have been slow to recognize that ordinary people have no determinate metaethical positions. Terms like “objective” and “relative” have colloquial analogs to their philosophical counterparts, unlike many of the technical terms used in e.g., physics or mathematics. Scholars studying metaethics may have developed the misperception that ordinary people have substantive, determinate metaethical stances due in part to the fact that ordinary people use similar terms to those used among academics. Such usage may not even have evolved independently; it’s possible notions devised in the academy leak into public discourse. If so, there may be a degree of terminological and conceptual cross-pollination between academic and everyday contexts that bolsters the illusion that ordinary people are “realists” and “antirealists.” Instead, it may be that ordinary people have a superficial or muddled understanding of the meaning of the relevant terms and concepts, and are employing them for largely argumentative purposes, *not* to express sober stances towards abstract philosophical issues about the metaphysics of morality. Of course, these considerations are speculative, and gathering support for them is a task for future research.

There are likely a number of limitations with these studies that have gone unnoticed, or that others will point out in due time. One significant limitation with these studies is how little they tell us. Thus, in describing future directions I am simultaneously addressing the study's central shortcomings.

First, future versions of the third person paradigm should develop more robust measures by expanding on the number of items that appear in each subscale, and make a more concerted effort to assess their psychometric quality by e.g., analyzing their factor structure and recruiting experts familiar with the relevant concepts to judge their validity. Assessing how ordinary people interpret these questions using the methods introduced in earlier chapters would likewise serve a valuable role in refining our measures. It would also be worthwhile to make a second attempt at devising valid measures of noncognitivism, though this may ultimately prove infeasible. Even if efforts to devise adequate measures for noncognitivism, future research could expand on the range of metanormative concepts by devising new subscales. This could include efforts to distinguish cultural relativism from individual subjectivism, naturalist from non-naturalist conceptions of moral realism, and, though it may prove difficult, less well-studied and potentially more complicated metaethical positions, such as error theory, constructivism, and ideal observer theory. Such efforts will be challenging, if not futile, at least using ordinary survey items. Such concepts may require more sophisticated paradigms. Researchers could also directly assess whether ordinary people explicitly endorse metaethical pluralism or indeterminacy, by directly or indirectly asking them about each.

Future research could also refine or expand on the third-personal aspect of the paradigm. One shortcoming with the current set of studies is that they provide little information about *who* participants had in mind when thinking about a typical person or judging the metaethical beliefs of a person with no information other than that they made a particular moral claim. Researchers could incorporate additional measures that assess what kind of person participants have in mind by asking them to describe or make additional judgments about the imagined person. Future studies could also provide

richer and more detailed information about the target participants are asked to judge. This could include demographic information, such as age, gender, or political leanings, or via richer statements or vignettes that provide more context.

There were also only a limited number of statement types: I only asked people to judge scenarios in which a person stated that an action was “right” or “wrong.” Moral vocabulary is far more expansive than this. Future efforts could assess a wider range of moral statements, including statements that an action is “obligatory,” “prohibited,” “permissible,” “required,” “acceptable,” and so on. Nuance could be added to such claims by including qualifiers, e.g., that an action is *always* prohibited, or *sometimes* permissible. Such qualifiers may mislead participants, but could also highlight ways in which participants interpret items in unintended ways, e.g., people may judge that a person who says that an action is *sometimes permissible* is more likely to endorse moral relativism, even though such a claim is logically consistent with moral realism and universalism.

It may also be of interest to ask participants to assess what metaethical standards people *outside* their culture endorse, or even to ask participants to speculate about the metaethical beliefs of nonhumans, e.g., extraterrestrials or artificial intelligences. While people would not, of course, have any way of knowing the answer to these questions, such findings could reveal features of how people think about the nature of morality that might otherwise be difficult to assess.

In addition to alternating features of the target of judgment in third person paradigms, it would also be helpful to recruit participants from more diverse backgrounds. The present studies only recruited MTurk participants, but results showed that, consistent with research on the demographic characteristics of workers on MTurk, these participants were less politically conservative and much less religious than the general population of the United States. Since these variables play a significant role in how people think about metaethics, the present studies do not generalize well to people in the United States. It would also be helpful to focus on members of particular religions (or even atheists

or agnostics), or to assess demographically distinct subpopulations, e.g., members of distinct religious communities such as Mormons, though it would be especially interesting to survey non-WEIRD populations (Henrich et al., 2010).

Future studies could also incorporate additional nonmoral domains, such as epistemic and prudential norms. It could also turn out that existing domains have not been conceptualized or operationalized adequately. It may be, for instance, that people do not treat aesthetic judgments about art and beauty the same way as food preferences. Some normative domains may also exhibit intradomain variation. There may be systematic differences in how people judge moral issues that reflect Haidt's moral foundations. Davis (2021) has already addressed this topic using an adapted version of the disagreement paradigm, though given its methodological shortcomings, a look with a new paradigm would be worthwhile. Other domains may likewise exhibit subdomains worthy of targeted inquiry, e.g., people may think differently about scientific claims versus other sorts of "factual" claims, such as math or religion.

It would also be helpful to expand on the number of concrete items in the moral domain. Study 2 only employed 4, yet future research could incorporate a broader and more representative sampling of distinct moral issues. In addition, there were *no* concrete measures for nonmoral domains. These domains are typically neglected relative to the moral domain, yet this could be addressed by providing sufficient attention to testing a well-designed set of concrete issues for these domains as well. It would also be worthwhile to explore differences in judgments about nonmoral domains using a within-subjects design.

Much of the present research focuses on identifying whether or not there is a disparity between third and prison judgments, and speculating about what might cause such a disparity. Yet the present findings are incapable of directly identifying what mechanisms could be driving this difference, or explaining why participants endorse various metaethical positions in general. While I will abstain from

sketching any particular research programs for identifying the factors that explain people's responses, I believe exploring these factors is one of the most critical pieces of the broader puzzle in the psychology of metaethics.

One way to contribute to this task, and that could be of considerable independent interest, would be to explore the relationship between responses to the third person paradigm and a variety of psychological variables that could be related, e.g., Big Five personality traits or other philosophical beliefs, such as belief in free will.

Finally, there are a number of avenues for future research that are distinct to Study 4. However, I should first acknowledge one limitation with this study. My main test of whether normative entanglement occurred consisted of a variety of measures assessing the character of Sam, the target who made the statement. While this is one element of normative entanglement, and the most important one from a social psychological perspective, I did not directly assess judgments about Sam's normative moral position. That is, I did not ask whether Sam thinks the action in question is morally right or wrong. While it may serve as little more than an attention check for the realism and cancellation conditions, since Sam explicitly states that they think the action is wrong, it is important to confirm that the proportion of participants who judge that Sam does *not* think the action is morally wrong is lower in the no cancellation condition compared to the other two conditions. I intend to rectify this lacuna in subsequent research. In addition, I plan to include a variety of additional measures designed to bolster the case for normative entanglement. In particular, I plan to ask participants whether Sam would be more or less likely to engage in the action in question, how Sam would react to someone committing the action, whether Sam would favor laws against the action, and whether Sam would want to punish people who commit the action. In short, I want to evaluate whether participants go beyond merely judging Sam's character negatively in the no cancellation condition, but

whether they also differ in their judgments about Sam's normative moral beliefs and Sam's attitudes and behavioral disposition.

Future research could also assess whether phenomena similar to normative entanglement occur in everyday circumstances. In particular, such studies could assess how people judge arguments that employ various types of metaethical language, in both direct ways and in ways designed to serve rhetorical purposes (e.g., implying that someone who does not endorse moral realism is evil). Do they find such arguments persuasive? Would they employ such arguments themselves? Do they employ these arguments when given the opportunity? Can we identify instances of their use among e.g., politicians, legal contexts, or academic work? Once we reorient our perspective on the functions moral language, judgment, and norms play in everyday life, whole new avenues of research open up. For instance, Sperber and Baumard (2012) maintain that one of the central functions of moral behavior is to develop a positive reputation in order to facilitate cooperation with others. They acknowledge that a genuine motivation to do what is morally good (or at least considered morally good in one's community) may be one of the surest ways to maintain a positive reputation. However, they argue that if, from an adaptive perspective, the target of such behavior is our reputation (rather than e.g., "doing the right thing"), it may be too costly to leave the fortunes of our reputation to our conscience alone. They propose that we may have an evolved predisposition to be *directly* concerned with managing reputation. If so, much of our motivation in moral disputes may be instrumental, with an eye towards navigating social hierarchies, and avoiding loss in power, status, and cooperative opportunities.

Such proposals synergize well with the argumentative theory of reasoning, and with the potential argumentative role of metaethical terms and concepts. In short, the claim that metaethical discourse may play a largely argumentative and rhetorical role has not emerged within a theoretical vacuum, but dovetails with emerging theories about the social function of moral reasoning and

behavior. Future studies can capitalize on this by directly assessing what role (if any) metaethical terms and concepts play in argumentative contexts and other contexts where people's reputation is at stake.

6.6 Conclusion

Previous chapters offered a bleak picture of the state of research on folk metaethics. This chapter offers a more optimistic outlook. I suspect the most constructive changes will emerge from a reorientation in our general approach to studying how ordinary people think about traditionally philosophical topics like moral realism and antirealism (and their nonmoral analogs). Rather than identifying the distinctions of interest to academic philosophers, directly operationalizing these into a set of measures, and then throwing them at participants with little or no context, then expecting people's responses to be informative, those studying the psychology of metaethics should appreciate that morality is a deeply social and practical enterprise. Real moral judgment and behavior doesn't occur in the imaginary worlds dreamed up by philosophers, where space pirates threaten annihilation and people must contemplate hurling their grandma into a volcano to save a dozen doctors. Real moral judgments involve messy situations that are contextually rich, and often feature people we know and care about. Our reputations are often on the line, and we stand to gain or lose a great deal if we say or do the wrong thing. Stripped of this context, much existing research on moral psychology is ecologically impoverished in ways researchers are only beginning to recognize and detect (see e.g., Navarro-Plaza et al., 2020).

The third person paradigm goes some way in slightly ameliorating these concerns by focusing judgments on others rather than asking for self-reports. Yet the real potential of the third person paradigm is that it can be expanded in the many ways highlighted here: we can add richer and more detailed context than is typical of much existing research in moral psychology. In addition, by driving a wedge between third and first person judgments, the disagreement paradigm could serve as the start of a research project designed to assess the role that normative confluences and other methodological

concerns may play in undermining the validity of previous research. Finally, Study 4 assesses the possibility that implications about a person's character may influence how participants respond to metaethical stimuli.

Once we internalize the fact that participants may be unable to simply turn off concerns about how others might perceive them based on their response to stimuli, we can begin to appreciate that participants may not be responding to metaethics stimuli like a philosopher, trained (with a questionable degree of success) to set aside their emotions, concerns about their reputation, and fears about the implications of facing the truth, to render a sober, hard-nosed judgment about the fundamental nature of reality. That just isn't how *ordinary* people think. And yet research on the psychology of metaethics has treated ordinary people as if they were philosophical automata capable of readily distinguishing abstract and unfamiliar philosophical concerns about, of all things, the metaphysical status of truth claims from far more familiar, mundane questions about whether they think an action is morally right or wrong. And we expect them to do that while holding in abeyance millions of years of evolution and decades of cultural conditioning that impel them to prioritize their reputation when considering moral issues.

This is, to put it mildly, absurd. Progress in the psychology of metaethics, and in moral psychology more generally, will be best served by recognizing that morality, as a psychological phenomenon, has not been accurately represented by focusing so much on the rarefied ratiocinations of an insular coterie of academics whose thought processes have been transmogrified by their shared induction into a canon of idiosyncratic thinkers. No, morality, *actual* morality, occurs in the real world, with its rich social context, and with all the attendant social motivations, reputational stakes, fear of shame and punishment, identifying cooperative opportunities, signaling of our commitment to our ingroups, distancing ourselves from outgroups, and coming out on top in disputes by winning arguments, that occur in that world.

CHAPTER 7:

Conclusion

Many readers will recognize that the title of this dissertation is a reference to Schrödinger's cat. Schrödinger's cat is a thought experiment devised to illustrate the bizarre and seemingly paradoxical nature of quantum mechanics (Schrödinger; 1935; Villars, 1986). Erwin Schrödinger imagined a scenario in which a macroscopic event—the fate of a cat—could be entangled with the outcome of a seemingly indeterminate quantum event. Wineland (2013) provides a summary of the scenario:

Erwin Schrödinger [...] realized that, in principle, quantum mechanics should apply to a macroscopic system in a more complex way, which could then lead to bizarre consequences. In his specific example, the system is composed of a single radioactive particle and a cat placed together with a mechanism such that if the particle decays, poison is released, which kills the cat. Quantum mechanically we represent the quantum states of the radioactive particle as undecayed = $|\uparrow\rangle$ or decayed = $|\downarrow\rangle$, and live and dead states of the cat as $|L\rangle$ and $|D\rangle$. If the system is initialized in the state represented by the wave function $|\uparrow\rangle|L\rangle$, then after a duration equal to the half life of the particle, quantum mechanics says the system evolves to a superposition state where the cat is alive and dead simultaneously, expressed by the superposition wave function:

$$\Psi = \frac{1}{\sqrt{2}}[|\uparrow\rangle|L\rangle + |\downarrow\rangle|D\rangle].$$

Schrödinger dubbed this an entangled state because the state of the particle is correlated with the state of the cat. That is, upon measurement, if the particle is observed to be undecayed, one can say with certainty that the cat is alive, and visa versa. But before measurement, the particle and cat exist in both states. (p. 1103)

For our purposes, the critical element of this thought experiment is that the cat exists in a state of superposition, simultaneously both dead and alive, and this state of indeterminacy collapses only when the wave function itself does. Prior to the collapse, *there is no determinate fact of the matter* about whether the cat is alive or dead.

I didn't draw on an analogy between this scenario and metaethical indeterminacy just because I found it amusing, or because I really wanted to include a bad pun in the title.¹¹⁵ I drew on it for a different reason: I believe that it's time for philosophers to begin a similar paradigm shift. Physicists have had a century to grapple with the unsettling implications of quantum mechanics, and while they have yet to reach a consensus on how (or even whether) to interpret quantum mechanics, enough time has passed for them to have become comfortable with the idea that the world, as described by fundamental physics, is radically at odds with our ordinary ways of thinking and speaking. Yet contemporary analytic philosophy has yet to make a comparable shift. Philosophers have not adequately grappled with the possibility that the way they think about the world is also radically at odds with our ordinary ways of thinking and speaking.

I suspect that this is due to an entrenched methodological focus on the analysis of terms and concepts. Since the "linguistic turn," analytic philosophy developed a pathological obsession with language, and metaethics was swept up in this endeavor. MacIntyre (1957), one of central figures of 20th century metaethics, put it bluntly: "The central task to which contemporary moral philosophers have addressed themselves is that of listing the distinctive characteristics of moral utterances" (p. 325). Philosophers in the 20th century appeared to believe they could solve fundamental questions in ethics by studying how ordinary people talk, and the inertia behind this approach has continued into the present.

Stranger still, discovering the "distinctive characteristics of moral utterances" looks like an empirical task. Yet philosophers did not get out the clipboards and pens and start surveying people. Instead, they were content to address how ordinary people think and speak from the armchair, on the assumption that their education endowed them with unique insight into the nature of ordinary thought and language. I've never encountered any good reason to think armchair methods are adequate to

¹¹⁵ Though I admit these motivations played some role.

determine how ordinary people speak or think about morality specifically, and the only way to decisively resolve such disputes is with empirical data, e.g., from linguistics, psychology, and anthropology.

I'm not alone in holding such a view. Experimental philosophy emerged in the 21st century as a proposed remedy to exclusively armchair analyses. Rather than waiting for psychologists to pick up the slack, experimental philosophers began conducting their own studies. Social scientists quickly followed suit, conducting their own studies on folk philosophy. Yet regardless of their disciplinary backgrounds, all of these researchers made the same fundamental mistake: rather than addressing the study of folk philosophy with a bottom-up approach, they came equipped with an arsenal of philosophical terms, concepts, and distinctions from a tradition that had never engaged in *any* meaningful empirical research on folk morality.

As a result, researchers studying folk philosophy have taken their cues from contemporary analytic philosophy: they have co-opted the terms, concepts, and distinctions devised by academic philosophers, and have adapted these into a host of scales and paradigms designed to capture ordinary thought, as though these philosophical concepts could be converted wholesale into psychological constructs. This whole enterprise rests on the dubious notion that the categories and distinctions that characterize contemporary analytic philosophy are merely the domesticated forms of philosophical views that could be found in the “wild” in the words, thoughts, and deeds of ordinary people. In short, researchers have simply presumed that their accounts map onto psychological constructs that are present in ordinary thought with virtually no theoretical justification, and no substantive empirical evidence beyond appeals to anecdotes and personal experience.

I have argued for an alternative possibility: that this simply isn't the case. Perhaps philosophical notions of realism and antirealism aren't refined accounts of what ordinary people are saying and doing when they reason that murder is wrong. Instead, these philosophical positions are academic

inventions that don't reflect what people are actually saying and doing at all when they engage in moral judgment, reasoning, and behavior.

The goal of this dissertation has been to argue that this hypothesis, *folk metaethical indeterminacy*, offers a plausible explanation of existing patterns of data in folk metaethics research. In chapters two and three (and the accompanying supplements), I argued that existing methods are not valid, and have thus not provided us with good evidence of metaethical pluralism or uniformism. In chapters four and five, I presented qualitative and quantitative evidence that ordinary people do not interpret questions about metaethics as intended. At the end of chapter five, I showed that even when we have good reason to believe that a given set of concepts do not plausibly figure into ordinary thought and language, ordinary people can still be induced to respond in a nonrandom way. Were researchers so inclined, studies like this could be used as evidence of "folk quantum mechanics." Yet I have provided two better explanations for these results. In some cases, these findings may be due to *spontaneous theorizing*, i.e., participants developing a philosophical stance in the course of the study. In other situations, results may fail to reflect what participants actually think, since they cannot express their views due studies employing forced choice designs. Either way, we have no good reason to think ordinary people *outside* of the study are "implicit Copenhagians" or "implicit Many Worlders." The observed patterns of data are better understood as empirical artifacts.

I propose that the same is true of folk metaethics. First, we have the unusual pattern of data obtained when such research is conducted. Our best explanation for these results is not that ordinary people are metaethical pluralists, or that they are uniformists making systematic errors. The best explanation is that most participants are not interpreting stimuli in the way researchers intend.

Our next question is *why* people are not interpreting questions as intended. Some of this is undoubtedly due to researcher error, but it's not plausible that it's *entirely* due to inadequate study design. There is no reason to suppose researchers studying folk metaethics are uniquely bad at

designing studies. On the contrary, researchers studying folk metaethics have been highly receptive to the methodological shortcomings of earlier research, and have done far more than is typical in other areas of research to minimize conflation and enhance the proportion of intended interpretations. This has proven so difficult that researchers have resorted to giving participants extensive instructions, training exercises, and extremely complicated response options. This may appear to be a solution to the problem, but it isn't. It simply trades one problem for another. It would appear that the only way to ensure that ordinary people interpret questions about metaethics is to give them an impromptu course in philosophy. Either such efforts fail, in which case we'd need to employ even more robust training protocols, or they succeed, in which case participants are no longer ordinary people, but junior philosophers.

This reveals an insurmountable paradox to folk metaethics research: our goal is to study how people without philosophical training think about a philosophical topic. It turns out that they don't seem to interpret what we're asking them as intended, and the only way we may be able to ensure that they do interpret what we're asking as intended is to train them to do philosophy, exactly as Kauppinen (2007) proposed. Yet once we do this, they are no longer people without philosophical training. This is why I drew on the metaphor of Schrödinger's cat. Once we open the box, we'll find a cat that is either dead or alive, but we won't be able to say that the cat was dead or alive *all along*. Prior to opening the box, there was no determinate fact of the matter about whether the cat was alive or dead. I propose that the same is true for folk metaethics: prior to participating in a study, there is no fact of the matter about whether ordinary people are moral realists or antirealists.

There is little reason to think such indeterminacy is unique to folk metaethics. Instead, I suspect indeterminacy may apply to many classical philosophical concepts and distinctions. Prior to engaging with philosophy, people may have an underlying disposition to think one way or another, but in many cases they have no position on the matter, nor do they speak or think in ways that commit

them to a particular philosophical position. With respect to philosophical stances, they exist in a state of philosophical superposition that resolves into a determinate position only by becoming sufficiently familiar with the philosophical issue in question to form a position on it. Likewise, with respect to philosophical commitments, people's linguistic practices need not conform to any particular philosophical position unless, and only unless, such a commitment actually played a role in regulating their linguistic practices. And we have no justification for presuming that it must play such a role in the absence of any evidence or an adequate theoretical rationale.

Maybe there are good reasons to believe people are implicitly committed to realism or antirealism, or hold substantive stances towards these positions. If so, it will be difficult to account for the existing research on folk metaethics. If people really have positions on these topics, it's unclear why it's so difficult to figure out what they are. And if they have implicit commitments, a century of philosophy and nearly two decades of empirical research have yet to reveal anything about them. Perhaps future studies will succeed where existing ones have failed, but metaethical indeterminacy offers a simple and straightforward alternative explanation: all of this research mistakenly assumed that there was something to measure in the first place. Why should we have ever supposed that ordinary people are moral realists or antirealists, any more than we imagine that they endorse or are implicitly committed to a particular decision theory, e.g., causal decision theory (CDT) or evidential decision theory (EDT) (McCarthy, 2016; Munier, 1988)? There may be no reason to think that the philosophical distinctions philosophers care about capture *any* stable features of ordinary thought and language. The same applies to folk philosophy more broadly. I expect future research in folk philosophy to reveal that researchers grossly underestimated the chasm between ordinary moral thought and philosophical thought, and further, that they incorrectly presume the former mirrors (even if in some nascent way) the latter.

Even if people did have determinate philosophical stances, such research may still be methodologically misguided. I find it profoundly strange that researchers seem not to appreciate just how tall an order it would be to imagine that we could readily solicit responses to complicated philosophical questions using extremely sparse stimuli, stimuli that often include little or no explanation of the relevant concepts, and with almost no efforts made to disambiguate the mountain of potential conflations that would impede people's interpretations of them. After all, one of the central tasks, if not *the* central task of many academic philosophers is to *clarify* the philosophical positions that they are putting forward. Yes, philosophers present arguments. But the bulk of many academic works consists in painstakingly disentangling a concept from a host of potential confusions and misunderstandings, or carefully proposing and distinguishing terminological distinctions with the conceptual equivalent of a surgeon's practiced hand. But after the final satisfying keystroke, one can rest assured that if you say anything interesting at all, some pedant from Neverheardofit University will spring up out of nowhere with an article explaining why you got it all wrong.

And they might have a point. Even professionals struggle to fully grasp the nuances of the topics they write about, and often err in their understanding of concepts they've studied for years. The content of philosophical accounts are highly refined structures. They often bear more resemblance to a Gothic cathedral than the ordinary person's couch. And yet, for some reason, philosophers imagine the equivalent of *Basilica Sancti Francisci Assisiensis* is concealed in everyday thought and speech. Perhaps the reason people struggle so much with questions about metaethics (or questions about philosophy in general) is because there's far less philosophical content implicit in the way people speak and think than researchers imagine. We should consider the sobering possibility that there's no flying buttresses and elaborate frescoes lurking beneath the couch cushions of ordinary thought and language, only stale popcorn and loose change.

This may seem like an extraordinarily bleak conclusion. After all, I'm suggesting that much of the folk philosophical research produced in the past two decades has been fundamentally misguided, delivering results that tell us little or nothing about how ordinary people actually think or speak. It's true that I'm a pessimist about much of this research, and that I'm highly critical of the current state of philosophy and psychology, but my pessimism isn't born out of bitter naysaying and a desire to tear down the establishment. Folk philosophical research started off on the wrong foot. But there is ample time to chart a new course. Such course-correction may be a little humbling and more than a little difficult, but I find the prospects of venturing into uncharted seas exhilarating. Far better to lay new foundations than to lay bricks in a crumbling edifice.

What should we expect to find in charting this new course? I don't know, and that's precisely why it's thrilling to consider a fundamental shift in our methods. While it's too soon to say what we'll find, I would like to end by speculating on how we might get there. The central flaw I've identified in contemporary folk philosophical research has been its emphasis on *top-down, a prioristic* methods. Naturally, then, I believe the way forward will require a *bottom-up* approach that emphasizes descriptive and observational research. This suggestion is precisely in line with the observations Rozin (2001) made in commenting on what he perceived to be social psychology's premature attempt to present itself as a mature science before proceeding through the necessary stages for any science to mature. As Rozin puts it,

I believe that social psychology, modeling itself in the mid-20th century primarily on the natural sciences and on sensory psychology, has concentrated on the advancement of a formal, precise, and experimental science. However, unlike the successful work in the natural sciences and sensory psychology, the work in social psychology has not been preceded by an extensive examination and collection of relevant phenomena and the description of universal or contingent invariances. In the more advanced sciences that social psychology would like to emulate, there is much more emphasis on phenomena and "description" than there is in social psychology, and there is less reliance on experiment. Such sciences, particularly the life sciences, also pay less attention to models and hypotheses and more attention to evidence as opposed to proof or "definitive" studies. (p. 3)

I believe the same is true, to an even greater extent, for research on folk philosophy, and folk metaethics in particular. Unfortunately, most contemporary folk philosophical research hasn't ventured far enough from armchair musings. Instead, it largely consists of soliciting the armchair musings of untrained survey respondents, as though this were an adequate substitute for studying actual moral thought and behavior. Such research may play an important role in our understanding of folk morality, but how is an almost exclusive reliance on such data supposed to yield insights into what people are *actually* thinking and doing in real world instances? Relying on such data is the methodological equivalent of ecologists studying nature by analyzing Bob Ross paintings.¹¹⁶

If we really want to understand folk morality, we should study moral thought, language, and behavior *as it occurs in the real world*. The grist for the philosophical mill must come from deliberate engagement with morality in the environments in which it occurs. In the case of folk moral philosophy, this should begin with a *bottom-up* approach that relies on gathering vast amounts of descriptive and observational data that exhibits a deep appreciation for cultural, linguistic, and sociodemographic variations in human populations and the recognition that morality is a *social* phenomenon that should be studied in the ecological contexts in which it occurs. Such research should deliberately eschew the imposition of philosophical concepts and categories onto the data in an attempt to categorize all responses in accordance with the categories and distinctions in analytic philosophy. If we begin empirical inquiry with the presumption that folk thought *must* conform to a predetermined set of categories and distinctions, we can overlook genuine alternatives within those categories. At the same time, we also risk failing to observe features of folk thought that have little or nothing to do with these categories. It may turn out that there are many features of the way ordinary people think about (e.g.,

¹¹⁶ Granted, enough observational research and engagement with empirical evidence can facilitate clever hypotheses from the armchair. Galileo's thought experiment disproving Aristotelian conceptions of falling bodies exemplifies this (Gruszczyński, 2020). However, philosophers rarely engage in rigorous, systematic observational research of ordinary moral thought and practice.

the nature of morality) that aren't reducible to whether they endorse realism or antirealism, cognitivism or noncognitivism, and so on. We'll never know if we only design studies that attempt to sort people according to traditional philosophical distinctions. I am not suggesting we never conduct experiments, or engage in theorizing, or attempt to categorize participants in usefully discrete ways. I am simply claiming that such efforts are the task of a more mature science that has a rich dataset to work with, not the anecdotes and armchair speculation of a handful of unrepresentative academics. Here are a few elements that would be critical to the success of this research:

1. *Extensive pretesting to assess interpretation rates for traditionally philosophical topics.*

Researchers should not presume people understand the terms and concepts used in folk philosophical research. Instead, they should engage in extensive pretesting to assess how people interpret stimuli in order to determine whether people are interpreting stimuli as intended. This will require doing far more than including a single, trivially easy comprehension check, then declaring that participants understood what was asked because most answered this question correctly. It may involve using open response questions, cognitive interviewing, focus groups, and a variety of qualitative and quantitative methods to assess how well people understand questions. Note that evidence of adequate interpretation rates in one population doesn't justify the inference that *other* populations would interpret items as intended. Variables such as age, language, culture, and socioeconomic status could influence how different populations interpret stimuli.

2. *Account for interpretative variation.*

Researchers should be mindful of the possibility that participants interpret the same stimuli differently from one another, and factor this into their designs. Interpretative variation may contribute an unavoidable degree of noise to studies, and it may be substantial in cases where items are vague or ambiguous.

3. *More interdisciplinary collaboration*

Psychologists studying folk metaethics have often made serious errors in how they've operationalized metaethical concepts. This could have been avoided if they'd worked more closely with trained philosophers. Philosophers, in turn, often employ poorly designed studies due to an insensitivity to considerations that psychologists are keenly aware of. Researchers from a variety of backgrounds should work together to complement one another's skillsets.

4. *Expand methods beyond traditional surveys*

Researchers rely far too much on traditional surveys. Researchers should instead conduct more field studies and engage in more observational research, drawing on or engaging in ethnographic research by e.g., consulting content from the HRAF, or engaging in corpus analysis (for a recent example of the latter, see Stojanovic & McNally, 2022).

5. *Conduct more descriptive research*

Researchers often attempt to manipulate philosophical positions *as if* they were psychological constructs, without first establishing that they are, and that these attributes function in the way psychological constructs do. This is a mistake. You can't manipulate something that isn't there. While there is nothing wrong with manipulating folk philosophical beliefs and attitudes that *are* present, you have to first demonstrate that they're there before you can manipulate them.

6. *Don't assume one-to-one mapping between philosophical concepts and psychological constructs*

Researchers should not presume that a particular philosophical concept automatically corresponds to an analogous psychological construct, such that there's a one-to-one correspondence between philosophical positions and ordinary psychology. We should not assume, for instance, that since philosophers tend to endorse compatibilism, incompatibilism, or libertarian free will, that these are reflected in ordinary human psychology as three distinct, corresponding psychological profiles. Nor can we presume that because philosophers use a particular term to refer to a particular concept, that belief or disbelief in this concept is represented by a single, continuous variable in the population. To again draw on the concept of free will, it makes no sense to think that *all* ordinary people can be placed on a single spectrum of greater or lesser belief in "free will," since such measures presume that there is a shared concept that people exhibit varying degrees of belief or disbelief in.

7. *Don't assume associations that philosophers endorse are present in folk philosophy*

Philosophers often believe that particular philosophical positions or concepts are closely associated with other positions and concepts, or that philosophical concepts have entailments or implications for other philosophical positions. But they should not assume that such associations are present in ordinary thought. For instance, philosophers may perceive a very close association between academic disputes about free will and moral responsibility, so much so that one could almost define academic notions of free will as those features of agency necessary for certain forms of moral praise and blame. Yet there is little evidence ordinary people psychologically link free will and moral responsibility in the same way as philosophers.

8. *Avoid forced choice paradigms*

Researchers should avoid forced choice paradigms that require participants to rigidly fit into a narrow subset of categories. Such studies can force people to respond to questions in ways that don't represent what they think, or exacerbate spontaneous theorizing.

9. *Develop the concept of spontaneous theorizing and test for it*

The notion of spontaneous theorizing proposed here emerged over the course of my research. There may be similar notions already present in the literature that I'm not aware of. If so, perhaps they are highly developed. But if they are absent or underdeveloped, greater effort should be made to expand on the concept of spontaneous theorizing, and to devise ways of testing for the role that it plays in any given study. It may be that a much wider range of folk philosophical research is capturing conclusions people reached by participating in the study, rather than stances or commitments they held prior to doing so. If so, this could represent a significant threat to a great deal of research.

10. *Conduct more cross-cultural research*

One of the more obvious shortcomings with existing research is the lack of cross-cultural research. Growing evidence suggests WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations are one of the least psychologically representative populations in the world. Findings that would generalize to humanity as a whole already require more diverse sampling on these grounds alone. However, the need for more culturally diverse samples is especially important for folk philosophy, since most studies draw on categories and distinctions derived from an extremely narrow subset of WEIRD populations.

11. *Study broader array of philosophical traditions, and don't ignore less popular philosophical views*

Folk philosophical research often opts to simplify studies by exclusively assessing folk stances and commitments towards the most popular or simplified versions of a philosophical position. This is why, for instance, researchers have focused almost exclusively on realism and antirealism, or in some cases realism and cultural relativism or individual subjectivism. It took several years before researchers began to include noncognitivism. Yet even this doesn't cover the variety of views that are actually out there. Noncognitivism has been exclusively portrayed as a crude form of emotivism, when contemporary expressivist theories are richer and more sophisticated. And so far, no studies have incorporated measures of constructivism. There is little justification for the exclusion of constructivism. The 2020 PhilPapers survey showed that 20.8% of philosophers endorse or lean towards constructivism, which was greater than expressivism (10.6%) and error theory (5.3%) (Bourget & Chalmers, ms). Why, then, do the

most recent studies include response options for expressivism and error theory, but not constructivism?

12. *Draw on psychological theories and approaches to inform hypotheses*

Lastly, researchers can and should draw on the wealth of psychological approaches, theories, and phenomena to inform folk philosophical hypotheses, rather than on philosophical tradition alone.

I want to end on this last suggestion by pointing to one example of how this could be done. One of the most obvious shortcomings to my argument for metaethical indeterminacy is that it consists solely of a negative thesis. I'm claiming that ordinary people have no determinate stance towards realism and antirealism. This entails that when ordinary people engage in everyday moral judgment, reasoning, and behavior, they're *not* making claims that commit them to realism, and it further entails that views about realism and antirealism don't typically figure into their judgments and reasoning at all. Yet this doesn't tell us what people *are* doing.

The traditional philosophical view presumes when people make moral claims and engage in moral judgment, that they are implicitly appealing to a principled and logically consistent set of moral standards, or are at least *attempting* to do so. As a result, the primary purpose of moral judgment is to assert what's *true*, and the primary purpose of moral reasoning is to enhance moral knowledge by identifying and resolving inconsistencies and errors. While this may be the purpose of moral *philosophy*, researchers seem to imagine that ordinary moral psychology functions in the same way, and that to the extent that ordinary moral thought is biased or inconsistent, this is an unfortunate byproduct of faulty and unrefined thinking.

There is good reason to question these assumptions. Perhaps the apparent biases and inconsistencies in ordinary moral thought are a feature, and not a bug. Moral reasoning is a specific instance of our more general capacity for reasoning. And while researchers have traditionally supposed that the purpose of reasoning is to enable us to have more accurate beliefs, this may not be true. As

Sperber and Mercier (2011; 2017) point out, reasoning is subject to a variety of biases and distortions that seem inconsistent with a psychological system optimized for delivering true beliefs. They propose that reason functions primarily to enable us to persuade others, and is thus best suited to constructing and evaluating arguments. They marshal an impressive body of reasons and evidence for taking this proposal seriously, which they summarize as follows:

Reasoning so conceived is adaptive given the exceptional dependence of humans on communication and their vulnerability to misinformation. A wide range of evidence in the psychology of reasoning and decision making can be reinterpreted and better explained in the light of this hypothesis. Poor performance in standard reasoning tasks is explained by the lack of argumentative context. When the same problems are placed in a proper argumentative setting, people turn out to be skilled arguers. Skilled arguers, however, are not after the truth but after arguments supporting their views. This explains the notorious confirmation bias. This bias is apparent not only when people are actually arguing, but also when they are reasoning proactively from the perspective of having to defend their opinions. (2011, p. 57)

I suspect that incorporating this account of reasoning into the way we analyze ordinary moral thought could go a long way towards better understanding what people are actually doing in everyday moral discourse. Philosophers have observed that people use moral language in a variety of ways. It sometimes appears to reference stance independent moral facts, while at other times it serves to make claims about our standards or the standards of our culture. In still other cases, it serves to express our emotional states or to issue commands. Philosophers look at this mess and imagine that one of these uses is the proper function of moral language, while the rest are aberrant or parasitic on its primary and central function (Gill, 2008). This presumes that ordinary moral language must conform to a particular metaethical thesis, such that the role of moral claims is implicitly driven by a principled commitment to logical consistency and the goal of accurately expressing one's moral beliefs. This is why philosophers have either never entertained pluralism, or objected to it when it's been proposed (Johansson & Olson, 2015). Indeed, Colebrook (2021) argues that if pluralism is true, this would entail that folk metaethical judgments are *irrational*.

The argumentative theory provides an alternative explanation. To the extent that ordinary moral language appears in some contexts to involve reference to stance-independent facts, in others to appeal to the standards of people or groups, and in others to serve a primarily expressive function, these are not genuine reflections of an underlying commitment to an inconsistent or irrational set of metaethical standards. Instead, such language is superficial, and functions primarily to serve *argumentative* and *rhetorical* purposes. In everyday contexts, moral judgments have consequences. Our moral attitudes towards others can lead to praise and blame, loss of friends, increases or decreases in our status, and all manner of practically relevant consequences. In many cases, the outcome of a moral judgment is a literal matter of life and death. Why, then, should we imagine that people are primarily driven to say what's *true*? Why not, instead, imagine that the primary function of moral judgment, reasoning, and discourse is to facilitate our sociofunctional goals? Humans are a social species. Our reproductive fitness is typically enhanced not by doing philosophy well, but by saying and doing things that will allow us to enhance our status and prestige, to form fruitful relationships with others, to insulate ourselves from threats to our health and success, and to avoid punishment, blame, and loss of reputation. *This* is the context in which moral thought and discourse occurs. Why assume that none of this is relevant to its function?

Once we reorient our perspective on the role of moral thought and language in everyday life, we can see why the apparent inconsistency of moral language is nothing of the sort. While ordinary moral language may not consistently fit with traditional philosophical accounts, this is because such accounts appeal to *semantic* consistency. Yet the argumentative theory centers consistency in the *functional role* of moral language: if using realist-sounding language would serve one's argumentative goal in one context, while using antirealist-sounding language would serve those same goals in a different context, people would be using such language consistently *with respect to the goal of persuading others*. If so, ordinary moral language may appear to be used in ways that fit with one or another

metaethical analysis, not because such language reflects a genuine commitment to a bizarre, highly variable form of moral pluralism, but because people vary how they talk about morality to serve different argumentative ends in different contexts. If so, apparent commitments to metaethical positions aren't so much *built into* the semantics of moral language, so much as they are external semantic drifters capable of fusing with moral language in particular conversational contexts to serve various social goals, without being incorporated into the deeper commitments that characterize moral language *itself*.

This suggestion is little more than a speculative hypothesis that may or may not be vindicated by future research. I don't bring up the argumentative theory of reasoning to defend its application to folk metaethics. I present it as an example of the kinds of empirically grounded paradigm shifts that can fundamentally alter the course of folk philosophical research. My criticisms of folk metaethics may give the impression that I believe there is little work to be done. This example also serves to highlight that nothing could be further from the truth. My critique of folk metaethics does not show that studying how ordinary people think about philosophical topics is a waste of time. On the contrary, it highlights how ordinary thought is deeply embedded in the practical and purpose-driven nature of everyday life. Philosophy was once seen as a method for reflecting, not just on abstract and esoteric matters with little practical relevance, but as a tool for assessing how to live well and pursue a life of purpose, meaning, and value. Far from the most insular and irrelevant of disciplines, it served as the foundation not only for science and academia, but for the ordinary person. If I've demonstrated anything, it's that philosophy has strayed far from this path. While I'm advocating substantial reform and innovation, I am, in a somewhat paradoxical way, also suggesting that philosophy return to its roots.

REFERENCES

- Abshire, J. (2011). *The history of Singapore*. Santa Barbara, CA: ABC-CLIO.
- Aczel, B., Bago, B., Szollosi, A., Foldes, A., & Lukacs, B. (2015). Is it time for studying real-life debiasing?. Evaluation of the effectiveness of an analogical intervention technique. *Frontiers in Psychology*, 6, 1120. <https://doi.org/10.3389/fpsyg.2015.01120>
- Adams, F., & Steadman, A. (2004). Intentional action and moral considerations: Still pragmatic. *Analysis*, 64(3), 268-276.
- Aharoni, E., Sinnott-Armstrong, W., & Kiehl, K. A. (2012). Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *Journal of Abnormal Psychology*, 121(2), 484-497.
- Aikins, H. A. (1902). *The principles of logic*. New York: Henry Holt and Company.
- Alexander, J. (2012). *Experimental philosophy: An introduction*. Cambridge, UK: Polity Press.
- Alexander, J., Mallon, R., & Weinberg, J. M. (2010). Accentuate the negative. *Review of Philosophy and Psychology*, 1(2), 297-314.
- Alexander, S. (2009, April 28). Generalizing from one example. *LessWrong*. <https://www.lesswrong.com/posts/baTWMegR42PAsH9qJ/generalizing-from-one-example>
- Alexander, S. (2013, April 12). Noisy poll results and reptilian muslim climatologists from Mars. *LessWrong*. <https://www.lesswrong.com/posts/CsKrQdQJJCFPjfkjF/noisy-poll-results-and-reptilian-muslim-climatologists-from-mars/>
- Alicke, M., Gordon, E., & Rose, D. (2013). Hypocrisy: what counts?. *Philosophical Psychology*, 26(5), 673-701.
- Allen, K., Quinlan, P., Andow, J., & Fischer, E. (2021). What is it like to be colour-blind? A case study in experimental philosophy of experience. *Mind & Language*, 1-26. <https://doi.org/10.1111/mila.12370>
- Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment*, 38(1), 1-5.
- Allidina, S., & Cunningham, W. A. (2018). Moral cues from ordinary behaviour. *Behavioral and Brain Sciences*, 43. <https://doi.org/10.1017/S0140525X18000018>
- Altman, D. G., & Bland, J. M. (1996). Absence of evidence is not evidence of absence. *British veterinary journal*, 74(4), 311-311.

- Amira, K., Cooper, C. A., Knotts, H. G., & Wofford, C. (2018). The southern accent as a heuristic in American campaigns and elections. *American Politics Research*, 46(6), 1065-1093.
- Andow, J. (2016). Qualitative tools and experimental philosophy. *Philosophical Psychology*, 29(8), 1128-1141.
- Apter, M. J. (1982). The experience of motivation: The theory of psychological reversals. London, UK: Academic Press.
- Ardila-Rey, A., & Killen, M. (2001). Middle class Colombian children's evaluations of personal, moral, and social-conventional interactions in the classroom. *International Journal of Behavioral Development*, 25(3), 246-255.
- Arnett, J. J. (2009). The neglected 95%, a challenge to psychology's philosophy of science. *American Psychologist*, 64(6), 571-574.
- Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., & Bong, C. H. (2014). Predicting survey responses: How and why semantics shape survey statistics on organizational behaviour. *PloS One*, 9(9), e106361. <https://doi.org/10.1371/journal.pone.0106361>
- Asch, S. E. (1987). *Social psychology*. New York, NY: Oxford University Press. (Original work published 1952)
- Ash, K., Johnson, W., Lagadinos, G., Simon, S., Thomas, J., Wright, E., & Gainous, J. (2020). Southern accents and partisan stereotypes: Evaluating political candidates. *Social Science Quarterly*, 101(5), 1951-1968.
- Ayars, A., & Nichols, S. (2020). Rational learners and metaethics: Universalism, relativism, and evidence from consensus. *Mind & Language*, 35(1), 67-89.
- Babushkina, D. (2019). Bradley's "my station and its duties" and its moral (in) significance. *Zeitschrift für Ethik und Moralphilosophie*, 2(2), 195-211.
- Bagnoli, C. (2021). Constructivism in metaethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 ed.). <https://plato.stanford.edu/archives/spr2021/entries/constructivism-metaethics>
- Baguley, T. (2012, June 21). *The stimuli-as-fixed-effect fallacy*. R-bloggers. <https://www.r-bloggers.com/2012/06/the-stimuli-as-a-fixed-effect-fallacy/>
- Ballester, J., Abdi, H., Langlois, J., Peyron, D., & Valentin, D. (2009). The odor of colors: Can wine experts and novices distinguish the odors of white, red, and rosé wines?. *Chemosensory Perception*, 2(4), 203-213.

- Barnard, R., & Ulatowski, J. (2013). Truth, correspondence, and gender. *Review of Philosophy and Psychology*, 4(4), 621-638.
- Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9), 536-554.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323-370.
- 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(2), 260-268.
- Baz, A. (2012). *When words are called for*. Cambridge, MA: Harvard University Press.
- Baz, A. (2015). Questioning the method of cases fundamentally—Reply to Deutsch. *Inquiry*, 58(7-8), 895-907.
- Baz, A. (2017). *The crisis of method in contemporary analytic philosophy*. Oxford, UK: Oxford University Press.
- Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287-311.
- Becklooyd, D. (2021). *Is common sense neither compatibilist nor incompatibilist?* [Master's thesis, Open Access Te Herenga Waka-Victoria University of Wellington].
<https://doi.org/10.26686/wgtn.14927007.v1>
- Beebe, J. R. (2012). Experimental epistemology. In A. Cullison (Ed.), *Continuum companion to epistemology* (pp. 248-269). London, UK: Continuum.
- Beebe, J. R. (2014). How different kinds of disagreement impact folk metaethical judgments. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 167-187). London, UK: Bloomsbury.
- Beebe, J. R. (2015). The empirical study of folk metaethics. *Etyka*, 50, 11-28.
- Beebe, J. R. (2020). The empirical case for folk indexical moral relativism. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 4, pp. 81-111). Oxford, UK: Oxford University Press.
- Beebe, J. R., & Sackris, D. (2016). Moral objectivism across the lifespan. *Philosophical Psychology*, 29(6), 912-929.

- Beebe, J. R., Qiaoan, R., Wysocki, T., & Endara, M. A. (2015). Moral objectivism in cross-cultural perspective. *Journal of Cognition and Culture*, 15(3-4), 386-401.
- Bergenholtz, C., Busch, J., & Praëm, S. K. (2021). Exclusion Criteria in Experimental Philosophy. *Erkenntnis*, 86(6), 1531-1545.
- Berniūnas, R. (2014). Delineating the moral domain in moral psychology. *Problemos*, (Supplement), 90-101.
- Berniūnas, R. (2020). Mongolian yos surtakhuun and WEIRD “morality”. *Journal of Cultural Cognitive Science*, 4(1), 59-71.
- Berniūnas, R., Beinorius, A., Dranseika, V., Silius, V., & Rimkevičius, P. (2021). The weirdness of belief in free will. *Consciousness and Cognition*, 87, 103054. <https://doi.org/10.1016/j.concog.2020.103054>
- Berniūnas, R., Silius, V., & Dranseika, V. (2022). Moralization East and West: Moralizing different transgressions among Chinese, Americans, and Lithuanians. *Asian Journal of Social Psychology*, 25(2), 185-197.
- Bervoets, J., Milton, D., & Van de Cruys, S. (2021). Autism and intolerance of uncertainty: an ill-fitting pair. *Trends in Cognitive Sciences*, 25(12), 1009-1010.
- Bex-Priestley, G. (2018). Error and the limits of quasi-realism. *Ethical Theory and Moral Practice*, 21(5), 1051-1063.
- Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18(5), 382-386.
- Birch, S. A., Brosseau-Liard, P. E., Haddock, T., & Ghrear, S. E. (2017). A ‘curse of knowledge’ in the absence of knowledge? People misattribute fluency when judging how common knowledge is among their peers. *Cognition*, 166, 447-458.
- Björnsson, G. (2012). Do ‘objectivist’ features of moral discourse and thinking support moral objectivism?. *The Journal of ethics*, 16(4), 367-393.
- Blackburn, S. (1991). Just causes. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 61(1/2), 3-17.
- Blackburn, S. (1993). *Essays in quasi-realism*. Oxford, UK: Oxford University Press.
- Blackburn, S. (2000). *Ruling passions: A theory of practical reasoning*. Oxford, UK: Oxford University Press.

- Blair, J. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1-29.
- Blair, J., Jones, L., Clark, F., & Smith, M. (1995). Is the psychopath 'Morally Insane'? *Personality and Individual Differences*, 19(5), 741-752.
- Blair, R. J. R. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences*, 22(5), 731-739.
- Bojanowski, J. (2012). Is Kant a moral realist?. *Kant Yearbook*, 4(1), 1-22.
- Bonard, C., Cova, F., & Humbert-Droz, S. (2021). De gustibus est disputandum: An empirical investigation of the folk concept of aesthetic taste. In *Perspectives on Taste* (pp. 77-108). Routledge.
- Bonge, D. R., Schuldt, W. J., & Harper, Y. Y. (1992). The experimenter-as-fixed-effect fallacy. *The Journal of Psychology*, 126(5), 477-486.
- Booth, A. R., & Rowbottom, D. P. (Eds.) (2014). *Intuitions*. Oxford, UK: Oxford University Press.
- Bourget, D. & Chalmers, D. J. (ms) *Philosophers on Philosophy: The PhilPapers 2020 Survey*.
- Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe?. *Philosophical Studies*, 170(3), 465-500.
- Bower, B. (2005, December 4). The Pirahã Challenge. *ScienceNews*
<https://www.sciencenews.org/article/pirah%C3%A3-challenge> (accessed June 10, 2022)
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Braun, V., & Clarke, V. (2014). What can “thematic analysis” offer health and wellbeing researchers?. *International Journal of Qualitative Studies on Health and Well-being*, 9(1), 26152.
<https://doi.org/10.3402/qhw.v9.26152>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589-597.
- Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, 21(1), 37-47.
- Brogaard, B. (2008). Moral contextualism and moral relativism. *The Philosophical Quarterly*, 58(232), 385-409.

- Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361-379). New York, NY: Guilford Press.
- Buchtel, E. E., Guan, Y., Peng, Q., Su, Y., Sang, B., Chen, S. X., & Bond, M. H. (2015). Immorality east and west: Are immoral behaviors especially harmful, or especially uncivilized?. *Personality and Social Psychology Bulletin*, 41(10), 1382-1394.
- Buckley, C. B. (1902). *An anecdotal history of Singapore (vol. 1)*. Singapore: Fraser & Neave, Limited.
- Buckwalter, W., & Turri, J. (2015). Inability and obligation in moral judgment. *PLoS One*, 10(8), e0136589. <https://doi.org/10.1371/journal.pone.0136589>
- Bush, L. S., & Moss, D. (2020). Misunderstanding metaethics: Difficulties measuring folk objectivism and relativism. *Diametros* 17(64): 6-21.
- Bush, L.S. (2016). Cultural & moral relativism. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-16999-6_518-1
- Bush, L.S., & Millhouse, T. (2016, June). *Operationalizing metaethics: The disagreement paradigm in empirical moral psychology* [conference session]. The 42nd Annual Meeting of the Society for Philosophy and Psychology, Austin, TX. <https://www.socphilpsych.org/spp2016program.pdf>
- Çağaptay, S. (2014). *The rise of Turkey: The twenty-first century's first Muslim power*. Dulles, VA: Potomac Books.
- Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The concept of external validity. *Journal of Consumer Research*, 9(3), 240-244.
- Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97(5), 1232-1254.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Cappelen, H. (2012). *Philosophy without intuitions*. Oxford, UK: Oxford University Press.
- Cappelen, H. (2014). X-phi without intuitions?. In A. Booth and D. Rowbottom (Eds.). *Intuitions* (pp. 269-286). New York, NY: Oxford University Press.
- Carnes, N. C., Allmon, B., Alva, J., Cousar, K. A., & Varnam, Z. D. (2022). How morality signals, benefits, binds, and teaches. *Journal of Experimental Social Psychology*, 101, 104313. <https://doi.org/10.1016/j.jesp.2022.104313>

- Carroll, L. (2022). *Jabberwocky*. Poetry Foundation.
<https://www.poetryfoundation.org/poems/42916/jabberwocky> (Original work published in 1871)
- Carter, D. B., & Patterson, C. J. (1982). Sex roles as social conventions: The development of children's conceptions of sex-role stereotypes. *Developmental Psychology*, 18(6), 812-824.
- Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology nursing forum*, 41(5), 545-547.
- Caspar, E. A., Verdin, O., Rigoni, D., Cleeremans, A., & Klein, O. (2017). What do you believe in? French translation of the FAD-plus to assess beliefs in free will and determinism and their relationship with religious practices and personality traits. *Psychologica Belgica*, 57(1), 1-16.
<https://doi.org/10.5334/pb.321>
- Caspar, E. A., Vuillaume, L., Magalhães De Saldanha da Gama, P. A., & Cleeremans, A. (2017). The influence of (dis)belief in free will on immoral behavior. *Frontiers in Psychology*, 8, Article 20.
<https://doi.org/10.3389/fpsyg.2017.00020>
- Cassady, J. C., & Finch, W. H. (2014). Confirming the factor structure of the Cognitive Test Anxiety Scale: Comparing the utility of three solutions. *Educational Assessment*, 19(3), 229-242.
- Castensen, M. W., Sellmaier, S., Taylor, P. C. J., Deroy, O. (2022). Free will properly understood: Folk conceptions are incompatibilist, but not hard incompatibilist. SSRN.
<http://dx.doi.org/10.2139/ssrn.4024722>
- Caven, P. (2015). *Moral Disagreement-A Psychological Account and the Political Implications*. (uk.bl.ethos.658757) [Doctoral dissertation, University of Sheffield]. White Rose eTheses Online.
- Caverni, J. P., Fabre, J. M., Gonzalez, M. (1990). *Cognitive biases*. Amsterdam, NL: Elsevier.
- Chalmers, D. J. (2014). Intuitions in philosophy: A minimal defense. *Philosophical Studies*, 171(3), 535-544.
- Chan, H. M., & Yan, H. K. (2007). Is there a geography of thought for east-west differences? Why or why not?. *Educational Philosophy and Theory*, 39(4), 383-403.
- Chance, Z., & Norton, M. I. (2015). The what and why of self-deception. *Current Opinion in Psychology*, 6, 104-107.
- Chappell, R. Y. (2007, June 19). *Context and relativism*. Philosophy, et cetera.
<https://www.philosophyetc.net/2007/06/context-and-relativism.html>

- Chartrand, L. (2022). Modeling and corpus methods in experimental philosophy. *Philosophy Compass*, e12837. <https://doi.org/10.1111/phc3.12837>
- Chituc, V., Henne, P., Sinnott-Armstrong, W., & De Brigard, F. (2016). Blame, not ability, impacts moral “ought” judgments for impossible actions: Toward an empirical refutation of “ought” implies “can”. *Cognition*, 150, 20-25.
- Choi, I., Koo, M., & Choi, J. A. (2007). Individual differences in analytic versus holistic thinking. *Personality and Social Psychology Bulletin*, 33(5), 691-705.
- Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2015). Waste management: How reducing partiality can promote efficient resource allocation. *Journal of Personality and Social Psychology*, 109(2), 210-231.
- Chudnoff, E. (2013). *Intuition*. Oxford, UK: Oxford University Press.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335-359.
- Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist*, 26(2), 120--123.
- Climenhaga, N. (2018). Intuitions are used as evidence in philosophy. *Mind*, 127(505), 69-104.
- Cline, B. (2018). The tale of a moderate normative skeptic. *Philosophical Studies*, 175(1), 141-161.
- Cohen, Y. (2018). An analysis of recent empirical data on ‘ought’ implies ‘can’. *Philosophia*, 46(1), 57-67.
- Cohnitz, D., & Häggqvist, S. (2010). The role of intuitions in philosophy. *Studia Philosophica Estonica*, 2.2, 1-14. <https://doi.org/10.12697/spe.2009.2.2.01>
- Colebrook, R. (2021). The irrationality of folk metaethics. *Philosophical Psychology*, 34(5), 684-720.
- Colebrook, R., & Sarkissian, H. (2018). Objectivity. In T. Shackelford & V. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science*. Springer International Publishing.
- Collier-Spruel, L., Hawkins, A., Jayawickreme, E., Fleeson, W., & Furr, R. M. (2019). Relativism or tolerance? Defining, assessing, connecting, and distinguishing two moral personality features with prominent roles in modern societies. *Journal of Personality*, 87(6), 1170-1188.
- Compen, B., Pitthan, F., Schelfhout, W., & De Witte, K. (2022). How to elicit and cease herding behaviour? On the effectiveness of a warning message as a debiasing decision support system. *Decision Support Systems*, 152, 113652. <https://doi.org/10.1016/j.dss.2021.113652>

- Cooper, D. E. (1978). Moral relativism. *Midwest Studies in Philosophy*, 3(1), 97-108.
- Cosmides, L., & Tooby, J. (2016). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91-115). New York, NY: Guilford.
- Countering Moral Relativism (2015, April 21). *California Catholic conference*.
<https://cacatholic.org/article/countering-moral-relativism-spiritual-poverty-our-time>
- Cova, F., & Pain, N. (2012). Can folk aesthetics ground aesthetic realism?. *The Monist*, 95(2), 241-263.
- Cova, F., Garcia, A., & Liao, S. Y. (2015). Experimental philosophy of aesthetics. *Philosophy Compass*, 10(12), 927-939.
- Cova, F., Olivola, C. Y., Machery, E., Stich, S., Rose, D., Alai, M., ... & Zhu, J. (2019). De Pulchritudine non est Disputandum? A cross-cultural investigation of the alleged intersubjective validity of aesthetic judgment. *Mind & Language*, 34(3), 317-338.
- Crescioni, A. W., Baumeister, R. F., Ainsworth, S. E., Ent, M., & Lambert, N. M. (2016). Subjective correlates and consequences of belief in free will. *Philosophical Psychology*, 29(1), 41-63.
- Crone, D. L., & Levy, N. L. (2019). Are free will believers nicer people? (Four studies suggest not). *Social Psychological and Personality Science*, 10(5), 612-619.
- Curry, O. S., Alfano, M., Brandt, M. J., & Pelican, C. (2021). Moral molecules: Morality as a combinatorial system. *Review of Philosophy and Psychology*, 1-20. <https://doi.org/10.1007/s13164-021-00540-x>
- Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, 78, 106-124.
- Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, 60(1), 47-69.
- da Vela, B., & Earley, B. (2015). Senatus and Sénat: the reception of the Roman Senate during the Radical Stage of the French Revolution (1792-4). *Classical Reception Journal*, 7(1), 46-63.
- Damon, W. (1977). *The social world of the child*. San Francisco, CA: Jossey-Bass.
- Dancy, J. (2004). *Ethics without principles*. Oxford, UK: Clarendon Press.
- Dancy, J. (2017) Moral particularism. In E. N Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2017 ed.). <https://plato.stanford.edu/archives/win2017/entries/moral-particularism/>

- Darwall, S. (2006). Morality and practical reason: A Kantian approach. In D. Copp, (Ed.), *The Oxford handbook of ethical theory* (pp. 282-320). New York, NY: Oxford University Press.
- Davidson, D. (1973). On the very idea of a conceptual scheme. *Proceedings and Addresses of the American Philosophical Association*, 47, 5-20.
- Davis, T. (2021). Beyond objectivism: new methods for studying metaethical intuitions. *Philosophical Psychology*, 34(1), 125-153.
- De Freitas, J., Tobia, K. P., Newman, G. E., & Knobe, J. (2017). Normative judgments and individual essence. *Cognitive Science*, 41, 382-402.
- Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology*, 32(1), 1-12.
- de Muynck, W. M. (2004) *Foundations of quantum mechanics, an empiricist approach*. Dordrecht, NL: Kluwer Academic Publishers.
- Denman, D. C., Baldwin, A. S., Betts, A. C., McQueen, A., & Tiro, J. A. (2018). Reducing "I don't know" responses and missing survey data: Implications for measurement. *Medical Decision Making*, 38(6), 673-682.
- Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.
- Deutsch, M. (2009). Experimental philosophy and the theory of reference. *Mind & Language*, 24(4), 445-466.
- Deutsch, M. (2010). Intuitions, counter-examples, and experimental philosophy. *Review of Philosophy and Psychology*, 1(3), 447-460.
- Deutsch, M. (2015). *The myth of the intuitive*. Cambridge, MA: MIT Press.
- Devaine, M., Hollard, G., & Daunizeau, J. (2014). Theory of mind: did evolution fool us?. *PloS One*, 9(2), e87619. <https://doi.org/10.1371/journal.pone.0087619>
- Dhaliwal, N. A., Patil, I., & Cushman, F. (2021). Reputational and cooperative benefits of third-party compensation. *Organizational Behavior and Human Decision Processes*, 164, 27-51.
- Díaz, R. (2019). Using fMRI in Experimental Philosophy: Exploring the Prospects. In E. Fischer & M. Curtis (Eds.) *Methodological advances in experimental philosophy* (pp. 131-152). London, UK: Bloomsbury Publishing.

- Dink, J. W., & Rips, L. J. (2017). Folk teleology and its implications. In D. Rose (Ed.), *Experimental metaphysics* (pp. 207-235). London, UK: Bloomsbury Academic.
- Dionysius of Halicarnassus. (1940). *Roman Antiquities, Volume III: Books 5-6.48* (E. Cary, Trans.). Cambridge, MA: Harvard University Press.
- Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. *Psychology of Learning and Motivation*, 50, 307-338.
- Doğruyol, B., Alper, S., & Yilmaz, O. (2019). The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences*, 151, 109547. <https://doi.org/10.1016/j.paid.2019.109547>
- Doing reflexive TA. (n.d.) *Thematic analysis*. <https://www.thematicanalysis.net/doing-reflexive-ta/>
- Dominici, L. (2018, July 24). *Moral relativism and its effects*. Seven Pillars Institute for Global Finance and Ethics. <https://sevenpillarsinstitute.org/moral-relativism-and-its-effects/>
- Doral, M. (2010). Turkey: Another \$1 Trillion Emerging Economy?. *Journal of Emerging Knowledge on Emerging Markets*, 2(November 2010), 71-83.
- Dorsey, D. (2006). A coherence theory of truth in ethics. *Philosophical Studies*, 127(3), 493-523.
- Dranseika, V., Berniūnas, R., & Silius, V. (2018). Immorality and bu daode, unculturedness and bu wenming. *Journal of Cultural Cognitive Science*, 2(1), 71-84.
- Dreier, J. (2004). Meta-ethics and the problem of creeping minimalism. *Philosophical Perspectives*, 18, 23-44.
- Drennan, J. (2003). Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*, 42(1), 57-63.
- Dupré, J. (2016). Against maladaptationism: or what's wrong with evolutionary psychology?. In M. Mazzotti (Ed.), *Knowledge as social order: Rethinking the sociology of Barry Barnes* (pp. 165-180). New York, NY: Routledge.
- Dweck, C. (2006). *Mindset: The new psychology of success*. Philadelphia, PA: Psychology Press.
- Dwyer, S. (1999). Moral competence. In K. Murasugi & R. Stainton (Eds.), *Philosophy and linguistics* (pp. 169-190). Boulder, CO: Westview Press.
- Dwyer, S. (2004). How good is the linguistic analogy. In Carruthers, P., Laurence, S., & Stich, S. (Eds.), *The innate mind: Culture and cognition* (Vol. 2, pp. 237-256). New York, NY: Oxford University Press.

- Dwyer, S., Huebner, B., & Hauser, M. D. (2010). The linguistic analogy: Motivations, results, and speculations. *Topics in Cognitive Science*, 2(3), 486-510.
- Earlenbaugh, J., & Molyneux, B. (2009). Intuitions are inclinations to believe. *Philosophical Studies*, 145(1), 89-109.
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133-140.
- Enoch, D. (2014). Why I am an objectivist about ethics (and why you are, too). *The ethical life*, 192-205.
- Enoch, D. (2005). Why idealize?. *Ethics*, 115(4), 759-787.
- Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current anthropology*, 46(4), 621-646.
- Everett, D. (2008). Don't sleep, there are snakes: Life and language in the Amazonian jungle. New York, NY: Pantheon books.
- Everett, D. L. (1983). *A língua Pirahã ea teoria da sintaxe: Descrição, perspectivas e teoria* [Doctoral dissertation, Unicamp Campinas].
- Everett, D. L. (1986). *Pirahã*. In by D. Derbyshire and G. Pullum (Eds.), *Handbook of Amazonian languages* (Vol. 1, pp. 200-326). Berlin: Mouton de Gruyter.
- Everett, D. L. (2009b). Pirahã culture and grammar: A response to some criticisms. *Language*, 405-442.
- Everett, D. L. (2012). What does Pirahã grammar have to teach us about human language and the mind?. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(6), 555-563.
- Everett, J. A., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, 79, 200-216.
- Evers, D. (2014). Moral contextualism and the problem of triviality. *Ethical Theory and Moral Practice*, 17(2), 285-297.
- FAQs. (n.d.). *Thematic analysis*. <https://www.thematicanalysis.net/faqs/>
- Fazekas, K. (2016). Special Relativity, Multiple B-series, and the Passage of Time. *American Philosophical Quarterly*, 215-229.
- Feldman, G., Chandrashekar, S. P., & Wong, K. F. E. (2016). The freedom to excel: Belief in free will predicts better academic performance. *Personality and Individual Differences*, 90, 377-383.

- Feldman, G., Farh, J. L., & Wong, K. F. E. (2018). Agency beliefs over time and across cultures: Free will beliefs predict higher job satisfaction. *Personality and Social Psychology Bulletin*, 44(3), 304-317.
- 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. *Consciousness and Cognition*, 18(1), 342-350.
- Feltz, A., & Cokely, E. T. (2013). Predicting philosophical disagreement. *Philosophy Compass*, 8(10), 978-989.
- Feltz, A., Cokely, E. T. (2008). *The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions* [paper presentation]. In B.C. Love, K. McRae & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1771-1776). Austin, TX: Cognitive Science Society.
- Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural compatibilism versus natural incompatibilism: Back to the drawing board. *Mind & Language*, 24(1), 1-23.
- Fessler, D. M., Barrett, H. C., Kanovsky, M., Stich, S., Holbrook, C., Henrich, J., Bolyanatz, A. H., Gervais, M. M., Gurven, M., Kushnick, G., Pisor, A. C., von Rueden, C., Laurence, S. (2015). Moral parochialism and contextual contingency across seven societies. *Proceedings of the Royal Society B*, 282(1813), 20150907. <https://doi.org/10.1098/rspb.2015.0907>
- Fessler, D. M., Holbrook, C., Kanovsky, M., Barrett, H. C., Bolyanatz, A. H., Gervais, M. M., Gurven, M., Henrich, J., Kushnick, G., Pisor, A. C., Stich, S., von Rueden, C., & Laurence, S. (2016). Moral parochialism misunderstood: a reply to Piazza and Sousa. *Proceedings of the Royal Society B*, 283(1823), 20152628. <https://doi.org/10.1098/rspb.2015.2628>
- Findley, M. G., Kikuta, K., & Denly, M. (2021). External validity. *Annual Review of Political Science*, 24, 365-393.
- Finlay, S. (2008). The error in the error theory. *Australasian Journal of Philosophy*, 86(3), 347-369.
- Finlay, S. (2011). Errors upon errors: A reply to Joyce. *Australasian Journal of Philosophy*, 89(3), 535-547.
- Firth, R. (1952). Ethical absolutism and the ideal observer. *Philosophy and Phenomenological Research*, 12(3), 317-345.
- Fischer, E., & Curtis, M. (Eds.). (2019). *Methodological advances in experimental philosophy*. New York, NY: Bloomsbury Academic.
- Fischhoff, B. (1981). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty* (pp. 331-339). New York, NY: Cambridge University Press.

- Fisher, M., Knobe, J., Strickland, B., & Keil, F. C. (2017). The influence of social interaction on intuitions of objectivity and subjectivity. *Cognitive Science*, 41(4), 1119-1134.
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456-465.
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 370-378.
- Foot, P. (1978a). Moral arguments. In Foot (Ed.), *Virtues and vices and other essays in moral philosophy* (pp. 96-109). Oxford, UK: Basil Blackwell.
- Foot, P. (1978b). Moral beliefs. In Foot (Ed.), *Virtues and vices and other essays in moral philosophy* (pp. 110-131). Oxford, UK: Basil Blackwell.
- Formosa, P. (2013). Is Kant a moral constructivist or a moral realist?. *European Journal of Philosophy*, 21(2), 170-196.
- Forsyth, D. R. (1980). A taxonomy of ethical ideologies. *Journal of Personality and Social Psychology*, 39(1), 175-184.
- Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, 108(3), 819-824.
- Fuqua, J. (2021). Ethical Mooreanism. *Synthese*, 199(3), 6943-6965.
- Gabennesch, H. (1990). The perception of social conventionality by children and adults. *Child Development*, 61(6), 2047-2059.
- Gaesser, B., Campbell, D., & Young, L. (2022). Moral psychology from the lab to the wild: Relief registries as a paradigm for studying real-world altruism. PsyArXiv. 10.31234/osf.io/2xfsz
- Galef, J. (2021). *The scout mindset: Why some people see things clearly and others don't*. Canada: Penguin.
- Gendler, T. S., & Liao, S. Y. (2016). The problem of imaginative resistance. In N. Carroll & J. Gibson (Eds.), *The Routledge companion to philosophy of literature* (pp. 405-418). New York, NY: Routledge.
- Gibbard, A (2003). *Thinking how to live*. Cambridge, MA: Harvard University Press.
- Gill, M. B. (2008). Metaethical variability, incoherence, and error. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The cognitive science of morality* (Vol. 2, pp. 387-402). Cambridge, MA: MIT Press.
- Gill, M. B. (2009). Indeterminacy and variability in meta-ethics. *Philosophical Studies*, 145(2), 215-234.

- Goh, C.M.C (2017). English as lingua franca: Singapore's common tongue. In O.S. Tan & E.L. Low (Eds.), *Lee Kuan Yew's educational legacy: The challenge of success* (pp. 57-67). Singapore: Springer.
- Goldbeck-Wood, S. (1999). Evidence on peer review—scientific quality control or smokescreen?. *BMJ*, 318(7175), 44-45.
- Goldman, A. I. (2007). Philosophical intuitions: Their target, their source, and their epistemic status. *Grazer Philosophische Studien*, 74(1), 1-26. https://doi.org/10.1163/9789401204651_002
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, 106(3), 1339-1366.
- Goodwin, G. P., & Darley, J. M. (2010). The perceived objectivity of ethical beliefs: Psychological findings and implications for public policy. *Review of Philosophy and Psychology*, 1(2), 161-188.
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others?. *Journal of Experimental Social Psychology*, 48(1), 250-256.
- Google Ngram Viewer (2019a). “ moral realism ”. Google Books Ngram Viewer. https://books.google.com/ngrams/graph?content=%22+moral+realism+%22&year_start=1800&year_end=2019&corpus=26&smoothing=3&case_insensitive=true
- Google Ngram Viewer (2019b). “ moral relativism ”. Google Books Ngram Viewer. https://books.google.com/ngrams/graph?content=%22+moral+relativism+%22&year_start=1800&year_end=2019&corpus=26&smoothing=3&case_insensitive=true
- Gopnik, A., & Schwitzgebel, E. (1998). Whose concepts are they, anyway? The role of philosophical intuition in empirical psychology. In M. R. DePaul & W. Ramsey (Eds.), *Rethinking intuition: The psychology of intuition and its role in philosophy inquiry* (pp. 75-91). Lanham, MD: Rowman & Littlefield.
- Gordon, D. S., & Lea, S. E. (2016). Who punishes? The status of the punishers affects the perceived success of, and indirect benefits from, “moralistic” punishment. *Evolutionary Psychology*, 14(3), 1-14. <https://doi.org/10.1177/1474704916658042>
- Gowans, C. (2021). Moral relativism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 ed.). <https://plato.stanford.edu/archives/spr2021/entries/moral-relativism/>
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029-1046.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 55-130). San Diego, CA: Academic Press.

- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, 8, 125-130.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366-385.
- Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143(4), 1600-1615.
- Greco, J. (2014). Common Sense in Thomas Reid. *Canadian Journal of Philosophy* 41(Sup. 1): 142-155.
- Greene, J. (2003). From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology?. *Nature Reviews Neuroscience*, 4(10), 846-850.
- Greene, J. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*. New York, NY: Penguin.
- Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *Ethics*, 124(4), 695-726.
- Greene, J.D. (2008). The secret joke in Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology, volume 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35-80). Cambridge, MA: MIT Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics III: Speech acts* (pp. 41-58). New York: Academic Press.
- Grimm, P. (2010). Social desirability bias. In J. Sheth & N. Malhotra (Eds.), *Wiley international encyclopedia of marketing*. <https://doi.org/10.1002/9781444316568.wiem02057>
- Gruszczyński, R. (2020). Parts of falling objects: Galileo's thought experiment in mereological setting. *Erkenntnis*, 1-22. <https://doi.org/10.1007/s10670-020-00263-y>
- Guerin, B., Leugi, G. B., & Thain, A. (2018). Attempting to overcome problems shared by both qualitative and quantitative methodologies: Two hybrid procedures to encourage diverse research. *The Australian Community Psychologist*, 29(2), 74-90.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, 108(4), 814-834.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4), 55-66.

- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog?. *Journal of Personality and Social Psychology*, 65(4), 613-628.
- Hawley, P. (2008). Moral absolutism defended. *The Journal of Philosophy*, 105(5), 273-275.
- Hawley, P. H., Stump, K. N., & Ratliff, J. (2010). Sidestepping the jingle fallacy: Bullying, aggression, and the importance of knowing the difference. In D. L. Espelage & S. M. Swearer (Eds.), *Bullying in North American schools* (pp. 121-136). New York, NY: Routledge.
- Heiphetz, L., & Young, L. L. (2017). Can only one person be right? The development of objectivism and social preferences regarding widely shared and controversial moral beliefs. *Cognition*, 167, 78-90.
- Heiphetz, L., Spelke, E. S., Harris, P. L., & Banaji, M. R. (2013). The development of reasoning about beliefs: Fact, preference, and ideology. *Journal of Experimental Social Psychology*, 49(3), 559-565.
- Heiphetz, L., Strohminger, N., & Young, L. L. (2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science*, 41(3), 744-767.
- Heiphetz, L., Strohminger, N., Gelman, S. A., & Young, L. L. (2018). Who am I? The role of moral beliefs in children's and adults' understanding of identity. *Journal of Experimental Social Psychology*, 78, 210-219.
- Helwig, C. C. (2006). Rights, civil liberties, and democracy across cultures. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 185-210). Mahwah, NJ: Lawrence Erlbaum Associates.
- Henne, P., Chituc, V., De Brigard, F., & Sinnott-Armstrong, W. (2016). An empirical refutation of 'ought' implies 'can'. *Analysis*, 76(3), 283-290.
- Henne, P., Semler, J., Chituc, V., De Brigard, F., & Sinnott-Armstrong, W. (2019). Against some recent arguments for 'ought' implies 'can': Reasons, deliberation, trying, and furniture. *Philosophia*, 47(1), 131-139.
- Henrich, J. (2022). *Q&A on WEIRD*. The WEIRDist people in the world. <https://weirdpeople.fas.harvard.edu/qa-weird>
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q., & Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795-815.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, 33(2-3), 61-83.

- Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, 5(2), 205-221.
- Hofer, H. & Mills, G. (1998). *Hyaenas : status survey and conservation action plan*. International Union for Conservation of Nature (IUCN).
<https://policycommons.net/artifacts/1372774/hyaenas/1986984/>
- Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.
- Hofstede, G. H. (n.d.). *The 6-D model of national culture*. <https://geerthofstede.com/culture-geerthofstede-gert-jan-hofstede/6d-model-of-national-culture/>
- Hok, H., Martin, A., Trail, Z., & Shaw, A. (2020). When children treat condemnation as a signal: The costs and benefits of condemnation. *Child Development*, 91(5), 1439-1455.
- Holleman, G. A., Hooge, I. T., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, 11, Article 721.
<https://doi.org/10.3389/fpsyg.2020.00721>
- Hollos, M., Leis, P. E., & Turiel, E. (1986). Social reasoning in Ijo children and adolescents in Nigerian communities. *Journal of Cross-Cultural Psychology*, 17(3), 352-374.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.
- Hopster, J. (2019). The meta-ethical significance of experiments about folk moral objectivism. *Philosophical Psychology*, 32(6), 831-852.
- Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology*, 97(6), 963-976.
- Horvath, J., & Grundmann, T. (Eds.). (2012). *Experimental philosophy and its critics*. London, UK: Routledge.
- Horvath, J., & Koch, S. (2021). Experimental philosophy and the method of cases. *Philosophy Compass*, 16(1), e12716. <https://doi.org/10.1111/phc3.12716>
- Horwich, P. (2012). *Wittgenstein's metaphilosophy*. Oxford, UK: Oxford University Press.
- Hsieh, N. S. & Andersson H. (2021). Incommensurable values. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2021 ed.).
<https://plato.stanford.edu/archives/fall2021/entries/value-incommensurable/>

- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99-114.
- Huebner, B., Lee, J. J., & Hauser, M. D. (2010). The moral-conventional distinction in mature moral competence. *Journal of Cognition and Culture*, 10(1-2), 1-26.
- Huebner, B., Lee, J. J., & Hauser, M. D. (2010). The moral-conventional distinction in mature moral competence. *Journal of Cognition and Culture*, 10(1-2), 1-26.
- Huemer, M. (2007). Compassionate phenomenal conservatism. *Philosophy and Phenomenological Research*, 74(1), 30-55.
- Huff, G., & Huff, G. (2020). The Second World War Japanese Occupation of Singapore. *Journal of Southeast Asian Studies*, 51(1-2), 243-270.
- Hughes, G. D. (2009). The impact of incorrect responses to reverse-coded survey items. *Research in the Schools*, 16(2), 78-88.
- Hunt, S. D., Sparkman Jr, R. D., & Wilcox, J. B. (1982). The pretest in survey research: Issues and preliminary findings. *Journal of Marketing Research*, 19(2), 269-273.
- Hursthouse, R. (2002). *On virtue ethics*. New York, NY: Oxford University Press.
- Hursthouse, R. (2016). Virtue ethics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2018 ed.). <https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/>
- Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, 3(2), 166-184.
- Iurino, K., & Saucier, G. (2020). Testing measurement invariance of the Moral Foundations Questionnaire across 27 countries. *Assessment*, 27(2), 365-372.
- James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt and Company.
- Jamovi* (Version 2.3.21) [Computer Software]. <https://www.jamovi.org>
- Janooff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, 96(3), 521-537.
- JASP Team* (Version 0.17). JASP (Version 0.17) <https://jasp-stats.org/>
- Jebari, J., & Huebner, B. (2018). From objectivized morality to objective morality. *Behavioral and Brain Sciences*, 41. 10.1017/S0140525X18000109

- Jian, Z., Zhang, W., Tian, L., Fan, W., & Zhong, Y. (2019). Self-deception reduces cognitive load: The role of involuntary conscious memory impairment. *Frontiers in Psychology*, 10, Article 1718. <https://doi.org/10.3389/fpsyg.2019.01718>
- Johansson, J., & Olson, J. (2015). Against pluralism in metaethics. In C. Daly (Ed.), *The Palgrave handbook of philosophical methods* (pp. 593-609). Hampshire, UK: Palgrave Macmillan.
- Johansson-Stenman, O. (2012). Are most people consequentialists?. *Economics Letters*, 115(2), 225-228.
- Johnson, J. M., Bristow, D. N., & Schneider, K. C. (2004). Did you not understand the question or not? An investigation of negatively worded questions in survey research. *Journal of Applied Business Research (JABR)*, 20(1), 75-86.
- Johnson, S. G., Rodrigues, M., & Tuckett, D. (2021). Moral tribalism and its discontents: How intuitive theories of ethics shape consumers' deference to experts. *Journal of Behavioral Decision Making*, 34(1), 47-65.
- Jordan, J. J., & Kteily, N. S. (2022). *Reputation fuels moralistic punishment that people judge to be questionably merited*. PsyArXiv. 10.31234/osf.io/97nhj
- Jordan, J. J., & Rand, D. G. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology*, 118(1), 57-88.
- Joyce, R. (2001). *The myth of morality*. Cambridge, UK: Cambridge University Press.
- Joyce, R. (2006). *The evolution of morality*. Cambridge, MA: MIT Press.
- Joyce, R. (2011a). Moral fictionalism. *Philosophy Now*, 82, 14-17.
- Joyce, R. (2011b). The error in 'The error in the error theory'. *Australasian Journal of Philosophy*, 89(3), 519-534.
- Joyce, R. (2012). Enough with the errors! A final reply to Finlay. [Unpublished manuscript].
- Joyce, R. (2015). Moral objectivity and moral relativism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy archive*. <https://stanford.library.sydney.edu.au/archives/spr2017/entries/moral-anti-realism/moral-objectivity-relativism.html>
- Joyce, R. (2019). Moral fictionalism: How to have your cake and eat it too. In R. Garner & R. Joyce (Eds.), *The end of morality: Taking moral abolitionism seriously* (pp. 150-166) New York, NY: Routledge.
- Joyce, R. (2020). Moral and epistemic normativity: The guilty and the innocent. In C. Cowie & R. Rowland (Eds.), *Companions in guilt arguments in metaethics* (pp. 53-72). Abingdon, UK: Routledge.

- Joyce, R. (2021). Moral anti-realism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2021 ed.). <https://plato.stanford.edu/entries/moral-anti-realism/>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54-69.
- Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., Rennie, D., & PEER Investigators. (1998). Does masking author identity improve peer review quality?: A randomized controlled trial. *JAMA*, 280(3), 240-242.
- Kagan, S. (1997). *Normative ethics. Dimensions of philosophy series*. Boulder, CO: Westview Press.
- Kahane, G. (2013). Must metaethical realism make a semantic claim?. *Journal of Moral Philosophy*, 10(2), 148-178.
- Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, 125(2), 131-164.
- Kahlke, R. M. (2014). Generic qualitative approaches: Pitfalls and benefits of methodological mixology. *International Journal of Qualitative Methods*, 13(1), 37-52.
- Kalderon, M. E. (2005). *Moral fictionalism*. New York, NY: Oxford University Press.
- Kallestrup, J. (2013). *Semantic externalism*. London, UK: Routledge.
- Kelly, D., & Stich, S. (2007). Two theories about the cognitive architecture underlying morality. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 348-366). New York, NY: Oxford University Press.
- Kelly, D., Stich, S., Haley, K. J., Eng, S. J., & Fessler, D. M. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22(2), 117-131.
- Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer review in scientific publications: benefits, critiques, & a survival guide. *EJIFCC*, 25(3), 227-243.
- Kern, F. G., & Gleditsch, K. S. (2017). Exploring pre-registration and pre-analysis plans for qualitative inference. *ResearchGate*.
https://www.researchgate.net/publication/319141144_Exploring_Pre-registration_and_Pre-analysis_Plans_for_Qualitative_Inference
- Khoo, J., & Knobe, J. (2018). Moral disagreement and moral semantics. *Noûs*, 52(1), 109-143.

- Killen, M., McGlothlin, H., & Lee-Kim, J. (2002). Between individuals and culture: Individuals' evaluations of exclusion from social groups. In H. Keller, Y. Poortinga, & A. Schölmerich (Eds.), *Between biology and culture: Perspectives on ontogenetic development* (pp. 159-190). Cambridge, UK: Cambridge University Press.
- Kim, M., & Yuan, Y. (2015). No cross-cultural differences in the Gettier car case intuition: A replication study of Weinberg et al. 2001. *Episteme*, 12(3), 355-361.
- King James Version Bible*. (2022). King James Bible Online.
<https://www.kingjamesbibleonline.org/Romans-Chapter-2/>
- Kirchin, S. (Host). (2022, June 26). PGS - Metaethics - Part ONE - Moral realism. In *Philosophy gets schooled*. <https://www.podbean.com/ew/pb-riik3-125c463>
- Kirwin, C. (2021). Value realism and idiosyncrasy. Unpublished manuscript. Retrieved from https://elucidations.vercel.app/value_realism_and_idiosyncrasy.pdf
- Kissinger-Knox, A., Aragon, P., & Mizrahi, M. (2018). Ought implies can,” framing effects, and “empirical refutations. *Philosophia*, 46(1), 165-182.
- Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32, 385-418.
- Knobe, J. & S. Nichols (2008). An experimental philosophy manifesto. In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 3-14). Oxford, UK: Oxford University Press.
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(3), 190-194.
- Knobe, J. (2007). Experimental philosophy. *Philosophy Compass*, 2(1), 81-92.
- Knobe, J. (2016). Experimental philosophy is cognitive science. In J. Sytsma & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 37-52) Malden, MA: Wiley-Blackwell.
- Knobe, J., & Nichols, S. (2017). Experimental philosophy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2017 ed.).
<https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/>
- Knobe, J., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H., & Sommers, T. (2012). Experimental philosophy. *Annual Review of Psychology*, 63, 81-99.
- Knowles, E. S., & Nathan, K. T. (1997). Acquiescent responding in self-reports: cognitive style or social concern?. *Journal of Research in Personality*, 31(2), 293-301.
- Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*, 12(2), 93-120.

- Komatsu, L. K., & Galotti, K. M. (1986). Children's reasoning about social, physical, and logical irregularities: A look at two worlds. *Child Development*, 57(2), 413-420.
- Koonin, E. V. (2009). The Origin at 150: is a new evolutionary synthesis in sight?. *Trends in Genetics*, 25(11), 473-475.
- Koons, R. C. (2022). A dynamic B theory of time. In C. J. Austin, A. Marmodoro, & A. Roselli (Eds.), *Powers, time, and free will* (pp. 69-82). Cham, CH: Springer.
- Korman, D. and Carmichael, C. (2017). What do the folk think about composition and does it matter?. In D. Rose (Ed.), *Experimental metaphysics* (pp. 187-206). London, UK: Bloomsbury Academic.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.
- Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28(2), 75-84.
- Kvale, S. (1994). Ten standard objections to qualitative research interviews. *Journal of Phenomenological Psychology*, 25(2), 147-173.
- Lassiter, D. (2008). Semantic externalism, language variation, and sociolinguistic accommodation. *Mind & Language*, 23(5), 607-633.
- Lavrakas, P. J. (2008). *Encyclopedia of survey research methods* (Vol. 2). Thousand Oaks, CA: Sage.
- Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *Sage Open*, 6(1), 2158244016636433.
- Levine, S., Rottman, J., Davis, T., O'Neill, E., Stich, S., & Machery, E. (2021). Religious affiliation and conceptions of the moral domain. *Social Cognition*, 39(1), 139-165.
- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Levy, N. (2003). Descriptive relativism: Assessing the evidence. *Journal of Value Inquiry*, 37, 165-177.
- Lewis, D. (1983). *Philosophical papers: Volume I*. New York, NY: Oxford University Press.
- Lewkowicz, D. J. (2001). The concept of ecological validity: What are its limitations and is it bad to be invalid?. *Infancy*, 2(4), 437-450.
- Li, J., Zhao, Y., Lin, L., Chen, J., & Wang, S. (2018). The freedom to persist: Belief in free will predicts perseverance for long-term goals among Chinese adolescents. *Personality and Individual Differences*, 121, 7-10.

- Liddle, J. R., Bush, L. S., & Shackelford, T. K. (2011). An introduction to evolutionary psychology and its application to suicide terrorism. *Behavioral Sciences of Terrorism and Political Aggression*, 3(3), 176-197.
- Lilienfeld, S. O., & Strother, A. N. (2020). Psychological measurement and the replication crisis: Four sacred cows. *Canadian Psychology/Psychologie Canadienne*, 61(4), 281-288.
- Lillehammer, H. (2007). Davidson on value and objectivity. *Dialectica*, 61(2), 203-217.
- Lim, D., & Chen, J. (2018). Is compatibilism intuitive?. *Philosophical Psychology*, 31(6), 878-897.
- Lindauer, M., & Southwood, N. (2021). How to cancel the Knobe effect: the role of sufficiently strong moral censure. *American Philosophical Quarterly*, 58(2), 181-186.
- Linnebo, O. (2018). Platonism in the philosophy of mathematics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2018 ed.). <https://plato.stanford.edu/entries/platonism-mathematics/>
- Lockhart, K. L., Abrahams, B., & Osherson, D. N. (1977). Children's understanding of uniformity in the environment. *Child Development*, 48(4), 1521-1531.
- Loeb, D. (2003). Gastronomic realism—A cautionary tale. *Journal of Theoretical and Philosophical Psychology*, 23(1), 30-49.
- Loeb, D. (2008). Moral incoherentism: How to pull a metaphysical rabbit out of a semantic hat. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The cognitive science of morality* (Vol. 2, pp. 355-386). Cambridge, MA: MIT Press.
- Lopez, J., & Hillygus, D. S. (2018). Why so serious?: Survey trolls and misinformation. *Why so serious*. [Unpublished manuscript]. <http://dx.doi.org/10.2139/ssrn.3131087>
- Ludwig, K. (2007). The epistemology of thought experiments: First person versus third person approaches. *Midwest Studies in Philosophy*, 31, 128-159.
- Lycan, W. G. (2019). *On evidence in philosophy*. Oxford, UK: Oxford University Press.
- Machery, E. & Stich, S. (2022). The Moral/conventional distinction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022 ed.). <https://plato.stanford.edu/archives/sum2022/entries/moral-conventional/>
- Machery, E. (2017). *Philosophy within its proper bounds*. Oxford, UK: Oxford University Press.
- Machery, E. (2018). Morality: A historical invention. In K. J. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 259-265). New York, NY: The Guilford Press.

- Machery, E., & Mallon, R. (2010). Evolution of morality. In J. M. Doris & The Moral Psychology Research Group (Eds.). *The moral psychology handbook* (pp. 3-47). New York, NY: Oxford University Press.
- Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition*, 92(3), B1-B12.
- Machery, E., & Stich, S. (2022). The moral/conventional distinction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022 ed.).
<https://plato.stanford.edu/archives/sum2022/entries/moral-conventional/>
- MacIntyre, A. (1957). What morality is not. *Philosophy*, 32(123), 325-335.
- Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. London, UK: Penguin.
- Madans, J., Miller, K., Maitland, A., & Willis, G. B. (2011). *Question evaluation methods: Contributing to the science of data quality*. Hoboken, NJ: John Wiley & Sons.
- Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, 6, Article 2, 1-7. <https://doi.org/10.3389/fpsyg.2015.00002>
- Mainwood, P. (2020). Richard Feynman said, "I think I can safely say that nobody understands quantum mechanics." What is your opinion or perspective about this? What did he mean?. *Quora*. <https://www.quora.com/Richard-Feynman-said-I-think-I-can-safely-say-that-nobody-understands-quantum-mechanics-What-is-your-opinion-or-perspective-about-this-What-did-he-mean/answer/Paul-Mainwood>
- Mallon, R., Machery, E., Nichols, S., & Stich, S. (2009). Against arguments from reference. *Philosophy and Phenomenological Research*, 79(2), 332-356.
- Martí, G. (2009). Against semantic multi-culturalism. *Analysis*, 69(1), 42-48.
- Martin, R. A. (2007). *The psychology of humor: An integrative approach*. Burlington, MA: Elsevier Academic Press.
- Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, 15(2), 51-69.
- Mavrogenis, A. F., Quaile, A., & Scarlat, M. M. (2020). The good, the bad and the rude peer-review. *International Orthopaedics*, 44(3), 413-415.
- Maxwell, N. (2006). Special relativity, time, probabilism and ultimate reality. *Philosophy and Foundations of Physics*, 1, 229-245.

- May, J., & Holton, R. (2012). What in the world is weakness of will?. *Philosophical Studies*, 157(3), 341-360.
- McCambridge, J., De Bruin, M., & Witton, J. (2012). The effects of demand characteristics on research participant behaviours in non-laboratory settings: a systematic review. *PloS One*, 7(6), e39116. <https://doi.org/10.1371/journal.pone.0039116>
- McCarthy, D. (2016). Probability in ethics. In A. Hájek & C. Hitchcock (Eds.), *The Oxford handbook of philosophy and probability* (pp. 705-737). Oxford, UK: Oxford University Press.
- McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science*, 21(8), 1141-1149.
- McGraw, A. P., Schiro, J. L., Fernbach, P. M. (2015). Not a problem: A downside of humorous appeals. *Journal of Marketing Behavior*, 1(2), 187-208.
- McGraw, A. P., Williams, L. E., & Warren, C. (2014). The rise and fall of humor: Psychological distance modulates humorous responses to tragedy. *Social Psychological and Personality Science*, 5(5), 566-572.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276-282.
- McNaughton, D. (1988). *Moral vision: An introduction to ethics*. Oxford, UK: Blackwell.
- McPherson Frantz, C. (2006). I AM being fair: The bias blind spot as a stumbling block to seeing both sides. *Basic and Applied Social Psychology*, 28(2), 157-167.
- Medinnus, G. R. (1962). Objective responsibility in children: A comparison with the Piaget data. *The Journal of Genetic Psychology*, 101(1), 127-133.
- Meisters, J., Hoffmann, A., & Musch, J. (2020). Can detailed instructions and comprehension checks increase the validity of crosswise model estimates?. *PloS One*, 15(6), e0235403. <https://doi.org/10.1371/journal.pone.0235403>
- Menold, N. (2020). Double barreled questions: An analysis of the similarity of elements and effects on measurement quality. *Journal of Official Statistics*, 36(4), 855-886.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57-74.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Merritt, J. (2016, March 25). The death of moral relativism. *The Atlantic*. <http://www.theatlantic.com/politics/archive/2016/03/the-death-of-moral-relativism/475221/>

- Mihailov, E. (2022). Measuring Impartial Beneficence: A Kantian Perspective on the Oxford Utilitarianism Scale. *Review of Philosophy and Psychology*, 1-16. <https://doi.org/10.1007/s13164-021-00600-2>
- Mijović-Prelec, D., & Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538), 227-240.
- Miller, C. (2009). The conditions of moral realism. *Journal of Philosophical Research*, 34, 123-155.
- Miller, J. G., & Bersoff, D. M. (1988). When do American children and adults reason in social conventional terms?. *Developmental Psychology*, 24(3), 366-375.
- Milo, R. (1995). Contractarian constructivism. *The Journal of Philosophy*, 92(4), 181-204.
- Mizrahi, M. (2015). Ought, can, and presupposition: An experimental study. *Method: Analytic Perspectives*, 4(6), 232-243.
- Molouki, S., & Bartels, D. M. (2017). Personal change and the continuity of the self. *Cognitive Psychology*, 93, 1-17.
- Monroe, A. E., & Malle, B. F. (2015). Free will without metaphysics. In A. R. Mele (Ed.), *Surrounding free will* (pp. 25-48). Oxford, UK: Oxford University Press.
- Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), 760-775.
- Moody-Adams, M. M. (2009). *Fieldwork in familiar places: Morality, culture, and philosophy*. Cambridge, MA: Harvard University Press.
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38(4), 379-387.
- Moomal, Z., & Henzi, S. P. (2000). The evolutionary psychology of deception and self-deception. *South African Journal of Psychology*, 30(3), 45-51.
- Moral parochialism misunderstood: a reply to Piazza and Sousa. *Proceedings of the Royal Society B*, 283(1823), 20152628. <https://doi.org/10.1098/rspb.2015.2628>
- Morreall, J. (2009). *Comic relief: A comprehensive philosophy of humor*. Malden, MA: Wiley-Blackwell.
- Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(2), 120-123.
- Moss, D. (2017). Experimental philosophy, folk metaethics and qualitative methods. *Teorema: Revista Internacional de Filosofía* 36(3), 185-203.

- Moss, D., & Bush, L. S. (2021). Measuring metaaesthetics: Challenges and ways forward. *New Ideas in Psychology*, 62, 100866. <https://doi.org/10.1016/j.newideapsych.2021.100866>
- Moss, D., Montealegre, A., Bush, L.S., Caviola, L., & Pizarro, D. (n.d.). *The reputational implications of relativism and objectivism* [Unpublished manuscript].
- Munier, B. R. (1988). A guide to decision-making under uncertainty. In B. R. Munier (Ed.), *Risk, decision and rationality* (pp. 1-34). Dordrecht, NL: D. Reidel Publishing Company.
- Murray, D., & Nahmias, E. (2014). Explaining away incompatibilist intuitions. *Philosophy and Phenomenological Research*, 88(2), 434-467.
- Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, 31(6), 678-701.
- Myers, R. H. (2004). Finding value in Davidson. *Canadian Journal of Philosophy*, 34(1), 107-136.
- Nadelhoffer, T., & Monroe, A. (2022). Experimental philosophy of free will and moral responsibility: An introduction. In T. Nadelhoffer & A. Monroe (Eds.), *Advances in experimental philosophy of free will and responsibility* (pp. 1-11). London, UK: Bloomsbury Academic.
- Nadelhoffer, T., & Nahmias, E. (2007). The past and future of experimental philosophy. *Philosophical Explorations*, 10(2), 123-149.
- Nadelhoffer, T., Murray, S., & Murry, E. (2021). Intuitions About Free Will and the Failure to Comprehend Determinism. *Erkenntnis*, 1-22. <https://doi.org/10.1007/s10670-021-00465-y>
- Nadelhoffer, T., Rose, D., Buckwalter, W., & Nichols, S. (2020). Natural compatibilism, indeterminism, and intrusive metaphysics. *Cognitive Science*, 44(8), e12873. <https://doi.org/10.1111/cogs.12873>
- Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. T. (2014). The free will inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition*, 25, 27-41.
- Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay denial of knowledge for justified true beliefs. *Cognition*, 129(3), 652-661.
- Nahmias, E. (2006). Folk fears about freedom and responsibility: Determinism vs. reductionism. *Journal of Cognition and Culture*, 6(1-2), 215-237.
- Nahmias, E. (2011). Intuitions about free will, determinism, and bypassing. In R. Kane (Ed.), *The Oxford handbook of free will* (2nd ed., pp. 555-576). New York, NY: Oxford University Press.

- Nahmias, E., & Murray, D. (2010). Experimental philosophy on free will: An error theory for incompatibilist intuitions. In J. Aguilar, A. Buckareff, & K. Frankish (Eds.), *New waves in philosophy of action* (pp. 189-216). New York, NY: Palgrave-Macmillan.
- Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive?. *Philosophy and Phenomenological Research*, 73(1), 28-53.
- Nahmias, E., Morris, S., & Nadelhoffer, T. (2004). The phenomenology of free will. *Journal of Consciousness Studies*, 11(7-8), 162-179.
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, 18(5), 561-584.
- Navarro-Plaza, M. C., Gomis-Pont, S., Mirete-Fructuoso, M., Molina-Rodríguez, S., & Martínez, L. M. (2020). To what extent is trading ecological validity for experimental control justified in moral research?. PsyArXiv. <https://doi.org/10.31234/osf.io/eykcd>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, 69, 511-534.
- Nevins, A., Pesetsky, D., & Rodrigues, C. (2009). Pirahã exceptionality: A reassessment. *Language* 85(2), 355-404.
- Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, 37(1), 1-19.
- Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, 135(2), 151-166.
- Nichols, S. (2004). After objectivity: An empirical study of moral judgment. *Philosophical Psychology*, 17(1), 3-26.
- Nichols, S. (2008). How can psychology contribute to the free will debate? In K. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 10-31). Oxford, UK: Oxford University Press.
- Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science*, 331(6023), 1401-1403.
- Nichols, S. (2012). The indeterminist intuition. *The Monist*, 95(2), 290-307.
- Nichols, S., & Bruno, M. (2010). Intuitions about personal identity: An empirical study. *Philosophical Psychology*, 23(3), 293-312.

- Nichols, S., & Folds-Bennett, T. (2003). Are children moral objectivists? Children's judgments about moral and response-dependent properties. *Cognition*, 90(2), B23-B32.
- Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, 41(4), 663-685.
- Nichols, S., Stich, S., & Weinberg, J. (2003). Metaskepticism: Meditations in ethno-epistemology. In S. Luper (Ed.), *The skeptics* (pp. 227-247), Burlington VT: Ashgate.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.
- Nisan, M. (1987). Moral norms and social conventions: A cross-cultural comparison. *Developmental Psychology*, 23(5), 719-725.
- Nisan, M. (1988). A story of a pot: or A cross-cultural comparison of basic moral evaluations: A response to the critique by Turiel, Nucci, and Smetana (1988). *Developmental Psychology*, 24(1), 144-146.
- Nisbett, R. (2004). *The geography of thought: How Asians and Westerners think differently... and why*. New York, NY: Simon & Schuster.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231-259.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*, 108(2), 291-310.
- Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, 26(5), 653-684.
- Nucci, L. & Nucci, M. (1982). Children's social interactions in the context of moral and conventional transgressions. *Child Development*, 53(2), 403-412.
- Nucci, L. & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, 49(2), 400-407.
- Nucci, L. P. (2001). *Education in the moral domain*. Cambridge, UK: Cambridge University Press.
- Nucci, L. P., Turiel, E., & Encarnacion-Gawrych, G. (1983). Children's social interactions and social concepts: Analyses of morality and convention in the Virgin Islands. *Journal of Cross-Cultural Psychology*, 14(4), 469-487.

- Nucci, L., & Turiel, E. (1993). God's word, religious rules, and their relation to Christian and Jewish children's concepts of morality. *Child Development*, 64(5), 1475-1491.
- Nucci, L., Camino, C., & Sapiro, C. M. (1996). Social class effects on northeastern Brazilian children's conceptions of areas of personal choice and social regulation. *Child Development*, 67(3), 1223-1242.
- Office of International Religious Freedom. (2021, May 12). *2020 report on International Religious Freedom: Brazil*. U.S. Department of State. <https://www.state.gov/reports/2020-report-on-international-religious-freedom/brazil/>
- Olsen, W. 2004. Triangulation in social research: qualitative and quantitative methods can really be mixed. In M. Holborn (Ed.), *Developments in sociology: An annual review* (Vol. 20, pp. 103-121). Ormskirk, UK: Causeway Press.
- Olson, J. (2014). *Moral error theory: History, critique, defense*. Oxford, UK: Oxford University Press.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. 10.1126/science.aac4716
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776-783.
- Ort, A., & Fahr, A. (2020). The effectiveness of a positively vs. negatively valenced PSA against sexually transmitted diseases—Evidence from an experimental study. *SCM Studies in Communication and Media*, 9(3), 341-366.
- Oxford University Press. (2021). *Objective*, adj. and n. In OED online. www.oed.com/view/Entry/129634
- Paden, R. (1994). The natural history of student relativism. *Journal of Thought*, 29(2), 47-58.
- Parfit, D. (2011). *On what matters* (Vol. 2). Oxford, UK: Oxford University Press.
- Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience*, 23(10), 3162-3180.
- Paul, L. A. (2010). Temporal experience. *The Journal of Philosophy*, 107(7), 333-359.
- Peters, U. (2022). What is the function of confirmation bias?. *Erkenntnis*, 87, 1351-1376.
- Peterson, J. (2018). *12 Rules for Life: An Antidote to Chaos*. Toronto, ON: Random House.

- Pethokoukis, J. (2011). *Paul Ryan Speaks! Part two! Moral relativism, the EU debt crisis, inflation and more* American Enterprise Institute. <https://www.aei.org/economics/paul-ryan-speaks-part-two-moral-relativism-the-eu-debt-crisis-inflation-and-more/>
- Pfister, J. (2019). Classification of strategies for dealing with student relativism and the epistemic conceptual change strategy. *Teaching Philosophy*, 42(3), 221-246.
- Phillips, J., Nyholm, S., & Liao, S. Y. (2014). The good in happiness. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 1, 253-293). Oxford, UK: Oxford University Press.
- Piaget, J. (1997). *The moral judgment of the child* (M. Gabain, Trans.). New York, NY: Simon & Schuster. (Original work published 1932)
- Piazza, J., & Sousa, P. (2016). When injustice is at stake, moral judgements are not parochial. *Proceedings of the Royal Society B*, 283(1823), 20152037. <https://doi.org/10.1098/rspb.2015.2037>
- Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, 107(supplement_2), 8993-8999.
- Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). American Psychological Association.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- Pölzler, T. (2017). Revisiting folk moral realism. *Review of Philosophy and Psychology*, 8(2), 455-476.
- Pölzler, T. (2018a). Can the empirical sciences contribute to the moral realism/anti-realism debate?. *Synthese*, 195(11), 4907-4930.
- Pölzler, T. (2018b). How to measure moral realism. *Review of Philosophy and Psychology*, 9(3), 647-670.
- Pölzler, T. (2018c). *Moral reality and the empirical sciences*. New York, NY: Routledge.
- Pölzler, T. (2021). Insufficient effort responding in experimental philosophy. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 4, pp. 214-256). Oxford, UK: Oxford University Press.
- Pölzler, T., & Wright, J. C. (2019). Empirical research on folk moral objectivism. *Philosophy Compass*, 14(5), e12589. <https://doi.org/10.1111/phc3.12589>

- Pözlner, T., & Cole Wright, J. C. (2020a). An empirical argument against moral non-cognitivism. *Inquiry*, 1-29. <https://doi.org/10.1080/0020174X.2020.1798280>
- Pözlner, T., & Wright, J. C. (2020b). Anti-realist pluralism: A new approach to folk metaethics. *Review of Philosophy and Psychology*, 11(1), 53-82.
- Pözlner, T., Zijlstra, L., & Dijkstra, J. (2022). Moral progress, knowledge and error: Do people believe in moral objectivity?. *Philosophical Psychology*, 1-37.
- Population Density (n.d.) The world bank. <https://data.worldbank.org/indicator/EN.POP.DNST>
- Potter, N. (1997). The synthetic a priori proposition of Kant's ethical philosophy. In B. S. Byrd, J. C. Joerden, & J. Hruschka (Eds.), *Jahrbuch für Recht und Ethik / Annual Review of Law and Ethics* (Vol. 5, pp. 437-459). Berlin, DE: Duncker & Humblot.
- Povee, K., & Roberts, L. D. (2014). Qualitative research in psychology: Attitudes of psychology students and academic staff. *Australian Journal of Psychology*, 66(1), 28-37.
- Pravato, G. (2020). Indeterminacy and Normativity. *Erkenntnis*, 1-23. <https://doi.org/10.1007/s10670-020-00293-6>
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11(1), 37-43.
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28(3), 369-381.
- Pust, J. (2017). Intuitions. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 ed.). <https://plato.stanford.edu/archives/sum2019/entries/intuition/>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71-90.
- Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies* 3(9), 369-387.
- Quintelier, K. J., & Fessler, D. M. (2012). Varying versions of moral relativism: The philosophy and psychology of normative relativism. *Biology & Philosophy*, 27(1), 95-113.
- Quintelier, K. J., De Smet, D., & Fessler, D. M. (2013). The moral universalism-relativism debate. *KLESIS—Revue Philosophique*, 27, 211-262.

- Quintelier, K., De Smet, D., & Fessler, D. (2014). Agent versus appraiser moral relativism: an exploratory study. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 209-230). London, UK: Bloomsbury Academic.
- Raaijmakers, J. G., Schrijnemakers, J. M., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41(3), 416-426.
- Rabb, N., Han, A., Nebeker, L., & Winner, E. (2020). Expressivist to the core: Metaaesthetic subjectivism is stable and robust. *New Ideas in Psychology*, 57, 100760. <https://doi.org/10.1016/j.newideapsych.2019.100760>
- Rabkin, E. S. (1979). Metalinguistics and Science Fiction. *Critical Inquiry*, 6(1), 79-97.
- Radziwill, N. (2015, March 16). One-proportion Z test in R. *Quality and innovation*. <https://qualityandinnovation.com/2015/03/16/one-proportion-z-test-in-r/>
- Rahim, L. Z. (2010). *Singapore in the Malay world: Building and breaching regional bridges*. New York, NY: Routledge.
- Rahman, M. S. (2020). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language “testing and assessment” research: A literature review. *Journal of Education and Learning* 6(1), 102-112.
- Rai, T. S., & Holyoak, K. J. (2013). Exposure to moral relativism compromises moral behavior. *Journal of Experimental Social Psychology*, 49(6), 995-1001.
- Reudell, R. (2021, April 8). *Why moral relativism is dangerous*. A Philosopher’s Stone. <https://medium.com/the-philosophers-stone/why-moral-relativism-is-dangerous-9b1e5da62bd7>
- Reuter, K., & Brun, G. (2022) Empirical studies in truth and the project of re-engineering truth. *Pacific Philosophical Quarterly*. <https://doi.org/10.1111/papq.12370>
- Reynolds, N., Diamantopoulos, A., & Schlegelmilch, B. (1993). Pre-testing in questionnaire design: A review of the literature and suggestions for further research. *Market Research Society Journal*, 35(2), 1-11.
- Rich, M. C. (1979). Verbal reports on mental processes: Issues of accuracy and awareness. *Journal for the Theory of Social Behaviour*, 9(1), 29-37.
- Rose, D., & Nichols, S. (2019). From punishment to universalism. *Mind & Language*, 34(1), 59-72.

- Rose, D., & Schaffer, J. (2017). Folk mereology is teleological. In D. Rose (Ed.), *Experimental metaphysics* (pp. 135-186). London, UK: Bloomsbury Academic.
- Rose, D., Schaffer, J., & Tobia, K. (2020). Folk teleology drives persistence judgments. *Synthese*, 197(12), 5491-5509.
- Rosenthal, G. S. (1978). *The many faces of Judaism: Orthodox, Conservative, Reconstructionist, & Reform*. New York: Behrman House.
- Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition*, 112(1), 159-174.
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5(1), 2-14.
- Ryazanov, A. A., Knutzen, J., Rickless, S. C., Christenfeld, N. J., & Nelkin, D. K. (2018). Intuitive probabilities and the limitation of moral imagination. *Cognitive Science*, 42, 38-68.
- Ryskin, R. A., & Brown-Schmidt, S. (2014). Do adults show a curse of knowledge in false-belief reasoning? A robust estimate of the true effect size. *PLoS One*, 9(3), e92406. <https://doi.org/10.1371/journal.pone.0092406>
- Sarkissian, H. (2016). Aspects of Folk Morality. In J. Sytsma, & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 212-224). West Sussex, UK: Wiley-Blackwell.
- Sarkissian, H., & Phelan, M. (2019). Moral objectivism and a punishing God. *Journal of Experimental Social Psychology*, 80, 1-7.
- Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is belief in free will a cultural universal?. *Mind & Language*, 25(3), 346-358.
- Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind & Language*, 26(4), 482-505.
- Satris, S. A. (1986). Student relativism. *Teaching Philosophy*, 9(3), 193-205.
- Sauer, H. (2019). The argument from agreement: How universal values undermine moral realism. *Ratio*, 32(4), 339-352.
- Savitsky, K., & Gilovich, T. (2003). The illusion of transparency and the alleviation of speech anxiety. *Journal of Experimental Social Psychology*, 39(6), 618-625.
- Sayre-McCord, G. (2014). Metaethics. In E. D. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2014 ed.). <https://plato.stanford.edu/archives/sum2014/entries/metaethics>

- Sayre-McCord, G. (2015). Moral realism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 ed.). <https://plato.stanford.edu/archives/sum2021/entries/moral-realism/>
- Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford, UK: Oxford University Press.
- Schimmack, U. (2021). The validation crisis in psychology. *Meta-Psychology*, 5, 1-9. <https://doi.org/10.15626/MP.2019.1645>
- Schlagel, R. H. (1974). Contra Wittgenstein. *Philosophy and Phenomenological Research*, 34(4), 539-550.
- Schmidt, M. F., Gonzalez-Cabrera, I., & Tomasello, M. (2017). Children's developing metaethical judgments. *Journal of Experimental Child Psychology*, 164, 163-177.
- Schrödinger, E. (1935). Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*, 23(50), 844-849.
- Schulz, J. F., Bahrami-Rad, D., Beauchamp, J., & Henrich, J. (2018, June 22). *The origins of WEIRD psychology*. [Unpublished manuscript]. <https://doi.org/10.31234/osf.io/d6qhu>
- Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, 3(10), 1055-1061.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, 61(2), 195-202.
- Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, 6(1), 38-47.
- Semler, J., & Henne, P. (2019). Recent experimental work on “ought” implies “can”. *Philosophy Compass*, 14(9), e12619. <https://doi.org/10.1111/phc3.12619>
- Shafer-Landau, R. (2003). *Moral realism: A defense*. Oxford, UK: Clarendon Press.
- Shakespeare, W. (2020). *The Tragedy of Julius Caesar*. B. A. Mowat & P. Westerine (Eds.). Folger Shakespeare Library. (Original work published 1599)
- Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2016). The development of inequity aversion: Understanding when (and why) people give others the bigger piece of the pie. *Psychological Science*, 27(10), 1352-1359.

- Shoemaker, D. & Tobia, K. (2022). Personal identity. In M. Vargas & J. M. Doris (Eds.), *Oxford handbook of moral psychology* (543-563). Oxford, UK: Oxford University Press.
- Shweder, R. A. (1990). In defense of moral realism: Reply to Gabennesch. *Child Development*, 61(6), 2060-2067.
- Shweder, R. A., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1-83). Chicago, IL: University of Chicago Press.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119-169). New York, NY: Routledge.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.
- Sinclair, N. (2012). Moral realism, face-values and presumptions. *Analytic Philosophy*, 53(2), 158-179.
- Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 4-31). New York, NY: Psychology Press.
- Sinnott-Armstrong, W. (2009). Mixed-up meta-ethics. *Philosophical Issues* 19(1), 235-256.
- Sinnott-Armstrong, W., & Wheatley, T. (2012). The disunity of morality and why it matters to philosophy. *The Monist*, 95(3), 355-377.
- Sinnott-Armstrong, W., & Wheatley, T. (2014). Are moral judgments unified?. *Philosophical Psychology*, 27(4), 451-474.
- Skitka, L. J., Bauman, C. W. and Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength of something more?. *Journal of Personality and Social Psychology*, 88(6), 895-917.
- Smetana, J. (1993). Understanding of social rules. In M. Bennett (Ed.), *The child as psychologist: An introduction to the development of social cognition* (pp. 111-141). New York, NY: Guilford Press.
- Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, 52, 1333-1336.
- Smetana, J. G., & Braeges, J. L. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly*, 36(3), 329-346.

- Smetana, J. G., Kelly, M., & Twentyman, C. T. (1984). Abused, neglected, and nonmaltreated children's conceptions of moral and social-conventional transgressions. *Child Development*, 55(1), 277-287.
- Smith, M. (1994). *The moral problem*. Malden, MA: Blackwell.
- Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99(4), 178-182.
- Smith, R. (2010). Classical peer review: an empty gun. *Breast Cancer Research*, 12(4), 1-4.
- Smith, E. R., & Miller, F. D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review*, 85(4), 355-362.
- Solzhenitsyn, A. (2009). *Détente, democracy and dictatorship* (3rd ed.). New York, NY: Routledge.
- Sommers, T. (2010). Experimental philosophy and free will. *Philosophy Compass*, 5(2), 199-212.
- Song, M. J., Smetana, J. G., & Kim, S. Y. (1987). Korean children's conceptions of moral and conventional transgressions. *Developmental Psychology*, 23(4), 577-582.
- Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies*, 132(1), 99-107.
- Sosa, E. (2009). A defense of the use of intuitions in philosophy. In D. Murphy & M. Bishop (Eds.) *Stich and his critics* (pp. 101-112). West Sussex, UK: Wiley-Blackwell.
- Sosa, E. (2014). Intuitions: Their nature and probative value. In A. R. Booth & D. P. Rowbottom (Eds.) *Intuitions* (pp. 36-49). Oxford, UK: Oxford University Press.
- Sousa, P., Allard, A., Piazza, J., & Goodwin, G. P. (2021). Folk moral objectivism: The case of harmful actions. *Frontiers in Psychology*, 12, Article 2776. <https://doi.org/10.3389/fpsyg.2021.638515>
- Sperber, D., & Baumard, N. (2012). Moral reputation: An evolutionary and cognitive perspective. *Mind & Language*, 27(5), 495-518.
- Stanford, P. (2018). The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation. *Behavioral and Brain Sciences*, 41, E95. doi:10.1017/S0140525X17001911
- Starmans, C., & Bloom, P. (2018a). If you become evil, do you die?. *Trends in Cognitive Sciences*, 22(9), 740-741.
- Starmans, C., & Bloom, P. (2018b). Nothing personal: What psychologists get wrong about identity. *Trends in Cognitive Sciences*, 22(7), 566-568.

- Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, 124(3), 272-283.
- Sterelny, K., & Fraser, B. (2016). Evolution and moral realism. *British Journal for the Philosophy of Science*, 68(4), 981-1006.
- Stevenson, C. L. (1963). *Facts and values*. New Haven, CT: Yale University Press.
- Stevenson, C. L. (1963). *Facts and values: Studies in ethical analysis*. New Haven, CT: Yale University Press.
- Stich, S. (2018). The moral domain. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 547-555). New York, NY: Guilford Press.
- Stich, S. (2019). The quest for the boundaries of morality. In A. Zimmerman, K. Jones, & M. Timmons (Eds.), *The Routledge Handbook of Moral Epistemology* (pp. 15-38). New York, NY: Routledge.
- Stich, S., & Tobia, K. (2016). Experimental philosophy and the philosophical tradition. In J. Sytsma & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 5-21). Malden, MA: Wiley-Blackwell.
- Stojanovic, I., & McNally, L. (2022). Are moral predicates subjective? A corpus study. In D. Bordonaba (Ed.), *Experimental philosophy of language: Perspectives, methods, and prospects* [Unpublished manuscript]. Springer. 10.13140/RG.2.2.23636.88965
- Stoljar, D. Damjanovic, N. (2007). The deflationary theory of truth. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2010 ed.).
<https://stanford.library.sydney.edu.au/archives/spr2010/entries/truth-deflationary/>
- Stroebe, W., Gadenne, V., & Nijstad, B. A. (2018). Do our psychological laws apply only to college students?: External validity revisited. *Basic and Applied Social Psychology*, 40(6), 384-395.
- Strohming, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131(1), 159-171.
- Strohming, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, 26(9), 1469-1479.
- Strohming, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition*, 119(2), 295-300.
- Stuenkel, O. (2011). Identity and the concept of the West: the case of Brazil and India. *Revista Brasileira de Política Internacional*, 54(1), 178-195.
- Suárez Álvarez, J., Pedrosa, I., Lozano, L. M., García Cueto, E., Cuesta Izquierdo, M., & Muñiz Fernández, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158.

- Swain, S., Alexander, J., & Weinberg, J. M. (2008). The instability of philosophical intuitions: Running hot and cold on truetemp. *Philosophy and Phenomenological Research*, 76(1), 138-155.
- Swee-Hock, S. (2012). *The population of Singapore* (3rd ed.). Pasir Panjang, Singapore: ISEAS Publishing.
- Sytsma, J., Bishop, R., & Schwenkler, J. (2022). Has the Side-Effect Effect been cancelled?(No, not yet.). *Synthese*, 200(5), 1-15.
- Sytsma, J., & Livengood, J. (2012). Experimental philosophy and philosophical disputes. *Essays in Philosophy*, 13(1), 145-161.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53-55.
- Taylor, S. E. (1982). The availability bias in social perception and interaction. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 190-200). New York, NY: Cambridge University Press.
- Tennant, J. P., & Ross-Hellauer, T. (2020). The limitations to our understanding of peer review. *Research Integrity and Peer Review*, 5(1), 1-14.
- Tenney, M. (2015, March 11). *The dangers of moral relativism*. United Families International. <https://www.unitedfamilies.org/life/abortion/the-dangers-of-moral-relativism/>
- Theriault, J., Waytz, A., Heiphetz, L., & Young, L. (2017). Examining overlap in behavioral and neural representations of morals, facts, and preferences. *Journal of Experimental Psychology: General*, 146(11), 1586-1605.
- Theriault, J., Waytz, A., Heiphetz, L., & Young, L. (2020). Theory of Mind network activity is associated with metaethical judgment: An item analysis. *Neuropsychologia*, 143, 107475. <https://doi.org/10.1016/j.neuropsychologia.2020.107475>
- Thompson, K. (2022). Qualitative methods show that surveys misrepresent “ought implies can” judgments. *Philosophical Psychology*, 1-29. <https://doi.org/10.1080/09515089.2022.2036714>
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., & Staake, T. (2018). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science*, 64(3), 1458-1476.
- Tisak, M. (1995). Domains of social reasoning and beyond. *Annals of Child Development*, 11(1), 95-130.
- Tisak, M. S., & Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology*, 24(3), 352-357.

- Tisak, M. S., Crane-Ross, D., & Tisak, J. (2000). Mothers' and teachers' home and school rules: Young children's conceptions of authority in context. *Merrill-Palmer Quarterly*, 46, 168-187.
- Tosi, J., & Warmke, B. (2016). Moral grandstanding. *Philosophy & Public Affairs*, 44(3), 197-217.
- Tosi, J., & Warmke, B. (2020). *Grandstanding: The use and abuse of moral talk*. New York, NY: Oxford University Press.
- Trainer, F. E. (1983). Ethical objectivism-subjectivism: a neglected dimension in the study of moral thought. *Journal of Moral Education*, 12(3), 192-207.
- Trask, L. (1997). Scare quotes. *University of Sussex*.
<https://www.sussex.ac.uk/informatics/punctuation/quotes/scare>
- Trivers, R. (2011). *The folly of fools: The logic of deceit and self-deception in human life*. New York, NY: Basic Books.
- Turek, F. [@DrFrankTurek]. (2020, August 3). *If there is no objective morality then love is no better than murder* [Tweet]. Twitter. <https://twitter.com/drfrankturek/status/1290347092486086658>
- Turiel, E. (1974). Conflict and transition in adolescent moral development. *Child Development*, 45(1), 14-29.
- Turiel, E. (1979). Distinct conceptual and developmental domains: social convention and morality. In H. Howe & C. Keasey (Eds.), *Nebraska symposium on motivation, 1977: Social cognitive development*. Lincoln: University of Nebraska Press.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. New York, NY: Cambridge University Press.
- Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. New York, NY: Cambridge University Press.
- Turiel, E. (2012). Moral reasoning, cultural practices, and social inequalities. *Innovación Educativa*, 12(59), 17-32.
- Turiel, E., Nucci, L. P., & Smetana, J. G. (1988). A cross-cultural comparison about what? A critique of Nisan's (1987) study of morality and convention. *Developmental Psychology*, 24(1), 140-143.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207-232.
- United Nations Development Programme (2020). *Human development report 2020: The next frontier: Human development and the Anthropocene*. New York, NY: United Nations.

- United States Census Bureau (2021, October 8). *Urban areas facts*. United States Census Bureau. <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/ua-facts.html>
- United States Census Bureau (2022, February 24). *Census bureau releases new educational attainment data*. United States Census Bureau. <https://www.census.gov/newsroom/press-releases/2022/educational-attainment.html>
- Uskul, A. K., Nisbett, R. E., & Kitayama, S. (2008). Ecoculture, social interdependence, and holistic cognition: Evidence from farming, fishing, and herding communities in Turkey. *Communicative & Integrative Biology*, 1(1), 40-41.
- Uttich, K., Tsai, G., & Lombrozo, T. (2014). Exploring metaethical commitments: Moral objectivity and moral progress. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 108-208). London, UK: Bloomsbury.
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin*, 134(3), 383-403.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science-Cambridge*, 17(6), 476-477.
- Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. *PloS One*, 7(11), e48693. <https://doi.org/10.1371/journal.pone.0048693>
- Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492.
- van den Berg, R. G. (2022). SPSS RM ANOVA – 2 Within-subjects factors. *SPSS Tutorials*. <https://www.spss-tutorials.com/spss-repeated-measures-anova-2-within-subjects-factors/>
- van Dongen, N., Colombo, M., Romero, F., & Sprenger, J. (2021). Intuitions about the reference of proper names: A meta-analysis. *Review of Philosophy and Psychology*, 12(4), 745-774.
- Van Inwagen, P. (1997). Materialism and the psychological-continuity account of personal identity. *Philosophical Perspectives*, 11, 305-319.
- Van Roojen, M. (2015). *Metaethics: A contemporary introduction*. New York, NY: Routledge.
- Van Roojen, M. (2018). Moral cognitivism vs. non-cognitivism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2018 ed.). <https://plato.stanford.edu/entries/moral-cognitivism/>

- van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2-12.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Väyrynen, P. (2021). Thick ethical concepts. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 ed.). <https://plato.stanford.edu/archives/spr2021/entries/thick-ethical-concepts/>
- Veluwenkamp, H. (2017). Parfit's and Scanlon's Non-Metaphysical Moral Realism as Alethic Pluralism. *Ethical Theory and Moral Practice*, 20(4), 751-761.
- Viciano, H., Hannikainen, I. R., & Gaitan Torres, A. (2019). The dual nature of partisan prejudice: Morality and identity in a multiparty system. *PloS One*, 14(7), e0219509. <https://doi.org/10.1371/journal.pone.0219509>
- Viciano, H., Hannikainen, I. R., & Rodríguez-Arias, D. (2021). Absolutely right and relatively good: Consequentialists see bioethical disagreement in a relativist light. *AJOB Empirical Bioethics*, 12(3), 190-205.
- Villars, C. N. (1986). The paradox of Schrodinger's cat. *Physics Education*, 21(4), 232-237.
- Vogt, P. J., & Goldman, A. (Hosts). (2016, May 12). On the inside (No. 64) [Audio podcast episode]. In *Reply all*. Gimlet. <https://gimletmedia.com/shows/reply-all/brho4v/64-on-the-inside>
- Wagner, J. M., Pölzler, T., & Wright, J. C. (2021). Implicit metaethical intuitions: Validating and employing a new IAT procedure. *Review of Philosophy and Psychology*, 1-31. <https://doi.org/10.1007/s13164-021-00572-3>
- Wainryb, C., Shaw, L. A., Langley, M., Cottam, K., & Lewis, R. (2004). Children's thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development*, 75(3), 687-703.
- Walters, G. (2020, April 3). *Why young people should reject moral relativism to lead a better life*. Universal Owl. <https://universalowl.com/2020/04/03/why-young-people-should-reject-moral-relativism-to-lead-a-better-life/>
- Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, 29(1/2), 429-460.

- Welsh, M., Begg, S., & Bratvold, R. (2007). Efficacy of bias awareness in debiasing oil and gas judgments. D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1647-1652). Cognitive Science Society. <https://hdl.handle.net/2440/41191>
- Wetzel, E., & Roberts, B. W. (2020). Commentary on Hussey and Hughes (2020): Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, 3(4), 505-508.
- White, P. (1980). Limitations on verbal reports of internal events: A refutation of Nisbett and Wilson and of Bem. *Psychological Review*, 87(1), 105-112.
- Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, 7, 1832. <https://doi.org/10.3389/fpsyg.2016.01832>
- Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, 39(4), 202-217.
- Wike, E. L., & Church, J. D. (1976). Comments on Clark's "The language-as-fixed-effect fallacy". *Journal of Memory and Language*, 15(3), 249-255.
- Wikforss, Å. (2008). Semantic externalism and psychological externalism. *Philosophy Compass*, 3(1), 158-181.
- Willig, C. (2008). *Introducing qualitative research in psychology*. New York, NY: Open University Press.
- Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.
- Wineland, D. J. (2013). Nobel Lecture: Superposition, entanglement, and raising Schrödinger's cat. *Reviews of Modern Physics*, 85(3), 1103-1114.
- Wolff, J. (2018). *An introduction to moral philosophy*. New York, NY: W.W. Norton & Company.
- Wong, D. B. (1995). Pluralistic relativism. *Midwest Studies in Philosophy*, 20, 378-399.
- Wong, D. B. (2006). *Natural moralities: A defense of pluralistic relativism*. New York, NY: Oxford University Press.
- Wright, J. C. (2015). Meta-ethical pluralism: A cautionary tale about cohesive moral communities. *Behavioral and Brain Sciences*, 38, e163, 44-44. doi:10.1017/S0140525X14001344

- Wright, J. C. (2018). The fact and function of meta-ethical pluralism: Exploring the evidence. In T. Lombrozo, J. Knobe, S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 2, pp. 119-150). Oxford, UK: Oxford University Press.
- Wright, J. C., Cullum, J., & Schwab, N. (2008). The cognitive and affective dimensions of moral conviction: Implications for attitudinal and behavioral measures of interpersonal tolerance. *Personality and Social Psychology Bulletin*, 34(11), 1461-1476.
- 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(3), 336-361.
- Wright, J. C., McWhite, C. B., & Grandjean, P. T. (2014). The cognitive mechanisms of intolerance: Do our meta-ethical commitments matter?. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 1, pp. 28-61). New York, NY: Oxford University Press.
- Wygod Cohen, S. (2021). Mind Independence versus Mind Nongroundedness: Two Kinds of Objectivism. *Ethics*, 132(1), 180-203.
- Yang, Z., Fu, D., Qi, Y., Zheng, Y., Li, Q., & Liu, X. (2019). Humor affects fairness considerations in the gain and loss contexts. *Frontiers in Psychology*, 9, 2679.
<https://doi.org/10.3389/fpsyg.2018.02679>
- Yasenchuk, K. (1997). Moral realism and the burden of argument. *The Southern Journal of Philosophy*, 35(2), 247-264.
- Yau, J., & Smetana, J. G. (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, 74(3), 647-658.
- Yilmaz, O., & Bahçekapili, H. G. (2015a). Without God, everything is permitted? The reciprocal influence of religious and meta-ethical beliefs. *Journal of Experimental Social Psychology*, 58, 95-100.
- Yilmaz, O., & Bahçekapili, H. G. (2015b). Without God, everything is permitted? The reciprocal influence of religious and meta-ethical beliefs. *Journal of Experimental Social Psychology*, 58(Suppl. A), 95-100.
- Yilmaz, O., & Bahçekapili, H. G. (2018). Meta-ethics and the mortality: Mortality salience leads people to adopt a less subjectivist morality. *Cognition*, 179, 171-177.
- Yilmaz, O., Bahçekapili, H. G., Harma, M., & Sevi, B. (2020). Intergroup tolerance leads to subjective morality, which in turn is associated with (but does not lead to) reduced religiosity. *Archive for the Psychology of Religion*, 42(2), 232-243.

- Young, L., & Durwin, A. J. (2013a). Moral realism as moral motivation: The impact of meta-ethics on everyday decision-making. *Journal of Experimental Social Psychology*, 49(2), 302-306.
- Young, L., & Durwin, A. J. (2013b). Moral realism as moral motivation: The impact of meta-ethics on everyday decision-making. *Journal of Experimental Social Psychology*, 49(2, Suppl.), 302-306.
- Zahn-Waxler, C., & Robinson, J. (1995). Empathy and guilt: Early origins of feelings of responsibility. In J. P. Tangney, & K. W. Fisher (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 143-173). New York, NY: Guilford Press.
- Zhu, J. H. (1996). 'I don't know' in public opinion surveys in China: Individual and contextual causes of item non-response. *Journal of Contemporary China*, 5(12), 223-244.
- Zijlstra, L. (2019). Folk moral objectivism and its measurement. *Journal of Experimental Social Psychology*, 84, 103807. <https://doi.org/10.1016/j.jesp.2019.04.005>
- Zijlstra, L. (2021). Are people implicitly moral objectivists?. *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-021-00593-y>

Appendix A: Glossary

Agent relativism

A form of relativism which relativizes moral truth to the goals, standards, or values of agents, rather than those judging the agent. For instance, whether it would be morally right or wrong for Alex to steal would depend on Alex's standards, not the moral standards of whoever is judging Alex. (Note that this description focuses on the deontic properties of actions but could also be used for evaluative claims, judgments about character traits, thick moral concepts, and nonmoral norms.)

Appraiser relativism

A form of relativism which relativizes moral truth to whoever is judging an action, rather than the agent performing the action. For instance, whether it would be morally right or wrong for Alex to steal would depend on the standards of whoever is judging Alex's actions, not Alex's standards (Note that this description focuses on the deontic properties of actions but could also be used for evaluative claims, judgments about character traits, thick moral concepts, and nonmoral norms.)

Clear intended interpretation

An interpretation of the given stimuli that appears consistent with researcher intent and is sufficiently clear to be confidently classified as an instance of an intended interpretation.

Clear interpretation

An interpretation of the given stimuli that is sufficiently clear that it can be confidently classified as an instance of either an intended or unintended interpretation.

Cognitivism

The metaethical position that moral claims are propositional, and are therefore capable of being true or false.

Commitment

(See *philosophical commitment*)

Conflation

In the context of research on folk metaethics, conflations occur when researchers or participants conflate a particular metaethical concept or distinction with another concept or distinction (whether it be a different metaethical concept/distinction or a non-metaethical concept/distinction). (See *formal conflation* and *informal conflation*)

Correspondence theory of truth

A view of *truth* which holds that claims are true insofar as they *represent* (or “correspond”) *reality*. In the context of moral claims, a correspondence-theoretic conception of truth would hold that moral claims are true insofar as they *represent moral facts* (Pözlner, 2018b, p. 662).

Cultural relativism

A form of relativism that relativizes moral truth to the moral standards of different cultures.

Cultural relativism

The metaethical position that moral claims are true or false relative to different cultural standards.

Descriptive metaethics

A branch of metaethics that seeks to offer the best account of the metaethical stances and commitments of moral thought and language. Could be described more simply as *the study of folk metaethics*.

Descriptive relativism

The descriptive claim that different individuals or societies have different moral values. In its stronger form, descriptive relativism refers to the notion that there are widespread and ineliminable fundamental moral disagreements, i.e., moral disagreements that cannot be reduced to differences in nonmoral beliefs.

Determinate

The position that there is a discoverable fact of the matter about whether competing propositions with respect to a given topic are true or best supported by reasons and evidence.

Disagreement paradigm

A common paradigm used to assess folk moral realism and antirealism. The disagreement paradigm presents a participant with moral (and potentially nonmoral) disagreements and asks them to judge whether one or both people who disagree can be (or are) correct. Response options are intended to reflect different realist and antirealist positions.

Error theory | The error theory

A metaethical position which holds that all first-order moral claims are false. Typically, error theories maintain that all first-order moral claims are false because they are implicitly committed to one or more false presuppositions, e.g., all moral claims implicitly presuppose that God exists, but God does not exist, so all such claims are false.

Evaluative standard ambiguity

An ambiguity that arises whenever there is insufficient information to know which moral framework a given moral claim could (if relativized) be indexed to. For instance, if we are told that Alex and Sam disagree about whether stealing is morally wrong, a cultural relativist cannot judge whether they can both be correct or whether one must be mistaken without knowing which culture(s) Alex and Sam are from, or which cultures Alex and Sam are referencing when making their claims.

External validity

The degree to which findings obtained in a given study justify inferences about people and situations outside the context of the study.

Folk philosophy

The philosophical stances and commitments of ordinary people. (See *ordinary people*)

Forced choice

A feature of studies in which participants are required to choose from among a limited set of response options that do not represent the full range of possible responses.

Formal conflation

In the context of research on folk metaethics, this occurs when stimuli erroneously conflate metaethical concepts or distinctions with other, unintended concepts or distinctions. For instance, researchers may use an item intended to measure *moral realism*, but accidentally use an item that a person competent with the relevant distinctions would recognize as an expression of universalism, or at least to be sufficiently ambiguous so as to plausibly be interpreted this way. Formal conflations are not the same as informal conflations. In principle, a competent person could spot a formal conflation and ignore it. Conversely, even if there is no formal conflation, someone might nevertheless conflate one concept/distinction with another concept/distinction.

Fundamental moral disagreement

Any moral disagreement that results from a genuine and ineliminable difference in moral values, and thus cannot be reduced to a difference in nonmoral beliefs.

Indeterminacy / indeterminism

The position that there is *no* discoverable fact of the matter about whether competing propositions with respect to a given topic are true or best supported by reasons and evidence.

Indeterminacy-Variability Thesis | IV Thesis

The hypothesis that folk metaethics is some combination of indeterminate or variable. Initially proposed by Michael Gill as an alternative to the UD assumption. (See *indeterminacy* and *variability*.)

Informal conflation

This occurs whenever a participant conflates a metaethical concept/distinction with a different concept/distinction, not as a result of experimenter error, but due to the participant failing to interpret face valid stimuli as intended.

Intended interpretation

An interpretation of the given stimuli that *appears* consistent with researcher intent. Intended interpretations may be clear or unclear, so responses coded as intended interpretations may not genuinely reflect a genuine instance of an intended interpretation.

Interpersonal variability

Metaethical pluralism with respect to the moral stances or commitments of different individuals or groups.

Interpretative variation

An index of the degree to which participants differ from one another in how they interpret the stimuli presented in a study. High interpretative variation would occur when participants frequently interpret stimuli differently from one another.

Interpretative variation

Systematic variation in how participants interpret experimental stimuli. For instance, some participants answering questions about a “bank” may interpret the term *bank* to refer to a riverbank, while others may interpret *bank* to refer to a financial institution. Interpretative variation may or may not be significant and may or may not have meaningful implications for a given study.

Intrapersonal variability

Metaethical pluralism with respect to the moral stances or commitments of an individual.

Magic: The Gathering | MTG

A collectible card game in which players take on the role of dueling wizards who summon monsters and cast spells. The goal of the game is to defeat all opponents. Typically, this involves reducing the life total of the other player (or players) to zero, though there are a variety of alternative ways to win.

Metaethical indeterminacy

The descriptive metaethical position that ordinary people have no determinate metaethical stances and/or commitments.

Metaethical pluralism

The descriptive metaethical position that ordinary people have different metaethical stances or commitments. For example, some people endorse realism or speak like realists, while others endorse antirealism or speak like antirealists. Pluralism can be both interpersonal and intrapersonal. Pluralism is not the mere recognition that some ordinary people may speak or think in unconventional ways, but that substantive and ineliminable elements of ordinary moral thought and language are best characterized by traditionally competing philosophical accounts (e.g., cognitivism and noncognitivism).

Metaethical quietism

The view that the dispute between realism and antirealism is a pseudoproblem resulting from linguistic and conceptual confusions. Technically an antirealist position, since quietism rejects realism, quietism nevertheless rejects much of the substantive content of conventional antirealist accounts as well.

Metaethical variability

The descriptive metaethical position that people do not have uniform metaethical stances or commitments, but instead exhibit variable metaethical stances or commitments.

Mind-independence

Another term for stance-independence.

Modal operator scope ambiguity

This occurs whenever the scope of a modal operator is ambiguous in a given sentence, such that it is subject to more than one possible interpretation. For instance, “Alex and Sam could both win the game” is ambiguous between an inclusive reading and an exclusive reading. On an exclusive reading, Alex *or* Sam could win, *but not both* (they could be playing a game that only allows one person to win). On an inclusive reading, Alex and Sam could both win *at the same time* (perhaps they are playing on the same team). Typically, such ambiguities are resolved by context. For instance, “Alex and Sam could both be tall” would typically be interpreted in inclusive terms, while “Alex and Sam could both win the gold medal for the 200m breaststroke at the next Olympics” would be exclusive.

Moral antirealism

The metaethical position that there are no stance-independent moral facts.

Moral disagreement

Any instance in which two people hold conflicting moral stances about a first-order moral claim. Such disagreements do not require either person to be aware of the other’s position or for them to be actively engaged in a dispute. “Disagreement” simply refers to the difference in first-order moral stances, *not* to some social phenomenon, such as a debate or conflict.

Moral domain

Norms and evaluations that are distinctively *moral* and thus could in principle be distinguished from nonmoral norms. The moral domain could in principle be unified by one or more metanormative characteristics or its distinctness could be primitive (i.e., incapable of being reduced or defined by appeal to other terms or concepts).

Moral grandstanding

Strategically exploiting moral discourse to enhance social status (Grubbs et al., 2019).

Moral incoherentism | incoherentism

A form of moral antirealism advanced by Don Loeb (2008). Incoherentism is the metaethical position that (1) (the descriptive metaethical / linguistic thesis) folk metaethics is committed to conflicting metaethical presuppositions (e.g., both realism and antirealism) and that (2) (the metaphysical thesis) along with additional considerations, this suggests that there are no stance-independent moral facts (this glosses over the explanation and arguments for the view). For example, if a person asserting that “murder is wrong” is committed both to the notion that this claim refers to a stance-independent moral fact *and* that it *doesn’t*, such a claim would be incoherent.

Moral realism

The claim that there are stance-independent moral facts. (Note: moral realism is sometimes described in other ways. It can be broader, including all positions which hold that there are moral facts, or narrower, by including additional theses such as the claim that we possess at least some moral knowledge, or that we are capable of knowing at least some moral facts. I do not employ these uses of *moral realism* here.)

Moral relativism

The metaethical position that moral claims are true or false relative to different standards of evaluation, such as the standards of individuals or groups.

Naturalism | Metaethical naturalism

The claim that moral facts are a type of *natural fact*. Natural facts are facts that are identical with, reducible to, or consistent with the facts described by the natural sciences.

Noncognitivism

The metaethical position that moral claims are not propositional, and are therefore incapable of being true or false.

Non-naturalism | Metaethical non-naturalism

The claim that moral facts cannot be reduced to *natural facts*.

Nonmoral differences

Differences that are not attributable to different moral standards. For instance, two people with the same moral standards could disagree about the moral status of abortion because they disagree about empirical details related to the viability of the fetus outside the womb or its capacity for pain or conscious experience.

Nonphilosophers

People without significant training in academic philosophy.

Normative antirealism

The claim that there are no stance-independent normative facts.

Normative conflation

Normative conflations are a distinctive form of conflation that occurs in studies designed to evaluate metaethical stances or commitments. Participants presented with stimuli intended to solicit metaethical stances or commitments may instead interpret stimuli to be asking questions about their normative moral standards. When this occurs, their responses may reflect their normative moral standards, rather than their metaethical standards. Such instances render any measures predicated on an intended interpretation invalid.

Normative domain

A general term for referring to different normative/evaluative domains, such as *moral norms*, *epistemic norms*, *prudential norms*, and so on.

Normative entanglement

The conflation between metaethical and normative considerations, or mistaken inferences about necessary conceptual relations between metaethics and normative ethics, e.g., someone may insist that

one cannot have normative moral standards unless they endorse moral realism, or that if you endorse moral antirealism, that you cannot object to someone harming you.

This occurs whenever metaethical claims are embedded (implicitly or explicitly) in normative claims or claims with potentially significant normative implications. Normative entanglement may be used as a rhetorical strategy to imply that people who disagree with the speaker hold immoral or undesirable stances and attitudes. For instance, a moral realist may ask a moral antirealist:

“Do you think it’s objectively wrong to torture babies for fun?”

There is no way to respond to this question with a simple “yes” or “no” without expressing a stance towards moral realism *and* a direct or implied normative position on the moral status of torturing babies for fun. This is because this question embeds a metaethical claim inside a normative claim, and simultaneously asks about both:

- (1) Metaethical claim: Torturing babies for fun is *stance-independently* morally wrong
- (2) Normative claim: Torturing babies for fun is morally wrong

An antirealist cannot respond with a “yes,” since this would be a concession to (1), which they do not accept. This only leaves them with the option to respond with “no.” Technically, responding with a “no” only entails a rejection of (1). However, note that to respond with a “yes” simultaneously affirms both (1) and (2). To deny (1) may therefore be interpreted to *pragmatically imply* a rejection of (2).

Note, also, that in addition to expressing (1) and (2), this claim also carries pragmatic implications about a respondent’s non-metaethical and non-propositional *attitudes* towards the action in question. To respond with a “no,” may imply more than merely that one does not hold the normative moral stance that torturing babies is morally wrong, but that one does not personally object to torturing babies for fun, but is ambivalent, indifferent, or even in support of doing so. As a result, to respond with a “no” may give the impression that the antirealist has extremely deviant and antisocial standards, or may lack such standards altogether.

Moral realists often employ questions like these in order to imply that moral antirealists reject (2), and that they therefore have no objections to, and are not opposed to, various moral transgressions. The effectiveness of this ploy often revolves around how realistic and how serious the transgression is. As a result, using normative entanglement as a rhetorical strategy is typically most effective when the person employing the strategy appeals to an extremely serious moral transgression that causes widespread repugnance (such as abusing animals), but that isn’t too outlandish or implausible to lose its rhetorical force (such as joining the dark side and using the Force to slaughter Wookiees).

Normative realism

The claim that there are stance-independent normative facts.

Normative relativism

The moral stance that we should tolerate people and cultures with different moral standards. Not a genuine form of “relativism,” but a concept frequently conflated with relativism as a metaethical position and closely connected to it.

Objectivism

A common term often used interchangeably with *stance-independence*.

Objectivism

A term commonly used in metaethics. Used interchangeably with “realism” and “stance-independence.”

Ordinary people

Nonphilosophers, i.e., people without significant training in academic philosophy.

Philosophical commitment

An account of a speech act or behavior that is best explained in accordance with a particular philosophical position.

Philosophical induction

Inducing ordinary people to engage in philosophy to such an extent that they are no longer ordinary people.

Philosophical stance

A belief about the truth of a philosophical position.

Reflexive thematic analysis (RTA)

A method of thematic analysis pioneered by Braun and Clarke which purports to be methodologically distinct from other forms of thematic analysis. This approach emphasizes the active, fluid, and evolving process of coding and rejects the notion of a single, objectively correct analysis. As such, RTA tends to reject the value of interrater reliability (FAQs, n.d.).

Relation-designating account

Relation designating accounts are those for which the truth conditions of a given set of claims are contingent, and are made with respect to, a fixed standard of evaluation, such that all of those claims are stance-dependent but all relate to the same standard. As such, the truth conditions for a given statement *cannot* vary across speakers, appraisers, and so on. For instance, suppose you endorsed the theory that whether an action is morally right or wrong depends on the moral standards of the current ruler. All statements, such as “murder is morally wrong” and “kindness is morally good” relate to the standards of this ruler. As such, they are stance-dependently true or false, in that they are *made true* by the ruler’s standards, but since there is only one current ruler with one set of moral standards, the moral rules cannot be true relative to one standard, but false relative to another.

Relativism

The claim that moral claims are true or false relative to the moral standards of different individuals or groups. Relativism may also be construed as a semantic thesis about the meaning of moral claims, namely, that they contain implicit indexicals that allow for the truth of moral claims to vary in accordance with the moral standards they relativized to.

Response option

One of a set of responses available to participants participating in a study, e.g., “true” and “false” would be standard response options to a true or false question.

Scope ambiguity

(See *modal operator scope ambiguity*)

Spontaneous theorizing

A phenomenon which occurs whenever the experimental context causes participants who held no determinate stance or commitment prior to participating in the study adopt or express a stance or commitment due to the experience of participating in the study itself.

Stance

(See *philosophical stance*)

Stance-dependent

A feature of the truthmaking properties of a given proposition. A proposition is stance-dependent insofar as it is made true by a stance.

Stance-independent

A feature of the truthmakers of a given proposition or set of propositions whereby that proposition (or those propositions) is *not* made true by real or hypothetical stances, such as the stances of individuals or groups.

Stance-independent

A feature of the truthmaking properties of a given proposition. A proposition is stance-independent insofar as it is made not true by a stance.

Strategic metaethical pluralism

A form of metaethical pluralism in which people adopt or express a particular metaethical stance or commitment to achieve social goals, such as persuasion or signaling desirable character traits. May explain apparent inconsistencies with people’s metaethical views by redescribing them in rhetorical or argumentative terms.

Subjectivism

A form of moral relativism that relativizes moral truth to the moral standards of different individuals.

Subjectivism

A form of relativism which relativizes moral claims to the standards of individuals.

Thematic analysis

A method of analysis used in qualitative research that centers on the use of *themes* to organize and evaluate recurring patterns in a given body of data (FAQs, n.d.).

Themes

A distinct, recurring pattern within a given dataset (FAQs, n.d.).

Training paradigms

Studies that employ extensive instructions intended to teach participants about the concepts and distinctions captured by the psychological constructs the study seeks to measure. Such studies often include training exercises (such as practice questions) and comprehension checks, and may exclude participants that fail at these tasks.

Unclear interpretation

An interpretation of the given stimuli that is not sufficiently clear to be confidently classified as an instance of either an intended or unintended interpretation.

Uniformism

The claim that there is a single determinate account of folk metaethics.

Uniformity-Determinacy assumption | UD assumption

The assumption that there is a uniform and determinate account of folk metaethics. An idea originally put forward by Michael Gill, who attributed the UD assumption to much of 20th century descriptive metaethics.

Unintended interpretation

An interpretation of the given stimuli that does not appear consistent with researcher intent. Intended interpretations may be clear or unclear, so it is possible that a participant whose response was coded as unintended did interpret the stimuli as intended.

Universalism

The claim that there is a single correct moral standard (Ayars & Nichols, 2018). Alternatively, universalism could also be used to convey the notion that a given moral principle or rule applies to everyone, and is thereby concerned with the *scope* of the moral standard in question (Goodwin & Darley, 2008).

WEIRD

An acronym proposed by Henrich and colleagues to refer to the standard research subjects in most psychological research. WEIRD stands for **W**estern, **E**ducated, **I**ndustrialized, **R**ich, and **D**emocratic.

APPENDIX B: Bibliography of Folk Metaethics Studies

The following bibliography provides a complete list of all published studies which address folk moral realism and antirealism in some way in chronological order. This list was updated on February 3, 2023, and consists only of those studies that, to the best of my knowledge, provide empirical data. It is possible I've missed some studies.

Bibliography

- Trainer, F. E. (1983). Ethical objectivism-subjectivism: a neglected dimension in the study of moral thought. *Journal of Moral Education*, 12(3), 192-207.
- Nichols, S., & Folds-Bennett, T. (2003). Are children moral objectivists? Children's judgments about moral and response-dependent properties. *Cognition*, 90(2), B23-B32.
- Nichols, S. (2004). After objectivity: An empirical study of moral judgment. *Philosophical Psychology*, 17(1), 3-26.
- Wainryb, C., Shaw, L. A., Langley, M., Cottam, K., & Lewis, R. (2004). Children's thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development*, 75(3), 687-703.
- Cova, F., Ravat, J. (2008). Common sense and moral objectivism: Global or local? A case of experimental philosophy. *KLESIS - Revue philosophique*, 9, 180-202.
- Feltz, A., & Cokely, E. T. (2008). The fragmented folk: more evidence of stable individual differences in moral judgments and folk intuitions. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1771-1776). Austin, TX: Cognitive science society.
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, 106(3), 1339-1366.
- Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind & Language*, 26(4), 482-505.
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others?. *Journal of Experimental Social Psychology*, 48(1), 250-256.
- Quintelier, K. J., & Fessler, D. M. (2012). Varying versions of moral relativism: The philosophy and psychology of normative relativism. *Biology & Philosophy*, 27(1), 95-113.

- Young, L., & Durwin, A. J. (2013a). Moral realism as moral motivation: The impact of meta-ethics on everyday decision-making. *Journal of Experimental Social Psychology*, 49(2), 302-306.
- Rai, T. S., & Holyoak, K. J. (2013). Exposure to moral relativism compromises moral behavior. *Journal of Experimental Social Psychology*, 49(6), 995-1001.
- Quintelier, K. J., De Smet, D., & Fessler, D. M. (2013). The moral universalism-relativism debate. *KLESIS–Revue Philosophique*, 27, 211-262.
- Heiphetz, L., Spelke, E. S., Harris, P. L., & Banaji, M. R. (2013). The development of reasoning about beliefs: Fact, preference, and ideology. *Journal of Experimental Social Psychology*, 49(3), 559-565.
- 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(3), 336-361.
- Quintelier, K., De Smet, D., & Fessler, D. (2014). Agent versus appraiser moral relativism: an exploratory study. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 209-230). London, UK: Bloomsbury Academic.
- Uttich, K., Tsai, G., & Lombrozo, T. (2014). Exploring metaethical commitments: Moral objectivity and moral progress. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 108-208). London, UK: Bloomsbury.
- Wright, J. C., McWhite, C. B., & Grandjean, P. T. (2014). The cognitive mechanisms of intolerance: Do our meta-ethical commitments matter?. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 1, pp. 28-61). New York, NY: Oxford University Press.
- Beebe, J. R. (2014). How different kinds of disagreement impact folk metaethical judgments. In H. Sarkissian & J. C. Wright (Eds.), *Advances in experimental moral psychology* (pp. 167-187). London, UK: Bloomsbury.
- Beebe, J. (2015). The empirical study of folk metaethics. *Etyka*, 50, 11-28.
- Beebe, J., Qiaoan, R., Wysocki, T., & Endara, M. A. (2015). Moral objectivism in cross-cultural perspective. *Journal of Cognition and Culture*, 15(3-4), 386-401.
- Yilmaz, O., & Bahcekapili, H. G. (2015). Without God, everything is permitted? The reciprocal influence of religious and meta-ethical beliefs. *Journal of Experimental Social Psychology*, 58, 95-100.

- Sarkissian, H. (2016). Aspects of Folk Morality. In J. Sytsma, & W. Buckwalter (Eds.), *A companion to experimental philosophy* (pp. 212-224). West Sussex, UK: Wiley-Blackwell.
- Beebe, J. R., & Sackris, D. (2016). Moral objectivism across the lifespan. *Philosophical Psychology*, 29(6), 912-929.
- Heiphetz, L., & Young, L. L. (2017). Can only one person be right? The development of objectivism and social preferences regarding widely shared and controversial moral beliefs. *Cognition*, 167, 78-90.
- Fisher, M., Knobe, J., Strickland, B., & Keil, F. C. (2017). The influence of social interaction on intuitions of objectivity and subjectivity. *Cognitive science*, 41(4), 1119-1134.
- Theriault, J., Waytz, A., Heiphetz, L., & Young, L. (2017). Examining overlap in behavioral and neural representations of morals, facts, and preferences. *Journal of Experimental Psychology: General*, 146(11), 1586-1605.
- Schmidt, M. F., Gonzalez-Cabrera, I., & Tomasello, M. (2017). Children's developing metaethical judgments. *Journal of Experimental Child Psychology*, 164, 163-177.
- Khoo, J., & Knobe, J. (2018). Moral disagreement and moral semantics. *Noûs*, 52(1), 109-143.
- Wright, J. C. (2018). The fact and function of meta-ethical pluralism: Exploring the evidence. In T. Lombrozo, J. Knobe, S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 2, pp. 119-150). Oxford, UK: Oxford University Press.
- Yilmaz, O., & Bahçekapili, H. G. (2018). Meta-ethics and the mortality: Mortality salience leads people to adopt a less subjectivist morality. *Cognition*, 179, 171-177.
- Colebrook, R. T. (2018). Toward a science of morals. [Doctoral dissertation, City University of New York.] CUNY Academic Works. https://academicworks.cuny.edu/gc_etds/2648
- Ayars, A., & Nichols, S. (2020). Rational learners and metaethics: Universalism, relativism, and evidence from consensus. *Mind & Language*, 35(1), 67-89.
- Johnson, S. G., Rodrigues, M., & Tuckett, D. (2021). Moral tribalism and its discontents: How intuitive theories of ethics shape consumers' deference to experts. *Journal of Behavioral Decision Making*, 34(1), 47-65.
- Rose, D., & Nichols, S. (2019). From punishment to universalism. *Mind & Language*, 34(1), 59-72.

- Hopster, J.K.G. (2019). Moral objectivity: Origins and foundations (Quaestiones Inertiae, volume 119) [Doctoral dissertation, Utrecht University]. Utrecht University Repository. <https://dspace.library.uu.nl/handle/1874/374619>
- Sarkissian, H., & Phelan, M. (2019). Moral objectivism and a punishing God. *Journal of Experimental Social Psychology*, 80, 1-7.
- Becklloyd, D. and Sytsma, J.(2019). *Simulating metaethics: Consensus and the independence of moral beliefs*. PhilSci Archive. <http://philsci-archive.pitt.edu/16461/>
- Zijlstra, L. (2019). Folk moral objectivism and its measurement. *Journal of Experimental Social Psychology*, 84, 103807. <https://doi.org/10.1016/j.jesp.2019.04.005>
- Viciano, H., Hannikainen, I. R., & Gaitan Torres, A. (2019). The dual nature of partisan prejudice: Morality and identity in a multiparty system. *PloS One*, 14(7), e0219509. <https://doi.org/10.1371/journal.pone.0219509>
- Collier-Spruel, L., Hawkins, A., Jayawickreme, E., Fleeson, W., & Furr, R. M. (2019). Relativism or tolerance? Defining, assessing, connecting, and distinguishing two moral personality features with prominent roles in modern societies. *Journal of Personality*, 87(6), 1170-1188.
- Davis, T. (2021). Beyond objectivism: new methods for studying metaethical intuitions. *Philosophical Psychology*, 34(1), 125-153.
- Pölzler, T., & Cole Wright, J. C. (2020a). An empirical argument against moral non-cognitivism. *Inquiry*, 1-29. <https://doi.org/10.1080/0020174X.2020.1798280>
- Pölzler, T., & Wright, J. C. (2020b). Anti-realist pluralism: A new approach to folk metaethics. *Review of Philosophy and Psychology*, 11(1), 53-82.
- Stanley, M. L., Marsh, E. J., & Kay, A. C. (2020). Structure-seeking as a psychological antecedent of beliefs about morality. *Journal of Experimental Psychology: General*, 149(10), 1908-1918.
- Theriault, J., Waytz, A., Heiphetz, L., & Young, L. (2020). Theory of Mind network activity is associated with metaethical judgment: An item analysis. *Neuropsychologia*, 143, 107475. <https://doi.org/10.1016/j.neuropsychologia.2020.107475>
- Yilmaz, O., Bahcekapili, H. G., Harma, M., & Sevi, B. (2020). Intergroup tolerance leads to subjective morality, which in turn is associated with (but does not lead to) reduced religiosity. *Archive for the Psychology of Religion*, 42(2), 232-243.

- Sousa, P., Allard, A., Piazza, J., & Goodwin, G. P. (2021). Folk moral objectivism: The case of harmful actions. *Frontiers in Psychology*, 12, Article 2776. <https://doi.org/10.3389/fpsyg.2021.638515>
- Sytsma, J., Muldoon, R., & Nichols, S. (2021). The meta-wisdom of crowds. *Synthese*, 199(3), 11051-11074.
- Viciano, H., Hannikainen, I. R., & Rodríguez-Arias, D. (2021). Absolutely right and relatively good: Consequentialists see bioethical disagreement in a relativist light. *AJOB Empirical Bioethics*, 12(3), 190-205.
- Wagner, J. M., Pölzler, T., & Wright, J. C. (2021). Implicit Metaethical Intuitions: Validating and Employing a New IAT Procedure. *Review of Philosophy and Psychology*, 1-31. <https://doi.org/10.1007/s13164-021-00572-3>
- Zijlstra, L. (2021). Are people implicitly moral objectivists?. *Review of Philosophy and Psychology*, 1-19. <https://doi.org/10.1007/s13164-021-00593-y>
- Cova, F. (2022, June 15). The need for measure calibration in experimental philosophy: the case of measures of folk objectivism. <https://doi.org/10.31234/osf.io/kcft9>
- Lewry, C., & Lombrozo, T. (2022). Ethical Explanations. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44), 42-48. <https://escholarship.org/uc/item/2rv1r2s0>
- Chvaja, R., Horský, J., Lang, M., & Kundt, R. (2022). Positive Association Between Ritual Performance and Perceived Objectivity of Moral Norms. *The International Journal for the Psychology of Religion*, 1-21. <https://doi.org/10.1080/10508619.2022.2121454>
- Pölzler, T., Zijlstra, L., & Dijkstra, J. (2022). Moral progress, knowledge and error: Do people believe in moral objectivity?. *Philosophical Psychology*, 1-37.
- Rodríguez-Pérez, A., Rodríguez-Torres, R., Betancor, V., Chen-Xia, X. J., & Rodríguez-Gómez, L. (2022). Are Civility Norms Morality Norms' Little Sister? The Truth Value That Lay Thinking Associates with Civility and Morality Social Norms. *Social Sciences*, 11(12), 568.
- Wright, J. C., & Pölzler, T. (2022). Should morality be abolished? An empirical challenge to the argument from intolerance. *Philosophical Psychology*, 35(3), 350-385.

APPENDIX C:

Metaethics scales

Relativism Subscale of the Ethics Position Questionnaire (EPQ)

Forsyth (1980)

Source

Forsyth, D. R. (1980). A taxonomy of ethical ideologies. *Journal of Personality and Social psychology*, 39(1), 175-184.

Scale

A = strongly disagree, B = disagree, C = no opinion or neutral, D = agree, E = strongly agree

1. There are no ethical principles that are so important that they should be part of any code of ethics.
2. What is ethical varies from one situation and society to another.
3. Moral standards should be seen as being individualistic; what one person considers to be moral may be judged to be immoral by another person.
4. Different types of moralities cannot be compared as to "rightness."
5. Questions of what is ethical for everyone can never be resolved since what is moral or immoral is up to the individual.
6. Moral standards are simply personal rules which indicate how a person should behave, and are not to be applied in making judgments of others.
7. Ethical considerations in interpersonal relations are so complex that individuals should be allowed to formulate their own individual codes.
8. Rigidly codifying an ethical position that prevents certain types of actions could stand in the way of better human relations and adjustment.
9. No rule concerning lying can be formulated; whether a lie is permissible or not totally depends on the situation.
10. Whether a lie is judged to be moral or immoral depends upon the circumstances surrounding the action.

Objectivism/Subjectivism scale (TOS)

Trainer (1983)

Source

Trainer, F. E. (1983). Ethical objectivism-subjectivism: a neglected dimension in the study of moral thought. *Journal of Moral Education*, 12(3), 192-207.

Scale

Dichotomous choice between realist and antirealist response.

A = Realist response, B = Antirealist response

- 1A. You can say without any doubt in some situations that something is right or wrong, and you say that people who don't agree with you are wrong.
- 1B. It isn't possible for anyone to be really sure what is right or wrong. You can only say that others with different opinions are wrong.
- 2A. Some things are wrong no matter what anyone thinks and people should be told this if they don't know.
- 2B. People should be free to work out for themselves what they think is right or wrong.
- 3A. You can be quite sure that some of your basic moral principles are right.
- 3B. You can't be sure that your basic moral principles are right. Even 'It is wrong to kill' is just an assumption or an item of belief on your part.
- 4A. Some values or actions are objectively right, they are right in fact, whether or not individuals think so.
- 4B. All judgements about right and wrong state nothing more objective than the ideas or attitudes of individuals.
- 5A. There are fixed and eternal moral laws of nature.
- 5B. All moral laws can be made, altered or rejected by man.
- 6A. Moral laws can't be changed by human beings any more than the laws of physics can be.
- 6B. Moral laws can be changed by man because they are only man-made rules in the first place.
- 7A. Human beings can only discover moral laws; we can't make them. Just as we can't make up true laws of science to suit ourselves neither can we make up true moral laws.
- 7B. Moral laws are made up by humans as ways they choose for regulating behaviour.
- 8A. In any problem to do with justice there is in fact a just solution, even though no one may be able to say what it is.
- 8B. You can only talk about what you or others would accept as a just or fair solution, not about what is in fact the just solution.
- 9A. There would still be fundamental human rights even if no one on earth thought there were.
- 9B. Rights are entirely created by man.
- 10A. It is possible to know that your basic moral principles are the right criteria for evaluating things.
- 10B. Your basic moral principles can only be your best guess at the criteria for evaluating things; you can never know whether yours are the right criteria.

- 11A. My moral principles are not just things I prefer. They are things that ought to be preferred because they are morally Important.
- 11B. My moral principles can't be any more than the values I choose or prefer to live by; others choose other values and I can't say my choices are in fact the right ones.
- 12A. It is possible to claim that your moral principles are better than those of some other people, such as a thief or a sadist.
- 12B. You can't claim that your moral principles are better than anyone else's.
- 13A. Some things are in fact morally better than others no matter what anyone thinks or prefers.
- 13B. Our basic moral values are only our personal preferences. You might prefer kindness to cruelty but you can't say it is in fact morally better.
- 14A. We can say much more than that we do not like this. We could say cruelty to animals is in fact morally bad and should not be done whether or not anyone likes to do it.
- 14B. If we found someone who likes to torture animals for the fun of it we could not say this was a morally bad thing to do. We could only say that his interest disgusts us and that we wish he would not practise it.
- 15A. Honesty is in fact morally better than cheating. There is more involved here than my liking for one and my dislike of another.
- 15B. I can't say that even my most important values are objectively morally better than any other values at all. There is, for example, no objective difference in the moral quality of honesty and cheating. All I can say is that I like one and not the other.
- 16A. Whether or not you disapprove and whether or not men make laws against it, murder is in fact morally wrong.
- 16B. I cannot condemn even murder as being a morally bad action, because it breaks no moral law. There are no moral laws of nature, there are only laws men make up. All I can say are things like, I don't approve of murder and most people don't so they make laws against it.
- 17A. You can say more than that you approve of the first person. There is a real moral difference in the quality of their actions and this does not depend on what people think or approve.
- 17B. If one person works hard for charity while another loafs, and another works hard for a terrorist organization, you can't say there is any real moral difference in the quality of these actions, or that the first is in fact morally better than the other. You can only say you approve of the first and disapprove of the other.

3-item face valid metaethics scale (YB3)

Yılmaz & Bahçekapili (2015b)

Source

Yılmaz, O., & Bahçekapili, H. G. (2015b). Without God, everything is permitted? The reciprocal influence of religious and meta-ethical beliefs. *Journal of Experimental Social Psychology*, 58(Suppl. A), 95-100.

Scale

(Not specified)

1. There are no ethical principles that are so important that they should be a part of any code of ethics.
2. Different types of moralities cannot be compared as to "rightness."
3. Questions of what is ethical for everyone can never be resolved since what is moral or immoral is up to the individual.

New Meta-Ethics Questionnaire (NMQ)

Yılmaz & Bahçekapili (2015a)

Source

Yılmaz, O., & Bahçekapili, H. G. (2015a). Without God, everything is permitted? The reciprocal influence of religious and meta-ethical beliefs. *Journal of Experimental Social Psychology*, 58, 95-100.

Scale

1 = strongly disagree, 7 = strongly agree

1. What is moral varies on the basis of context and society.
2. Moral standards are personal, therefore something morally acceptable to one person might be immoral for another person.
3. Since moral rules are not absolute, no definite judgments about them are possible.
4. Different cultures adopt different values and no moral law is right or wrong in an absolute sense.
5. We can agree on 'what is moral for everyone' because what is moral and immoral is self-evident.
6. If morality were to differ from person to person, it would be impossible for people to live together.
7. Since the moral laws I believe in are universally true, they can be applied to everyone in the world regardless of culture, race or religion.
8. If a moral law is right and good for others, it is also right and good for us.

3-Item Moral Objectivism Scale (MO3)

Sarkissian & Phelan (2019)

Source

Sarkissian, H., & Phelan, M. (2019). Moral objectivism and a punishing God. *Journal of Experimental Social Psychology*, 80, 1-7.

Scale

(Not specified)

1. There exists a single moral code that is applicable to everyone, regardless of any individual person's beliefs or cultural identity.
2. If two people really disagree about a particular moral problem then at most one of them can be correct, since moral problems cannot have multiple correct answers.
3. It is possible to compare different cultures by a single, universal standard of moral rightness.

Moral Relativism Scale (MRS)

Collier-Spruel et al. (2019)

Source

Collier-Spruel, L., Hawkins, A., Jayawickreme, E., Fleson, W., & Furr, R. M. (2019). Relativism or tolerance? Defining, assessing, connecting, and distinguishing two moral personality features with prominent roles in modern societies. *Journal of Personality*, 87(6), 1170-1188.

Scale

[1 = Strongly Disagree, 5 = Strongly Agree]

1. Different people can have opposing views on what is moral and immoral without anyone being wrong
2. People can disagree on what is morally right without anyone being wrong
3. Two different cultures could have dissimilar moral rules and both be "right"
4. One's own culture determines whether that person's actions are "right" or "wrong"
5. The viewpoint of one's culture determines whether their actions are morally right
6. There is a moral standard that all actions should be held to, even if cultures disagree [reverse coded]
7. Each person is the final authority on whether his or her actions really are morally correct
8. An action is only morally wrong if a person believes it is morally wrong
9. There are moral rules that apply to everyone regardless of personal beliefs [reverse coded]
10. The same moral standards should be followed by people from all cultures [reverse coded]

Folk Moral Objectivism Scale (FMO)

Zijlstra (2019)

Source

Zijlstra, L. (2019). Folk moral objectivism and its measurement. *Journal of Experimental Social Psychology*, 84, 103807. <https://doi.org/10.1016/j.jesp.2019.04.005>

Scale

[1 = Strongly disagree, 6 = Strongly agree]

No truth subscale

1. Other than what people believe, are brought up to believe, or want to believe about it, there are no facts about what is morally right and wrong
2. All ideas about what is morally right and morally wrong are products of individuals, cultures, and communities and nothing more
3. What people believe to be morally right and wrong are merely social conventions that could have been different
4. It is an illusion to think that anything is really morally true or false

Relativism subscale

5. When two people have opposing beliefs about a moral issue, it is not necessarily the case that either or both are wrong
6. There is not one but many different answers to the question of what is morally right and wrong and these can be equally correct
7. What is ultimately morally right and wrong is different for people with different moral views and from different cultures and societies
8. What is morally right and wrong is relative to the moral beliefs of an individual, culture, or society

Universalism subscale

9. What is ultimately morally right or wrong is the same for all people at all times and places
10. Although people or cultures sometimes ignore moral concerns, moral norms apply anywhere and everywhere
11. What is morally right and wrong for me here and now is also morally right and wrong for people elsewhere, even for people living in different countries and part of different cultures
12. Despite the diversity of moral views between individuals, cultures, and societies, there are moral norms that should apply universally

Absolutism subscale

13. Although people disagree about what is morally right and wrong, I believe in the existence of specific moral principles that can settle any moral disagreement
14. Certain actions are morally wrong and they remain morally wrong even in the rare case that no one believes so
15. There are absolute moral rules that apply to all people, including those who do not acknowledge these principles
16. There is, in all circumstances, one correct answer about what is the morally right thing to do

Divine Command Theory (DCT) subscale

17. The correct answer to any moral issue can be found in a sacred book or text (for example, the Bible, the Qur'an, the Torah, or another)
18. The only actions that are ultimately morally right or wrong are those actions that God prescribes
19. God is the only true source of knowledge about what is morally right or wrong
20. Without the existence of God, nothing is truly morally right or wrong

Objectivity of Morality Scale (JRT5)

Johnson, Rodrigues, & Tuckett (2020)

Source

Johnson, S. G., Rodrigues, M., & Tuckett, D. (2021). Moral tribalism and its discontents: How intuitive theories of ethics shape consumers' deference to experts. *Journal of Behavioral Decision Making*, 34(1), 47-65.

Scale

(Not specified)

Normativity subscale

1. Every good person on earth, regardless of culture, holds these beliefs.
2. The truth of these beliefs is self-evident.
3. A society could not survive without its citizens holding these beliefs

Subjectivity subscale

4. If someone strongly disagreed with you about one of these beliefs, it is possible that neither you nor the other person are mistaken [reverse coded]
5. There are no clearly true or false answers to these questions. [reverse coded]

APPENDIX D:
Research materials

Open Science Statement

Data, code, and materials can be found on the Open Science Network at: <https://osf.io/2afky/>

Qualitative Comprehension Paradigms
(Open response)

- Study 1A: Reanalysis of Goodwin & Darley (2008) Experiment 1
- Study 1B: Reanalysis of Goodwin & Darley (2008) Experiment 2
- Study 1C: New test of disagreement paradigm
- Study 2: Interpreting other people's responses to metaethics stimuli
- Study 3: Charity & noncognitivism
- Study 4: Direct realism and relativism
- Study 5: Interpretation of the Moral Relativism Scale (MRS)
- Study 6: YB3 and New-Metaethics Questionnaire (NMQ)
- Study 7: Folk Moral Objectivism Scale (FMO)

Study 1A & 1B:

These studies were a reanalysis of the studies conducted by Goodwin and Darley (2008) and will not be presented here.

Study 1C:
New test of the disagreement paradigm

Study 1C

When two people disagree about a moral issue, do you think they can both be correct, or must at least one of them be incorrect?

- ☐ They can both be correct
- ☐ At least one of them must be incorrect

Please briefly explain why you chose this response.

Study 2:
Interpreting other people's responses to metaethics stimuli

(Note: These materials were included as part of a larger study. I have extracted the relevant materials from the study, but the actual formatting of the study was somewhat different. It included other measures prior to participants being presented with open response questions, so these questions did not immediately follow the measures, as they appear here.)

Study 2A

The respondent was asked the following question:

When two people disagree about a moral issue, do you think they can both be correct, or must at least one of them be incorrect?

John:

“When people disagree about a particular moral issue there can be at most only one correct answer”.

Think about the person who answered this question.

In your own words, what do you think the respondent means in the statement above?

Study 2B

The respondent was asked the following question:

When two people disagree about a moral issue, do you think they can both be correct, or must at least one of them be incorrect?

John:

“When people disagree about a particular moral issue each can be correct according to their own moral standards”.

Think about the person who answered this question.

In your own words, what do you think the respondent means in the statement above?

Study 3:
Charity & noncognitivism

(Note: These materials were included as part of a larger study. I have extracted the relevant materials from the study, but the actual formatting of the study was somewhat different. It included other measures prior to participants being presented with open response questions, so these questions did not immediately follow the measures, as they appear here.)

Study 3A

In your own words, what does it mean to say that “it is a fact that some charities do more good than others, not a matter of personal beliefs or values”?

Study 3B

In your own words, what does it mean to say that “the truth about which charities do the most good depends on the beliefs and values of each individual”?

Study 3C

In your own words, what does it mean to say that “there is no fact of the matter about which charities do the most good”?

Study 4:
Direct realism and relativism

Study 4A

[Objectivism | Realism | Murder]

In your own words, what does it mean to say that the truth of the moral claim “murder is morally wrong” is objective?

[Concrete | Relativism | Murder]

In your own words, what does it mean to say that the truth of the moral claim “murder is morally wrong” is relative?

[Concrete | Realism | Abortion]

In your own words, what does it mean to say that the truth of the moral claim “abortion is morally wrong” is objective?

[Concrete | Relativism | Abortion]

In your own words, what does it mean to say that the truth of the moral claim “abortion is morally wrong” is relative?

Study 4B

[Abstract | Realism]

In your own words, what does it mean to say that moral truth is objective?

[Abstract | Relativism]

In your own words, what does it mean to say that moral truth is relative?



**Study 5:
Interpretation of the Moral Relativism Scale (MRS)**

[instructions]

In this study you will be asked to rate how much you agree or disagree with a statement. Then you will be asked a few questions about the statement.

{-----Page break-----}

[Participants randomly assigned to one of the following statements for the measures below or to items on the moral tolerance scale (MTS) from Collier-Spruel et al. (2019) which isn't shown]

[Relativism 1]

Different people can have opposing views on what is moral and immoral without anyone being wrong.

[Relativism 2]

People can disagree on what is morally right without anyone being wrong.

[Relativism 3]

Two different cultures could have dissimilar moral rules and both be “right.”

[Relativism 4]

One’s own culture determines whether that person’s actions are “right” or “wrong.”

[Relativism 5]

The viewpoint of one’s culture determines whether their actions are morally right.

[Relativism 6] [Reverse-coded]

There is a moral standard that all actions should be held to, even if cultures disagree.

[Relativism 7]

Each person is the final authority on whether his or her actions really are morally correct.

[Relativism 8]

An action is only morally wrong if a person believes it is morally wrong.

[Relativism 9] [Reverse-coded]

There are moral rules that apply to everyone regardless of personal beliefs.

[Relativism 10] [Reverse-coded]

The same moral standards should be followed by people from all cultures.

[Agree]

Please rate how much you agree or disagree with the following statement.

[Statement]

[1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree]

[Why]

Please explain why you chose this response.

[Explain]

In your own words, please explain what this statement means:

[Statement]

**Study 6:
YB3 and New-Metaethics Questionnaire (NMQ)**

(Note: Participants were randomly assigned to items from the EPQ, NMQ, YB3, or FMO in one study, which was all collected at the same time, so the items collected in study 6 and 7 were conducted as part of the same study. Materials for the EPQ are not shown here)

[Participants randomly assigned to items from the YB3 or NMQ]

[YB3 1]

There are no ethical principles that are so important that they should be a part of any code of ethics.

[YB3 2]

Different types of moralities cannot be compared as to "rightness."

[YB3 3]

Questions of what is ethical for everyone can never be resolved since what is moral or immoral is up to the individual.

(Note: Items NMQ 2 and NMQ 5 were not run because they also appear on the EPQ, so rerunning them would be redundant)

[NMQ 1]

Since what is moral varies on the basis of context and society, there is no one true morality.

[NMQ 3]

Since moral rules are not true or false in an absolute sense, moral debates are bound to remain inconclusive.

[NMQ 4]

Different cultures may adopt different values and thus it is impossible to compare cultures on the basis of an objective standard.

[NMQ 6]

What makes it possible for people to live together in harmony is the fact that fundamental moral rules do not differ from person to person.

[NMQ 7]

Since moral laws are universally true, they can be applied to everyone in the world regardless of culture, race or religion.

[NMQ 8]

Fundamental moral principles are universally valid; therefore they can be transferred from one society to another without difficulty.

[Agree]

Please rate how much you agree or disagree with the following statement.

[Statement]

[1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree]

[Why]

Please explain why you chose this response.

[Explain]

In your own words, please explain what this statement means:

[Statement]

**Study 7:
Folk Moral Objectivism Scale (FMO)**

(Note: Participants were randomly assigned to items from the EPQ, NMQ, YB3, or FMO in one study, which was all collected at the same time, so the items collected in study 6 and 7 were conducted as part of the same study. Materials for the EPQ are not shown here)

[Participants randomly assigned to one of the following items from the FMO]

[FMO Relativism 1]

When two people have opposing beliefs about a moral issue, it is not necessarily the case that either or both are wrong.

[FMO Relativism 2]

There is not one but many different answers to the question of what is morally right and wrong and these can be equally correct

[FMO Relativism 3]

What is ultimately morally right and wrong is different for people with different moral views and from different cultures and societies

[FMO Relativism 4]

What is morally right and wrong is relative to the moral beliefs of an individual, culture, or society

[FMO Universalism 1]

What is ultimately morally right or wrong is the same for all people at all times and places

[FMO Universalism 2]

Although people or cultures sometimes ignore moral concerns, moral norms apply anywhere and everywhere

[FMO Universalism 3]

What is morally right and wrong for me here and now is also morally right and wrong for people elsewhere, even for people living in different countries and part of different cultures

[FMO Universalism 4]

Despite the diversity of moral views between individuals, cultures, and societies, there are moral norms that should apply universally

[FMO Absolutism 1]

Although people disagree about what is morally right and wrong, I believe in the existence of specific moral principles that can settle any moral disagreement

[FMO Absolutism 2]

Certain actions are morally wrong and they remain morally wrong even in the rare case that no one believes so

[FMO Absolutism 3]

There are absolute moral rules that apply to all people, including those who do not acknowledge these principles

[FMO Absolutism 4]

There is, in all circumstances, one correct answer about what is the morally right thing to do

[FMO DCT 1]

The correct answer to any moral issue can be found in a sacred book or text (for example, the Bible, the Qur'an, the Torah, or another)

[FMO DCT 2]

The only actions that are ultimately morally right or wrong are those actions that God prescribes

[FMO DCT 3]

God is the only true source of knowledge about what is morally right or wrong

[FMO DCT 4]

Without the existence of God, nothing is truly morally right or wrong

[Agree]

Please rate how much you agree or disagree with the following statement.

[Statement]

[1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree]

[Why]

Please explain why you chose this response.

[Explain]

In your own words, please explain what this statement means:

[Statement]

Quantitative Comprehension Paradigms

(Multiple choice, Checkbox, Likert scale)

- Study 1: Moral Relativism Scale (MRS)
- Study 2: Bush & Moss
- Study 3: Fisher et al. Explicit
- Study 4: Young & Durwin
- Study 5: Checkbox
- Study 6: Likert Scale

Study 1: Moral Relativism Scale (MRS)

[instructions]

INSTRUCTIONS

On the next page, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

These instructions will appear on the next page for your convenience.

{-----Page break-----}

[conditions] [randomized, between subjects, one per participant]

Below, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

[condition 1] [antirealism] [disagreement | relativism] [MRS]

“Different people can have opposing views on what is moral and immoral without anyone being wrong.”

- **Since moral truth is determined by what each individual believes, people who hold conflicting moral views can both be correct**
- Different people can have conflicting views about what is right or wrong without either of them having made an unjustified mistake
- People with different moral views can each be correct about some aspects of a moral issue without either being completely incorrect
- When people hold conflicting moral views sometimes there is no way to know which view is correct or incorrect
- People with opposing moral views cannot be wrong because it is not possible for people to be right or wrong about their moral views
- People have a right to their moral views even if other people hold opposing moral views

[condition 2] [antirealism] [disagreement | dependent truth] [MRS]

“The viewpoint of one’s culture determines whether their actions are morally right.”

- **The truth about whether an action is morally right can only be judged according to the standards of different cultures**

- People from different cultures can be justified in holding different moral beliefs
- Each culture adapts its views about what is morally right to fit with their traditions and meet local needs
- Since each culture has different views about what is morally right, a person's actions will be judged as moral by some cultures and immoral by other cultures
- Different cultures have the right to hold different moral beliefs, so it isn't acceptable to impose our moral values on them
- The attitudes of each person's culture influence whether their actions are judged as moral or immoral

[condition 3] [realism | universalism]

“There are moral rules that apply to everyone regardless of personal beliefs.”

- **The truth of some moral rules is not determined by what people believe is morally right or wrong**
- Some moral rules apply to all people regardless of the specific circumstances they are in
- Everyone should be held to the same moral standards even if they do not agree with those standards
- People should be held accountable for breaking some moral rules even if they believe their actions were justified
- We should judge everyone according to a strict moral code even if we believe that there ought to be an exception in some cases
- Some actions are always right or wrong regardless of the intentions of the person committing the action

[condition 4] [realism | universal/normative]

“The same moral standards should be followed by people from all cultures.”

- **There is a correct set of moral standards that does not depend on what cultures think is right or wrong**
- We would be better off if people from every culture agreed to follow the same moral standards
- All cultures should be held to the same moral standards, so we should not tolerate cultures with moral beliefs that conflict with those standards
- People from all cultures should adhere to the same moral standards even if they do not agree with those standards
- We should not consider a person's cultural background when we judge their actions as morally right or wrong
- No matter what culture you come from, there are certain things everyone should know are right or wrong

Study 2: Fisher et al. Explicit

[instructions]

Instructions

On the next page, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

These instructions will appear on the next page for your convenience.

{-----Page break-----}

[conditions] [random assignment] [one condition each]

Below, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

[condition 1] [realism] [explicit] [Fisher]

“Moral truth is objective, so there is an objectively true answer to whether violence should be seen as an appropriate response in conflicts.”

- **The truth about whether violence is an appropriate response in conflicts does not depend on the standards or values of individuals or cultures**
- Whether violence is an appropriate response in conflicts depends on the circumstances
- We should see violence as appropriate in all situations or no situations, rather than thinking it is appropriate in some situations but not others
- The truth about whether violence should be seen as an appropriate response in conflicts applies to everyone, regardless of their situation
- We can appeal to facts and data instead of our subjective opinions to determine whether violence is an appropriate response in conflicts
- Although people can be biased in judging actions as right or wrong, if we can set aside our biases we can judge in each case whether violence is an appropriate response to conflict

[condition 2] [relativism] [explicit] [Fisher]

“Moral truth is relative, so the truth about whether violence should be seen as an appropriate response in conflicts is relative.”

- **Claims about whether violence is an appropriate response to conflicts can only be true or false according to different moral frameworks**
- Different people and societies have different perspectives, so there more than one point of view about whether violence should be seen as an appropriate response in conflicts,

- There is no way to be sure whether violence is an appropriate response in conflicts because we don't have a single agreed-upon method for measuring the outcomes of violent responses
- Whether violence should be seen as an appropriate response in conflicts depends on the context, so it may be appropriate in some cases but not others
- Since each person and society has different views about whether violence is an appropriate response in conflicts is relative, the use of violence will be judged as moral by some people and immoral by other people
- There are no moral truths about whether violence should be seen as an appropriate response in conflicts, just different people's feelings about what's right or wrong

{-----Page break-----}

[comprehension check]

In response to the question on the previous page, which of the following were you doing?

- **Selecting the single statement which I think best matches the meaning of the bolded sentence**
- Selecting the single statement that I most agree with
- Selecting every statement that I agree with
- Selecting every statement which I think best matches the meaning of the bolded sentence

{-----Page break-----}

[age]

What is your age?

[gender]

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

Study 3: Young & Durwin

[instructions]

Instructions

On the next page, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

These instructions will appear on the next page for your convenience.

{-----Page break-----}

[conditions] [random assignment] [one condition each]

Below, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

[condition 1] [realism] [Y&D]

“Some things are just morally right or wrong, good or bad, wherever you happen to be from in the world.”

- **There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.**
- There are moral truths that are wrong in every time, place, and situation without exception
- There are some actions that are considered morally wrong in all societies
- Some moral rules apply to all people regardless of the specific circumstances they are in
- We should all be judged according to a strict moral code regardless of where we come from
- People from all cultures should adhere to the same moral standards even if they do not agree with those standards

[condition 2] [antirealism] [Y&D]

“Our morals and values are shaped by our culture and upbringing, so there are no absolute right answers to any moral question.”

- **Whether an answer to a moral question is correct or incorrect can only be judged according to the standards of different individuals or cultures**
- People have different life experiences and points of view, so we cannot impose a single moral standard on everyone without considering their particular circumstances
- No questions about what is morally right or wrong have a single correct answer that applies in all circumstances, because the same action can be right in some circumstances but wrong in others
- Each person’s views about what is morally right or wrong is shaped by their unique background and experiences, so we frequently reach different conclusions about moral questions

- There are no moral truths, just different opinions about what is right or wrong
- Everyone's perspective on morality is different, so there's no way to know for sure whether an action is always right or wrong

{-----Page break-----}

[comprehension check]

In response to the question on the previous page, which of the following were you doing?

- **Selecting the single statement which I think best matches the meaning of the bolded sentence**
- Selecting the single statement that I most agree with
- Selecting every statement that I agree with
- Selecting every statement which I think best matches the meaning of the bolded sentence

{-----Page break-----}

[age]

What is your age?

[gender]

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

Study 4: Bush & Moss Scale

[instructions]

Instructions

On the next page, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

These instructions will appear on the next page for your convenience.

{-----Page break-----}

[conditions] [random assignment] [one condition each]

Below, you will see a statement in bold and a set of multiple choice options. *Please choose the option that you believe is the best interpretation of the meaning of the statement.*

We are not asking you to choose the option that you agree with the most. We only want to know which of the options most closely matches the meaning of the statement. There are no correct or incorrect answers. We are only interested in what you think.

[condition 1] [realism] [B&M]

“There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.”

- **Moral truth does not depend on cultural standards or personal values**
- The truth about what is morally right or wrong applies to all people regardless of the specific circumstances they are in
- Everyone should be held to the same moral standards even if they do not agree with those standards
- There are objective ways to measure whether some moral rules have good or bad outcomes, so individuals and cultures who think differently could be mistaken
- People should be held accountable for breaking some moral rules even if they believe their actions were justified
- Some actions are always right or wrong regardless of the intentions of the person committing the action

[condition 2] [antirealism] [relativism] [B&M]

“The truth of all moral claims can vary depending on the moral standards of different individuals or cultures.”

- **Facts about what is morally right or wrong are made true by the standards and values of people and societies**
- No claims about what is morally right or wrong can be true in all circumstances, because the same action can be right in some circumstances but wrong in others
- Different people and societies have different beliefs about what is morally right or wrong
- Each society adapts its views about what is morally right to fit with their traditions and meet local needs
- There are no moral truths, just different opinions about what is right or wrong
- Since each person and society has different views about what is morally right or wrong, we can't be sure there is a single correct answer to every moral claim

{-----Page break-----}

[comprehension check]

In response to the question on the previous page, which of the following were you doing?

- **Selecting the single statement which I think best matches the meaning of the bolded sentence**
- Selecting the single statement that I most agree with
- Selecting every statement that I agree with
- Selecting every statement which I think best matches the meaning of the bolded sentence

{-----Page break-----}

[age]
What is your age?

[gender]
What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

Study 5: Checkbox

[Instructions]

Instructions

On the next page, you will see a sentence in bold and a set of additional statements. *Please select every statement which you believe matches the meaning of the bolded sentence.*

We are not asking you to choose the statements that you agree with. We only want to know how well each of these statements match the meaning of the bolded sentence. There are no correct or incorrect answers. We are only interested in what you think.

These instructions will appear on the next page for your convenience.

{-----Page break-----}

[conditions] [random assignment] [one condition each]

Below, you will see a sentence in bold and a set of additional statements. *Please select every statement which you believe matches the meaning of the bolded sentence.*

We are not asking you to choose the statements that you agree with. We only want to know how well each of these statements match the meaning of the bolded sentence. There are no correct or incorrect answers. We are only interested in what you think.

[condition 1] [antirealism] [relativism | disagreement] [MRS]

"Different people can have opposing views on what is moral and immoral without anyone being wrong."

- **Since moral truth is determined by what each individual believes, people who hold conflicting moral views can both be correct**
- Different people can have conflicting views about what is right or wrong without either of them having made an unjustified mistake
- People with different moral views can each be correct about some aspects of a moral issue without either being completely incorrect
- When people hold conflicting moral views sometimes there is no way to know which view is correct or incorrect
- People with opposing moral views cannot be wrong because it is not possible for people to be right or wrong about their moral views
- People have a right to their moral views even if other people hold opposing moral views

[condition 2] [antirealism] [relativism | dependent truth] [MRS]

"The viewpoint of one's culture determines whether their actions are morally right."

- **The truth about whether an action is morally right can only be judged according to the standards of different cultures**
- People from different cultures can be justified in holding different moral beliefs
- Each culture adapts its views about what is morally right to fit with their traditions and meet local needs
- Since each culture has different views about what is morally right, a person's actions will be judged as moral by some cultures and immoral by other cultures
- Different cultures have the right to hold different moral beliefs, so it isn't acceptable to impose our moral values on them
- The attitudes of each person's culture influence whether their actions are judged as moral or immoral

[condition 3] [realism] [B&M]

"There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise."

- **Moral truth does not depend on cultural standards or personal values**

- The truth about what is morally right or wrong applies to all people regardless of the specific circumstances they are in
- Everyone should be held to the same moral standards even if they do not agree with those standards
- There are objective ways to measure whether some moral rules have good or bad outcomes, so individuals and cultures who think differently could be mistaken
- People should be held accountable for breaking some moral rules even if they believe their actions were justified
- Some actions are always right or wrong regardless of the intentions of the person committing the action

[condition 4] [realism] [universal conflation] [MRS]

"There are moral rules that apply to everyone regardless of personal beliefs."

- **The truth of some moral rules is not determined by what people believe is morally right or wrong**
- Some moral rules apply to all people regardless of the specific circumstances they are in
- Everyone should be held to the same moral standards even if they do not agree with those standards
- People should be held accountable for breaking some moral rules even if they believe their actions were justified
- We should judge everyone according to a strict moral code even if we believe that there ought to be an exception in some cases
- Some actions are always right or wrong regardless of the intentions of the person committing the action

[condition 5] [realism] [universal and normative conflation] [MRS]

"The same moral standards should be followed by people from all cultures."

- **There is a correct set of moral standards that does not depend on what cultures think is right or wrong**
- We would be better off if people from every culture agreed to follow the same moral standards
- All cultures should be held to the same moral standards, so we should not tolerate cultures with moral beliefs that conflict with those standards
- People from all cultures should adhere to the same moral standards even if they do not agree with those standards
- We should not consider a person's cultural background when we judge their actions as morally right or wrong
- No matter what culture you come from, there are certain things everyone should know are right or wrong

[condition 6] [antirealism] [relativism] [B&M]

“The truth of all moral claims can vary depending on the moral standards of different individuals or cultures.”

- **Facts about what is morally right or wrong are made true by the standards and values of people and societies**
- No claims about what is morally right or wrong can be true in all circumstances, because the same action can be right in some circumstances but wrong in others
- Different people and societies have different beliefs about what is morally right or wrong
- Each society adapts its views about what is morally right to fit with their traditions and meet local needs
- There are no moral truths, just different opinions about what is right or wrong
- Since each person and society has different views about what is morally right or wrong, we can't be sure there is a single correct answer to every moral claim

{-----Page break-----}

[Comprehension check]

In response to the questions above, which of the following were you doing:

- **Selecting every statement which I think matches the meaning of the bolded sentence**
- Selecting the single statement that I most agree with
- Selecting every statement that I agree with
- Selecting the single statement which I think best matches the meaning of the bolded sentence

{-----Page break-----}

[age]

What is your age?

[gender]

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

Study 6: Likert scale

[Instructions]

Instructions

On the next page, you will see a sentence in bold and a set of additional statements. *Please select the degree to which you agree or disagree that the additional statements match the meaning of the bolded sentence.*

We are not asking you to select how much you agree or disagree with the statements. We only want to know how well each of these statements match the meaning of the bolded sentence. There are no correct or incorrect answers. We are only interested in what you think.

These instructions will appear on the next page for your convenience.

{-----Page break-----}

[conditions] [random assignment] [one condition each] [each condition will include the same correct/incorrect response options that the item was paired with above, and will include two additional items: each realism item will include the other two realism items, and each antirealism item will include the other two antirealism items]

Below, you will see a sentence in bold and a set of additional statements. *Please select the degree to which you agree or disagree that the additional statements match the meaning of the bolded sentence.*

We are not asking you to select how much you agree or disagree with the statements. We only want to know how well each of these statements match the meaning of the bolded sentence. There are no correct or incorrect answers. We are only interested in what you think.

[condition 1] [realism] [universal] [MRS]

“There are moral rules that apply to everyone regardless of personal beliefs.”

[condition 2] [realism] [universal and normative conflation] [MRS]

“The same moral standards should be followed by people from all cultures.”

[condition 3] [realism] [B&M]

“There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.”

[condition 4] [antirealism] [relativism | disagreement] [MRS]

“Different people can have opposing views on what is moral and immoral without anyone being wrong.”

[condition 5] [antirealism] [relativism | dependent truth] [MRS]

“The viewpoint of one’s culture determines whether their actions are morally right.”

[condition 6] [antirealism] [relativism] [B&M]

“The truth of all moral claims can vary depending on the moral standards of different individuals or cultures.”

[1 = strongly disagree, 7 = strongly agree]

{-----Page break-----}

[Comprehension check]

In response to the questions above, which of the following were you doing:

- **Indicating how much I agree or disagree that each statement matches the meaning of the bolded sentence**
- Indicating my agreement with the single statement I agree with most, and indicating disagreement with every other statement
- Indicating how much I agree with each statement
- Indicating my agreement with the single statement which I think best matches the meaning of the bolded sentence, and indicating disagreement with all other statements

{-----Page break-----}

[age]

What is your age?

[gender]

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

Forced Choice & Spontaneous Theorizing Paradigm (Quantum mechanics)

Study 1: Quantum Mechanics

[instructions]

Instructions: Part 1

In this study, we will present you with different perspectives about an issue. The topic you are presented with may be unfamiliar to you. However, no prior knowledge is necessary or expected. Please answer to the best of your ability.

{-----Page break-----}

[quantum mechanics instructions]

We will now ask you some questions about quantum mechanics. On subsequent pages we will ask you follow-up questions about specific theories related to quantum mechanics.

[familiarity questions]

[fam_1] [randomized]

Each of the following terms is related to quantum mechanics. For each term, please indicate if you are familiar with that term. You may select more than one.

- The Copenhagen interpretation
- The Many Worlds interpretation
- Mach–Zehnder interferometer
- Quantum harmonic oscillator
- Dirac equation
- Finite potential well
- Time-symmetric theory
- Bohm's interpretation

[fam_2]

What is the highest level of formal education **in physics** that you have completed?

- ☐ None
- ☐ Some coursework in grade school
- ☐ Some coursework in high school
- ☐ Some coursework in college
- ☐ Undergraduate degree in physics (BS, BA, etc.)
- ☐ Some graduate coursework in physics
- ☐ Graduate degree in physics (MS, PhD, etc.)

[fam_3]

How well do you understand different interpretations of quantum mechanics?

[1 = Very poorly, 7 = Very well]

[/jam_4]

What level of expertise most closely reflects your knowledge of quantum mechanics?

- ☐ No knowledge (I may have seen or heard some of these terms before but I have no further understanding of quantum mechanics)
- ☐ Little knowledge (I know a little, so I understand about as much as someone who has not studied the topic or has at most some passing familiarity with it)
- ☐ Novice (I have some knowledge, could perform some related mathematical equations and could answer basic questions about theoretical physics)
- ☐ Proficient (I have significant knowledge, equivalent to a college student who has completed considerable coursework in physics and could perform complex mathematical equations, and could answer advanced questions about theoretical physics)
- ☐ Expert (I have extensive knowledge, equivalent to comprehensive understanding of the mathematical equations underlying quantum mechanics, and have mastered the most advanced mathematical equations and questions about theoretical physics)

[/position]

Do you have an opinion about which interpretation of quantum mechanics is correct?

- ☐ Yes
- ☐ No

{-----Page break-----}

Quantum mechanics

Quantum mechanics is a branch of physics that deals with atomic and subatomic phenomena. It is a highly successful theory that makes precise predictions and has inspired new technologies. Despite this success, scientists disagree about how to interpret it as a literal description of the world.

One disagreement concerns the nature of some subatomic events. These events appear to have more than one possible outcome, but when we measure them, we only observe one outcome. There are two competing explanations for what occurs when we observe one of these outcomes.

The Copenhagen Interpretation of Quantum Mechanics

According to the Copenhagen interpretation, we only observe one outcome because there is only one outcome. According to this view, such events have probabilistic outcomes. For example, a subatomic particle could have a 50% chance of moving to the left, and a 50% chance of moving to the right. We will not know which way it will move until we observe it. Once we observe it, there is an equal chance it ends up moving in either direction, like a coin flip. There is no way to know whether the particle would go to the left or the right in advance, because the result is random. Thus, the Copenhagen interpretation holds that some subatomic events have random outcomes.

The Many Worlds Interpretation of Quantum Mechanics

According to the Many Worlds interpretation, we observe one outcome but there is actually more than one outcome. According to this view, events do not have probabilistic outcomes. Whenever a subatomic particle could move to either the left or the right, instead of it only moving in one or the other direction, the universe splits into two different universes. The particle moves to the left in one of the universes, and to the right in the other universe. Each of those universes contains an exact copy of us and everything else in the universe, with the only difference being the direction the particle moves. In one universe, we observe the particle move to the left, while in the other, we observe the particle move to the right. The reason the particle seems to have moved to either the left or the right is because we are in one of these universes, and cannot observe the other universe. But there is a copy of us in the other universe that saw the particle move in the opposite direction. As a result, the outcome of subatomic events is never random. Instead, all possible ways the particle could move are realized in different universes. Since many subatomic events occur every second, there are countless other universes, each slightly different from the one you are in.

{-----Page break-----}

Instructions: Part 2

Now we will ask you what you think about the **Copenhagen** and **Many Worlds** interpretations of quantum mechanics.

[correct]

Think about the Copenhagen and Many Worlds interpretations of quantum mechanics. Which of these interpretations is correct?

- The **Copenhagen** interpretation is the correct interpretation of quantum mechanics. Whenever a subatomic event appears to have more than one possible outcome, only one of those outcomes randomly occurs.
- The **Many Worlds** interpretation is the correct interpretation of quantum mechanics. Whenever a subatomic event appears to have more than one possible outcome, all possible outcomes are realized. The universe splits into a different universe for each possible outcome, and each of those possibilities occurs in one of those universes.

[everyday]

How often does the topic of quantum mechanics come up in conversation in your everyday life?

- ☐ Never (It has never come up in conversation)
- ☐ Rarely (a few times in my life)
- ☐ Sometimes (several times a year)
- ☐ Often (several times a month)
- ☐ Frequently (several times a week)

[ordinary_agree]

Do you think most ordinary people have an opinion on how to interpret quantum mechanics?

- ☐ Yes, most ordinary people believe the Copenhagen interpretation is correct
- ☐ Yes, most ordinary people believe the Many Worlds interpretation is correct
- ☐ No, most ordinary people have no opinion about which interpretation is correct

[ordinary_claims]

When ordinary people make claims about events in the past, present, or future, do their claims presuppose a particular view about how to interpret quantum mechanics?

- ☐ Yes, when people talk about events they presuppose that the Copenhagen interpretation is correct
- ☐ Yes, when people talk about events they presuppose that the Many Worlds interpretation is correct
- ☐ No, when people talk about events they do not presuppose any particular interpretation of quantum mechanics is correct

[ordinary_practical]

How much of an impact do people's opinions about how to interpret quantum mechanics have on their everyday life? (For instance, what kinds of decisions they make and how successful they are in pursuing their goals)

- ☐ No impact
- ☐ Very little impact
- ☐ Some impact
- ☐ Moderate impact
- ☐ Enormous impact

{-----Page break-----}

Please indicate which of the following statements about the Copenhagen and Many Worlds interpretations is true or false.

[comp_1]

According to the Copenhagen interpretation, some events have random outcomes.

- **True**
- False

[comp_2]

According to the Many Worlds interpretation, if we observe a subatomic particle go to the left, there is another universe where it went to the right.

- **True**
- False

[comp_3]

According to the Copenhagen interpretation, events can have more than one outcome.

- True
- **False**

[comp_4]

According to the Many Worlds interpretation, we can observe all possible outcomes of subatomic events.

- True
- **False**

[comp_5]

Quantum mechanics is a branch of physics that primarily studies very large objects, such as planets and stars.

- True
- **False**

[comp_6]

Experts who study quantum mechanics disagree about which interpretation is correct.

- **True**
- False

{-----Page break-----}

[forced_choice]

In this study, you were asked to choose between the Copenhagen interpretation and the Many Worlds interpretation. However, it is possible that you do not agree with either of these positions. Did you feel that you had to choose between one of these two answers, even though it did not accurately reflect what you thought?

- ☐ Yes
- ☐ No

[age]

What is your age?

[gender]

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

Third Person & Normative Entanglement Paradigms

- Study 1: Third Person Paradigm (pilot)
- Study 2: Third Person Paradigm with concrete moral issues
- Study 3: Third Person Paradigm (nonmoral domains)
- Study 4: Normative entanglement

Study 1: Third Person Paradigm (pilot)

[instructions]

[other conditions] [randomized, between subjects, one per participant]

[condition 1] [no statement]

Instructions: Part 1

Please read the following questions carefully. You will be asked several questions about what you think **a typical person in your society** believes. Try to interpret the statements in as normal and non-exceptional a way as possible.

Think about a typical person in your society. Please rate how likely or unlikely it is that they hold the following beliefs.

[condition 2] [wrong]

Instructions: Part 1

You will be asked several questions about what you think **a typical person in your society** believes. Try to interpret the statements in as normal and non-exceptional a way as possible.

Scenario

A person is having a discussion about moral issues. During this discussion the person says:

"That is morally wrong."

[condition 3] [right]

Instructions: Part 1

You will be asked several questions about what you think **a typical person in your society** believes. Try to interpret the statements in as normal and non-exceptional a way as possible.

Scenario

A person is having a discussion about moral issues. During this discussion the person says:

"That is morally right."

Think about the person in this scenario. Please rate how likely or unlikely it is that they hold the following beliefs.

[All conditions]

[other] [objectivism 1]

They believe that when two people disagree about a moral issue at least one of them must be incorrect.

[other] [objectivism 2]

They believe there are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.

[other] [objectivism 3]

They believe moral truth is independent of cultural standards and personal beliefs.

[other] [universalism 1]

They believe that there is one standard of moral truth by which all people should be judged.

[other] [universalism 2]

They believe the same moral rules apply to everyone.

[other] [universalism 3]

They believe there is a single set of moral standards for all people and cultures.

[other] [relativism 1]

They believe that when two people disagree about a moral issue they can both be correct.

[other] [relativism 2]

They believe that moral truth is simply what is true to each individual.

[other] [relativism 3]

They believe that things are only morally right or wrong according to different points of view.

[other] [noncognitivism 1]

They believe that moral views cannot be true or false.

[other] [noncognitivism 2]

They believe that judgments about whether an action was morally right or wrong can be correct or incorrect.

[other] [noncognitivism 3]

They believe that judgments about morality are about what is true or false and not just expressions of feelings.

[1 = Very unlikely, 2 = Moderately unlikely, 3 = Somewhat unlikely, 4 = Neither likely nor unlikely, 5 = Somewhat likely, 6 = Moderately likely, 7 = Very likely]

{-----Page break-----}

[self condition]

[instructions]

Instructions: Part 2

Now we will ask you some questions about what **you** believe about the statements we just asked about.

{-----Page break-----}

[instructions, repeated]

Instructions: Part 2

Now we will ask you some questions about what **you** believe about the statements we just asked about.¹¹⁷

How strongly do you agree or disagree with the following statements?

[self] [objectivism 1]

When two people disagree about a moral issue at least one of them must be incorrect.

¹¹⁷ Instructions were provided on their own page once, between two page breaks, in order to prompt participants to recognize that the questions are shifting from judgments about someone else to judgments about themselves. Instructions were then repeated on the page with the items. That is why these same instructions appear twice here, with page breaks.

[self] [objectivism 2]

There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.

[self] [objectivism 3]

Moral truth is independent of cultural standards and personal beliefs.

[self] [universalism 1]

There is one standard of moral truth by which all people should be judged.

[self] [universalism 2]

The same moral rules apply to everyone.

[self] [universalism 3]

There is a single set of moral standards for all people and cultures.

[self] [relativism 1]

When two people disagree about a moral issue they can both be correct.

[self] [relativism 2]

Moral truth is simply what is true to each individual.

[self] [relativism 3]

Things are only morally right or wrong according to different points of view.

[self] [noncognitivism 1]

Moral views cannot be true or false.

[self] [noncognitivism 2]

Judgments about whether an action was morally right or wrong can be correct or incorrect.

[self] [noncognitivism 3]

Judgments about morality are about what is true or false and not just expressions of feelings.

[1 = Strongly disagree, 2 = Moderately disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Moderately agree, 7 = Strongly agree]

{-----Page break-----}

[demographics]

[gender]

What is your gender?

☐ Male

☐ Female

☐ Other

[age]

What is your age?

[politics]

Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?

[1 = Extremely liberal, 2 = Liberal, 3 = Slightly liberal, 4 = Moderate, 5 = Slightly conservative, 6 = Conservative, 7 = Extremely conservative]

[religiosity]

How religious do you consider yourself?

[1 = Not religious at all, 7 = Very religious]

Study 2: Third Person Paradigm with concrete moral issues

[instructions] [all conditions]

Instructions: Part 1

On the pages that follow you will be presented with scenarios describing people and will then be asked how likely you think it is that they believe certain statements. Please read these scenarios and statements carefully. There are no right or wrong answers, please just provide your thoughtful first response.

{-----Page break-----}

[condition 1] [no statement]

Please read the following questions carefully. You will be asked several questions about what you think a typical person in your society believes. Try to interpret the statements in as normal and non-exceptional a way as possible.

Think about a typical person in your society. Please rate how likely or unlikely it is that they hold the following beliefs.

[metaethics measures]

[other] [objectivism 3]

They believe moral truth is independent of cultural standards and personal beliefs.

[other] [universalism 2]

They believe the same moral rules apply to everyone.

[other] [relativism 3]

They believe that things are only morally right or wrong according to different points of view.

[1 = Very unlikely, 2 = Moderately unlikely, 3 = Somewhat unlikely, 4 = Neither likely nor unlikely, 5 = Somewhat likely, 6 = Moderately likely, 7 = Very likely]

[character measures]

[moral character]

How would you rate the moral character of this person?

[1 = Very morally bad, 2 = Moderately morally bad, 3 = Somewhat morally bad, 4 = Neither morally good nor morally bad, 5 = Somewhat morally good, 6 = Moderately morally good, 7 = Very morally good]

[empathy]

How empathic would you consider this person?

[1 = Not empathic at all, 7 = Very empathic]

[moral seriousness]

How seriously do you think this person takes morality?

[1 = Not seriously at all, 7 = Very seriously]

[partner preference]

Please rate the extent to which this person would be a good person to have as a social partner (such as a coworker, neighbor, or close friend).

[1 = Not good at all, 7 = Very good]

{-----Page break-----}

[concrete conditions] [order randomized, within subjects, all participants assigned to all conditions]

[condition 2] [meat]

Scenario

A person is having a discussion about moral issues. During this discussion the person says:

“It is morally wrong to eat meat.”

Think about the person who made this statement. Please rate how likely or unlikely it is that they hold the following beliefs.

[metaethics measures]

[Same as condition 1]

[character measures]

[Same as condition 1]

{-----Page break-----}

[condition 3] [grade]

Scenario

A person is having a discussion about moral issues. During this discussion the person says:

“It is morally wrong for a professor to give a bad grade to a student just because they dislike the student.”

Think about the person who made this statement. Please rate how likely or unlikely it is that they hold the following beliefs.

[metaethics measures]

[Same as condition 1]

[character measures]

[Same as condition 1]

{-----Page break-----}

[condition 4] [mock]

Scenario

A person is having a discussion about moral issues. During this discussion the person says:

“It is morally wrong for a person to go to a funeral to mock the deceased person in front of their family.”

Think about the person who made this statement. Please rate how likely or unlikely it is that they hold the following beliefs.

[metaethics measures]

[Same as condition 1]

[character measures]

[Same as condition 1]

{-----Page break-----}

[condition 5] [alcohol]

Scenario

A person is having a discussion about moral issues. During this discussion the person says:

“It is morally wrong for a woman who knows she is pregnant to drink alcohol.”

Think about the person who made this statement. Please rate how likely or unlikely it is that they hold the following beliefs.

[metaethics measures]

[Same as condition 1]

[character measures]

[Same as condition 1]

{-----Page break-----}

[self condition]

[instructions]

Instructions: Part 2

Now we will ask you some questions about what **you** believe about the statements we just asked about.

{-----Page break-----}

[instructions, repeated]

Instructions: Part 2

Now we will ask you some questions about what **you** believe about the statements we just asked about.

How strongly do you agree or disagree with the following statements?

[self] [objectivism 3]

Moral truth is independent of cultural standards and personal beliefs.

[self] [universalism 2]

The same moral rules apply to everyone.

[self] [relativism 3]

Things are only morally right or wrong according to different points of view.

[1 = Strongly disagree, 2 = Moderately disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Moderately agree, 7 = Strongly agree]

{-----Page break-----}

[demographics]

[age]

What is your age?

[gender]

What is your gender?

☐ Male

☐ Female

☐ Other

Study 3: Third Person Paradigm (nonmoral domains)

[instructions]

Instructions: Part 1

Please read the following questions carefully. You will be asked several questions about what you think a typical person in your society believes. Try to interpret the statements in as normal and non-exceptional a way as possible. There are no right or wrong answers, please just provide your thoughtful first response.

{-----Page break-----}

[conditions] [randomized, between subjects, one per participant]

[condition 1] [morality]

[other]

[instructions]

Think about a typical person in your society. Please rate how likely or unlikely it is that they hold the following beliefs.

[other] [moral] [objectivism 1]

They believe that when two people disagree about a moral issue at least one of them must be incorrect.

[other] [moral] [objectivism 2]

They believe there are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.

[other] [moral] [objectivism 3]

They believe moral truth is independent of cultural standards and personal beliefs.

[other] [moral] [universalism 1]

They believe that there is one standard of moral truth by which all people should be judged.

[other] [moral] [universalism 2]

They believe the same moral rules apply to everyone.

[other] [moral] [universalism 3]

They believe there is a single set of moral standards for all people and cultures.

[other] [moral] [relativism 1]

They believe that when two people disagree about a moral issue they can both be correct.

[other] [moral] [relativism 2]

They believe that moral truth is simply what is true to each individual.

[other] [moral] [relativism 3]

They believe that things are only morally right or wrong according to different points of view.

[1 = Very unlikely, 2 = Moderately unlikely, 3 = Somewhat unlikely, 4 = Neither likely nor unlikely, 5 = Somewhat likely, 6 = Moderately likely, 7 = Very likely]

{-----Page break-----}

[self]

[instructions]

Instructions: Part 2

Now we will ask you some questions about what **you** believe about the statements we just asked about.

[self] [moral] [objectivism 1]

When two people disagree about a moral issue at least one of them must be incorrect.

[self] [moral] [objectivism 2]

There are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.

[self] [moral] [objectivism 3]

Moral truth is independent of cultural standards and personal beliefs.

[self] [moral] [universalism 1]

There is one standard of moral truth by which all people should be judged.

[self] [moral] [universalism 2]

The same moral rules apply to everyone.

[self] [moral] [universalism 3]

There is a single set of moral standards for all people and cultures.

[self] [moral] [relativism 1]

When two people disagree about a moral issue they can both be correct.

[self] [moral] [relativism 2]

Moral truth is simply what is true to each individual.

[self] [moral] [relativism 3]

Things are only morally right or wrong according to different points of view.

[1 = Strongly disagree, 2 = Moderately disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Moderately agree, 7 = Strongly agree]

[condition 2] [taste]

[other]

[instructions]

Think about a typical person in your society. Please rate how likely or unlikely it is that they hold the following beliefs.

[other] [taste] [objectivism 1]

They believe that when two people disagree about whether a painting is beautiful at least one of them must be incorrect.

[other] [taste] [objectivism 2]

They believe there are facts about which foods taste good or bad that are true even if some people or societies think otherwise.

[other] [taste] [objectivism 3]

They believe the truth about which food, music, and art are best is independent of cultural standards and personal beliefs.

[other] [taste] [universalism 1]

They believe that there is one standard of aesthetic truth by which all art and music should be judged.

[other] [taste] [universalism 2]

They believe that all art and music is subject to the same standards for judging whether it is good or bad.

[other] [taste] [universalism 3]

They believe there is a single standard for judging how good or bad food tastes that applies to all people and cultures.

[other] [taste] [relativism 1]

They believe that when two people disagree about whether a work of art is beautiful they can both be correct.

[other] [taste] [relativism 2]

They believe that truths about which people, landscapes, and architecture are most beautiful is simply what is true to each individual.

[other] [taste] [relativism 3]

They believe that claims about what is beautiful or ugly can only be true according to different points of view.

[1 = Very unlikely, 2 = Moderately unlikely, 3 = Somewhat unlikely, 4 = Neither likely nor unlikely, 5 = Somewhat likely, 6 = Moderately likely, 7 = Very likely]

{-----Page break-----}

[self]

[instructions]

[self] [taste] [objectivism 1]

When two people disagree about whether a painting is beautiful at least one of them must be incorrect.¹¹⁸

[self] [taste] [objectivism 2]

There are facts about which foods taste good or bad that are true even if some people or societies think otherwise.

[self] [taste] [objectivism 3]

The truth about which food, music, and art are best is independent of cultural standards and personal beliefs.

[self] [taste] [universalism 1]

There is one standard of aesthetic truth by which all art and music should be judged.

[self] [taste] [universalism 2]

All art and music is subject to the same standards for judging whether it is good or bad.

[self] [taste] [universalism 3]

There is a single standard for judging how good or bad food tastes that applies to all people and cultures.

[self] [taste] [relativism 1]

When two people disagree about whether a work of art is beautiful they can both be correct.

[self] [taste] [relativism 2]

Truths about which people, landscapes, and architecture are most beautiful is simply what is true to each individual.

[self] [taste] [relativism 3]

Claims about what is beautiful or ugly can only be true according to different points of view.

[1 = Strongly disagree, 2 = Moderately disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Moderately agree, 7 = Strongly agree]

[condition 3] [social convention]

[other]

[instructions]

Think about a typical person in your society. Please rate how likely or unlikely it is that they hold the following beliefs.

¹¹⁸ Note that in the study conducted and reported here, the wrong item was used for [self] [taste] [objectivism 1]. In place of the correct measure, which is featured here, I mistakenly used [self] [taste] [relativism 3].

[other] [convention] [objectivism 1]

They believe that when two people disagree about cultural practices, such as appropriate work attire, that at least one of them must be incorrect.

[other] [convention] [objectivism 2]

They believe there are facts about which customs and conventions are best even if some people or societies believe otherwise.

[other] [convention] [objectivism 3]

They believe that there are facts about which social conventions (such as shaking hands or making eye contact) are correct regardless of the preferences or values of different individuals or cultures.

[other] [convention] [universalism 1]

They believe that all people should adopt the same social conventions and customs, regardless of their culture or background.

[other] [convention] [universalism 2]

They believe the same social conventions and customs apply to everyone.

[other] [convention] [universalism 3]

They believe the same social norms (such as rules about how to greet others and how to dress) apply to everyone in the world.

[other] [convention] [relativism 1]

They believe that when two people disagree about which social conventions (such as greeting others by bowing or shaking hands) are correct they can both be correct.

[other] [convention] [relativism 2]

They believe that truths about social conventions are simply what is true according to different cultures and social groups.

[other] [convention] [relativism 3]

They believe that facts about which customs and practices are appropriate can vary according to the standards of different communities.

[1 = Very unlikely, 2 = Moderately unlikely, 3 = Somewhat unlikely, 4 = Neither likely nor unlikely, 5 = Somewhat likely, 6 = Moderately likely, 7 = Very likely]

[self]

[instructions]

[self] [convention] [objectivism 1]

When two people disagree about cultural practices, such as appropriate work attire, at least one of them must be incorrect.¹¹⁹

[self] [convention] [objectivism 2]

There are facts about which customs and conventions are best even if some people or societies believe otherwise.

[self] [convention] [objectivism 3]

There are facts about which social conventions (such as shaking hands or making eye contact) are correct regardless of the preferences or values of different individuals or cultures.

[self] [convention] [universalism 1]

All people should adopt the same social conventions and customs, regardless of their culture or background.

[self] [convention] [universalism 2]

The same social conventions and customs apply to everyone.

[self] [convention] [universalism 3]

The same social norms (such as rules about how to greet others and how to dress) apply to everyone in the world.

[self] [convention] [relativism 1]

When two people disagree about which social conventions (such as greeting others by bowing or shaking hands) are correct they can both be correct.

[self] [convention] [relativism 2]

Truths about social conventions are simply what is true according to different cultures and social groups.

[self] [convention] [relativism 3]

Facts about which customs and practices are appropriate can vary according to the standards of different communities.

[1 = Strongly disagree, 2 = Moderately disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Moderately agree, 7 = Strongly agree]

[condition 4] [facts]

¹¹⁹ Note that the version of the item used in Chapter 6, Study 3 had an error. The item used there stated, “When two people disagree about cultural practices, such as appropriate work attire, [that] at least one of them must be incorrect.” The inclusion of “that” is a mistake that should be excluded from future uses of this measure.

[other]

[instructions]

Think about a typical person in your society. Please rate how likely or unlikely it is that they hold the following beliefs.

[other] [facts] [objectivism 1]

They believe that when two people disagree about matters of science or history that at least one of them must be incorrect.

[other] [facts] [objectivism 2]

They believe there are facts about science that are true even if some people or societies believe otherwise.

[other] [facts] [objectivism 3]

They believe that claims that describe what the world is like, such as the claims made in geography, history, and physics, are correct regardless of the preferences or values of different individuals or cultures.

[other] [facts] [universalism 1]

They believe that facts about science and math are true everywhere and at all times, rather than only being true in some places and times.

[other] [facts] [universalism 2]

They believe the way the world is (not simply the way it seems) is the same for everyone.

[other] [facts] [universalism 3]

They believe the same scientific facts (such as the chemical composition of water and the shape of the earth) should be recognized by everyone in the world.

[other] [facts] [relativism 1]

They believe that when two people disagree about matters of science or history, they can both be correct.

[other] [facts] [relativism 2]

They believe that truths about what the world is like are simply what is true according to different cultures and social groups.

[other] [facts] [relativism 3]

They believe that if the people from a particular society share a belief about what the world is like, their beliefs make that claim about what the world is like true.

[1 = Very unlikely, 2 = Moderately unlikely, 3 = Somewhat unlikely, 4 = Neither likely nor unlikely, 5 = Somewhat likely, 6 = Moderately likely, 7 = Very likely]

[self]

[instructions]

[self] [facts] [objectivism 1]

When two people disagree about matters of science or history at least one of them must be incorrect.¹²⁰

[self] [facts] [objectivism 2]

There are facts about science that are true even if some people or societies believe otherwise.

[self] [facts] [objectivism 3]

Claims that describe what the world is like, such as the claims made in geography, history, and physics, are correct regardless of the preferences or values of different individuals or cultures.

[self] [facts] [universalism 1]

Facts about science and math are true everywhere and at all times, rather than only being true in some places and times.

[self] [facts] [universalism 2]

The way the world is (not simply the way it seems) is the same for everyone.

[self] [facts] [universalism 3]

The same scientific facts (such as the chemical composition of water and the shape of the earth) should be recognized by everyone in the world.

[self] [facts] [relativism 1]

When two people disagree about matters of science or history, they can both be correct.

[self] [facts] [relativism 2]

Truths about what the world is like are simply what is true according to different cultures and social groups.

[self] [facts] [relativism 3]

If the people from a particular society share a belief about what the world is like, their beliefs make that claim about what the world is like true.

[1 = Strongly disagree, 2 = Moderately disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Moderately agree, 7 = Strongly agree]

{-----Page break-----}

[demographics]

¹²⁰ Note that the version of the item used in Chapter 6, Study 3 had an error. The item used there stated, “When two people disagree about matters of science or history [that] at least one of them must be incorrect.” The inclusion of “that” is a mistake that should be excluded from future uses of this measure.

[age]
What is your age?

[gender]
What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

Study 4: Normative entanglement

[conditions] [randomized, between subjects, one per participant]

[instructions]

Instructions: Part 1

In the scenario below, you will be presented with a discussion between two people. You will then be asked several questions about what you think about a particular person, given what they say in a discussion.

Scenario

Alex and Sam are having a discussion about moral issues. During this discussion, the following interaction takes place:

[condition specific wording]

Think about what Sam said in this scenario. Please rate how likely or unlikely it is that Sam holds the following beliefs.

[condition 1] [realism | funeral]

Alex: "Sam, do you believe it is objectively wrong for a person to go to a funeral to mock the deceased person in front of their family?"

Sam: "Yes, I believe it is objectively wrong for a person to go to a funeral to mock the deceased person in front of their family."

[condition 2] [realism | alcohol]

Alex: "Sam, do you believe it is objectively wrong for a woman who knows she is pregnant to drink alcohol?"

Sam: "Yes, I believe it is objectively wrong for a woman who knows she is pregnant to drink alcohol."

[condition 3] [realism | torture]

Alex: "Sam, do you believe it is objectively wrong to torture and kill a baby just for fun?"

Sam: "Yes, I believe it is objectively wrong to torture and kill a baby just for fun."

[condition 4] [antirealism | no cancellation | funeral]

Alex: "Sam, do you believe it is objectively wrong for a person to go to a funeral to mock the deceased person in front of their family?"

Sam: "No, I don't believe it is objectively wrong for a person to go to a funeral to mock the deceased person in front of their family."

[condition 5] [antirealism | no cancellation | alcohol]

Alex: "Sam, do you believe it is objectively wrong for a woman who knows she is pregnant to drink alcohol?"

Sam: "No, I don't believe it is objectively wrong for a woman who knows she is pregnant to drink alcohol."

[condition 6] [antirealism | no cancellation | torture]

Alex: "Sam, do you believe it is objectively wrong to torture and kill a baby just for fun?"

Sam: "No, I don't believe it is objectively wrong to torture and kill a baby just for fun."

[condition 7] [antirealism | cancellation | funeral]

Alex: "Sam, do you believe it is objectively wrong for a person to go to a funeral to mock the deceased person in front of their family?"

Sam: "No, I don't believe it is objectively wrong to go to a funeral to mock the deceased person in front of their family. But I do think it is morally wrong. I just don't believe morality is objective. I still find it to be terrible and I am deeply opposed to it."

[condition 8] [antirealism | cancellation | alcohol]

Alex: "Sam, do you believe it is objectively wrong for a woman who knows she is pregnant to drink alcohol?"

Sam: “No, I don’t believe it is objectively wrong for a woman who knows she is pregnant to drink alcohol. But I do think it is morally wrong. I just don't believe morality is objective. I still find it to be terrible and I am deeply opposed to it.”

[condition 9] [antirealism | cancellation | torture]

Alex: “Sam, do you believe it is objectively wrong to torture and kill a baby just for fun?”

Sam: “No, I don’t believe it is objectively to torture and kill a baby just for fun. But I do think it is morally wrong. I just don't believe morality is objective. I still find it to be terrible and I am deeply opposed to it.”

[All conditions]

[Metaethics]

[Sam] [objectivism 1]

Sam believes that when two people disagree about a moral issue at least one of them must be incorrect.

[Sam] [objectivism 2]

Sam believes there are facts about what is morally right or wrong that are true even if some people or societies believe otherwise.

[Sam] [objectivism 3]

Sam believes moral truth is independent of cultural standards and personal beliefs.

[Sam] [universalism 1]

Sam believes that there is one standard of moral truth by which all people should be judged.

[Sam] [universalism 2]

Sam believes the same moral rules apply to everyone.

[Sam] [universalism 3]

Sam believes there is a single set of moral standards for all people and cultures.

[Sam] [relativism 1]

Sam believes that when two people disagree about a moral issue they can both be correct.

[Sam] [relativism 2]

Sam believes that moral truth is simply what is true to each individual.

[Sam] [relativism 3]

Sam believes that things are only morally right or wrong according to different points of view.

[1 = Very unlikely, 2 = Moderately unlikely, 3 = Somewhat unlikely, 4 = Neither likely nor unlikely, 5 = Somewhat likely, 6 = Moderately likely, 7 = Very likely]

[character judgments]

[moral character]

How morally good or bad is Sam?

[1 = Very morally bad, 2 = Moderately morally bad, 3 = Somewhat morally bad, 4 = Neither morally good nor bad, 5 = Somewhat morally good, 6 = Moderately morally good, 7 = Very morally good]

[trustworthiness]

How trustworthy is Sam?

[1 = Very untrustworthy, 2 = Moderately untrustworthy, 3 = Somewhat untrustworthy, 4 = Neither trustworthy nor untrustworthy, 5 = Somewhat trustworthy, 6 = Moderately trustworthy, 7 = Very trustworthy]

[empathy]

How empathic is Sam?

[1 = Not empathic at all, 7 = Very empathic]

[judgmentalness]

How judgmental is Sam?

[1 = Not judgmental at all, 7 = Very judgmental]

[commitment]

How committed is Sam to their moral values?

[1 = Not committed at all, 7 = Very committed]

[political views]

How liberal or conservative is Sam?

[1 = Very liberal, 2 = Moderately liberal, 3 = Somewhat liberal, 4 = Neither liberal nor conservative, 5 = Somewhat conservative, 6 = Moderately conservative, 7 = Very conservative]

[religiosity]

How religious is Sam?

[1 = Not religious at all, 7 = Very religious]

{-----Page break-----}

[demographics]

[age]

What is your age?

[gender]

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other

APPENDIX E: Metaethics & the moral conventional distinction

E1.1 The MCD is neither universal nor innate

The *moral/conventional task* (MCT) is the primary experimental paradigm used to support the claim that children and adults distinguish moral norms from conventional norms, and was developed for the explicit purpose of detecting this distinction in children. Before administering the MCT to participants, experimenters develop a list of prototypical transgressions of moral and conventional norms, which experimenters select based on their own *a priori* classification. Since the MCT was designed for children, prototypical moral norms have typically consisted of acts of aggression exhibited in school (e.g., one child hitting another child) and other common transgressions among children (e.g., refusing to share a toy; Smetana, 1981), though more recent versions include more adult-appropriate transgressions (e.g., stealing cars or inciting riots in Huebner et al. 2010, see also Fessler et al., 2015; Kelly et al., 2007). Since moral norms are thought of as non-arbitrary rules that apply to all communities and cannot be overridden by local authorities, they differ from prototypical conventional norms, which are conceived of as arbitrary social conventions that promote social coordination by reinforcing behavioral uniformity within a social group, regardless of whether those norms generalize to other groups (Nucci & Turiel, 1978). Like moral norms, conventional norms used in research have likewise tended to depict classroom behavior, such as designated seating in classrooms and putting toys away (Kelly et al., 2007; Nucci & Turiel, 1978).¹²¹

¹²¹ To provide one concrete example of the precise set of items used, Smetana (1981, p. 1334) used the following items:

Moral norms

- (1) one child hitting another child
- (2) a child not sharing a toy
- (3) a child shoving another child
- (4) a child throwing water at another child
- (5) a child taking another child's apple

Conventional norms

- (6) a child not participating in show and tell
- (7) a child not sitting in the designated place (on a rug) during story time
- (8) a child not saying grace before snack
- (9) a child putting a toy away in the incorrect place
- (10) a child not placing her belongings in the designated place

Participants are presented with descriptions of moral and conventional transgressions and are asked a series of questions designed to elicit the predicted pattern of responses for each:

- (1) *Does the transgression generalize to people in other times and places* (i.e., is it also impermissible for them)?
- (2) *Is the transgression wrong regardless of what authorities or custom dictate?*¹²²
- (3) *Why is the transgression impermissible?*¹²³

Kelly and Stich point out that over sixty studies have employed the MCT, with most finding that participants reliably judge prototypical moral transgressions to generalize to other times and places, to be independent of authority, and to be justified in terms of harm, justice, or rights, while prototypical conventional transgressions exhibit none of these traits. As they put it, prototypical moral norms exhibit a *signature response pattern*, such that these characteristics reliably cluster together (the *signature moral pattern*), while the absence of this pattern applies to conventional norms (the *signature conventional pattern*).

Evidence of the MCD not only appear to be highly replicable, but generalizes to a wide range of populations and age groups, ranging from children around three years old (Smetana, 1981; Smetana & Braeges, 1990) to adolescents (Hollis, Leis, & Turiel, 1986) and adults (Huebner, Lee, & Hauser, 2010), as well as children with autism¹²⁴ (Blair, 1996) and children from both urban and rural populations (Song, Smetana, & Kim, 1987), higher and low socioeconomic backgrounds, higher and

¹²² In practice, this is tested by asking if the transgression would be permissible if an authority figure stated that it was (e.g., Kelly et al., 2007; Nucci & Turiel, 1993; Tisak, Crane-Ross, & Tisak, 2000), if a law were passed that legalized the act (e.g., Huebner et al., 2010; Kelly et al., 2007), or if there were no rules against it (e.g., Nucci & Turiel, 1978; Smetana, 1981).

¹²³ Earlier versions of the MCD also asked about how *serious* a transgression was, with seriousness serving as a fourth characteristic distinguishing moral and conventional norms. However, seriousness could not be reliably associated with paradigmatic moral judgments, and was dropped from canonical versions of the MCD (Smetana, 1993; Stich, 2018, p. 554, endnote 6; 2019; Tisak & Turiel, 1988; Turiel, 1983). Some studies also ask about the amount of punishment transgressors deserve (e.g., Smetana, 1981). Stich (2019) points to a remark from Smetana that captures the exclusion of seriousness succinctly: “the severity of the transgression is not considered to be a formal criterion for distinguishing moral and conventional rules and transgressions” (Smetana, 1993, p. 117).

¹²⁴ Some evidence indicates that children with psychopathic tendencies and adult psychopaths do not distinguish moral and conventional norms (Blair, 1995, 1997; Blair et al., 1995), though these results have been challenged by a recent, modified version of the MCT that found that a population of incarcerated psychopaths did draw the MCD (Aharoni, Sinnott-Armstrong, & Kiehl, 2012). With so little data, it remains unclear whether psychopaths distinguish moral and conventional norms.

lower IQ, and better and worse living conditions at home, including homes where children are neglected or abused (Smetana, 1984). The MCD also appears in insular religious communities (Nucci & Turiel, 1993) and in a diverse array of cultures spanning every populated continent (Yau & Smetana, 2003; Hollos, Leis, & Turiel, 1986; Nucci, Camino, & Sapiro, 1996; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Song, Smetana, & Kim, 1987; for reviews, see Helwig, 2006; Killen, McGlothlin, & Lee-Kim, 2002; Nucci, 2001; Tisak, 1995; Turiel, 2002).

These findings seem to provide overwhelming evidence that the MCD is a ubiquitous feature of human cognition. There is little doubt that at least *some* prototypical normative transgressions conform to their respective signature response patterns in many populations. But it is less clear that the tendency to draw these distinctions adequately captures a genuine universal distinction between moral and conventional norms, much less an innate capacity for doing so. Several studies show that the signature response patterns don't emerge in every community or for every norm violation. In one of their own studies, Nucci and Turiel (1993) found that children from a Conservative Jewish¹²⁵ community in Chicago judged all of the religious rules they were given to be authority independent (e.g., working on the Sabbath, circumcision, and day of worship), despite these rules having little apparent relation to harm, justice, or rights. Likewise, Nisan (1987) found that Muslim Arab children in a traditional village in Israel treated all violations, including prototypical conventional norms (e.g., coed bathing and addressing teachers by first name) as authority independent (Kelly & Stich, 2007; but see Nisan, 1988; Turiel, Nucci, & Smetana, 1988). Conversely, a population of Colombian children treated both moral and conventional norms as contingent on authority (Ardila-Rey & Killen, 2001).

More recently, Kelly et al. (2007) found that many adult participants did not judge various harmful actions (e.g., spanking, physical abuse in the military) to be authority independent, and many

¹²⁵ Some references to this study (Kelly & Stich, 2007) describe the children as Orthodox Jews. However, Nucci and Turiel (1993) described the participants in their study as Conservative and distinguished this community from the Orthodox tradition (see Rosenthal, 1978).

judged slavery and corporeal punishment to be less serious transgressions when committed in the distant past than the present, suggesting that for some people, harmful and unjust acts do not generalize with the same strength to other cultures and historical periods. Unfortunately, there are some problems with this particular study. For instance, participants were slightly more likely to judge slavery in ancient Greece and Rome to be less bad than slavery in the southern United States 200 years ago. Yet Kelly et al. stated that slavery was important for both the economic and social systems of Greece and Rome, while they only mentioned the importance of slavery to economic system of the US. Coupled with the greater salience of historical knowledge of the latter, potential self-presentational and demand effects for not denouncing US slavery more thoroughly, and preexisting beliefs about the comparative seriousness of slavery in the ancient world compared to slavery in the southern states, participants may have simply reasoned that the two scenarios had substantive moral differences. Failing to regard slavery in the ancient world as equally morally bad may reflect the judgment that slavery was comparatively less severe than more recent instances of slavery. If so, this would not indicate a failure to generalize; it would simply reflect differences in the content of the moral actions themselves, rather than the cultural and temporal contexts in which they occurred. The same may hold true of their other scenario, involving corporal punishment of sailors. Sailors expecting and tolerating the practice in a culture which permits it may not feel that they are being treated unjustly or having their rights violated, whereas contemporary sailors presumably would object to being physically beaten. The cultural context in which an act occurs can influence how it is experienced, and this can hold true even for committed objectivists. Thus, once again, participants may have simply regarded the actions as having different moral standing, rather than being morally equivalent actions occurring in different contexts.

Although there may be credible alternatives to Kelly et al.'s (2007) account, their findings may still indicate a diminution in moral seriousness towards temporally distant moral acts and a willingness

to treat some prototypically moral acts as authority independent. Fessler et al. (2015) also offer compelling evidence that corroborates Kelly et al.'s findings. They presented transgressions involving harm, justice, and rights violations¹²⁶ to each of seven culturally distinct populations, from affluent urban populations in California to small-scale indigenous societies in South America, Fiji, Indonesia, and Melanesia, which exhibited varying degrees of social stratification. Participants were asked to judge how morally bad these actions were if they occurred in a distant time, a distant place, or if an appropriate local authority declared that the action was “not bad.” All societies judged actions as less bad when performed in the past, in other cultures, or when approved by authority figures, but they showed dramatic differences in degree, with some (e.g., the Yasawa of Fiji) showing relatively little change in attitude and others (residents of Storozhnitsa, a Ukrainian village) expressing substantial reductions.¹²⁷

Even early findings suggested that the signature response patterns of moral and conventional norms are neither stable nor reliable among young children. Gabennesch (1990) gathered a substantial collection of these findings. To take just one example, half of the second and fourth graders in one sample judged eating with your hands to be wrong independent of authority and to generalize to other societies (Carter & Patterson, 1982; see also Damon, 1977; Komatsu & Galotti, 1986; Lockhart, Abrahams, & Osherson, 1977; Miller & Bersoff, 1988, Tisak & Turiel, 1988). More recently, children in the United States judged disgust-inducing behaviors (e.g., pitting in a glass of water before drinking

¹²⁶ The seven items used described (adapted from Fessler et al., 2015, p. 3):

- (1) a man stealing a stranger's money
- (2) a man battering his wife without provocation
- (3) a man striking and injuring a friend after the friend unintentionally injured him
- (4) a man cheating a stranger in a market transaction
- (5) a man knowingly spreading a false rumour that his rival is a thief
- (6) the initiator of a fight bribing a witness to lie about who was at fault, resulting in the innocent party being punished
- (7) and a man raping an unfamiliar woman

¹²⁷ Surprisingly, urban Californians exhibited moderate reductions, which seems inconsistent with claims that Westerners uniformly draw sharp distinctions between moral and conventional norms (but see Piazza & Sousa, 2016 for a methodological critique and Fessler et al., 2016 for a reply).

it and picking your nose) to not only be authority independent, but to be serious transgressions that generalized to other populations (Nichols, 2004). College students did not judge the same acts to generalize to other populations, but likewise judged them to be serious and authority independent transgressions. Likewise, low-SES populations in the United States and Brazil treated victimless acts as serious moral violations, despite these behaviors appearing to be unrelated to harm, justice, and rights, such as cleaning the toilet with your nation's flag, or siblings kissing each other passionately in private (Haidt, Koller, & Dias, 1993).

These observations played a role in the development of Haidt's Moral Foundations Theory, which has gathered substantial evidence that some individuals and cultures moralize content domains other than harm, or the sorts of rights and justice violations that appear in Turiel's MCT, including violations related to purity, loyalty, and authority (Haidt, 2012; Haidt & Joseph, 2004; Graham et al., 2013). In fact, people appear capable of flexibly construing almost any otherwise trivial issue in either moral or nonmoral terms under the right circumstances, such as wearing a sweater-vest (Allidina & Cunningham, 2018; Van Bavel et al., 2012). This suggests that the scope of moralized beliefs may not be exclusively determined by their content, and that some individuals and cultures moralize beliefs that are not justified in terms of harm, justice, and rights. Nisan (1987) proposed one explanation for differences like these, arguing that, while people in the West may tend to cordon off norms associated with welfare, justice, and rights from social conventions, other cultures "may anchor norms in heteronomous commands, thereby diminishing this distinction" (p. 719). In support of this suggestion, Nisan appeals to Shweder, Mahapatra, and Miller (1987), whose interviews in India and the United States suggest that people in the former may not distinguish convention from morality because "the social order is not separated from the natural moral order. In these cultures, customary practices that we regard as conventions are viewed as part of the nature of things" (pp. 719-720). If so, then there may be no principled way for these cultures to distinguish conventional and moral

norms, since the two stem from the same fundamental source. This would explain why conservative Jewish children regarded their religious conventions as authority independent. If so, this would suggest that the MCD is not an innate feature of normative cognition, but is instead a culturally constructed distinction that is contingent on surrounding beliefs about the source and justification of norms. Since these beliefs may vary across cultures, the MCD may only emerge in some cultures.

Even if the signature moral and conventional patterns emerged with respect to some norms across most or even all cultures, this would not decisively establish that Turiel and colleagues had provided a satisfactory account of the moral domain. Turiel relies on an *a priori* conception of morality that defines it in advance as serious, authority independent, generalizable, concerned with harm, justice, and rights, and for Turiel (1983), morality *just is* “analytically independent of systems of social organization that coordinate interactions” (p. 39, as quoted in Quintelier et al., 2013, p. 222). In other words, moral norms are not mere social conventions *by definition*. This preconception about morality has led to a narrow and selectively generated set of transgressions that fit his and other researcher’s intuitions about what would represent prototypical moral and conventional norms, but excludes transgressions that might not neatly conform to their judgments. Sure enough, when Huebner et al. (2010) included a wider range of transgressions, they did find that moral items included still displayed the signature moral pattern, but conventional norms formed a continuum ranging from those displaying the signature conventional pattern to those exhibiting one or more qualities associated with the moral domain. This suggests that there may be no sharp dividing line between moral and conventional norms, despite Huebner and colleagues interpreting their findings as at least a partial vindication of the MCD.

Another issue with the *a priori* approach Turiel and colleagues take is that little effort is made to determine whether subjects *themselves* view these issues as moral. Evidence indicating that folk moralization is inconsistent with the MCD puts pressure on Turiel and colleagues to explain why they

are inconsistent. One way to respond to this is to argue that people lack introspective access to their own patterns of judgment, or otherwise fail to explicitly recognize and articulate a reliable distinction between categories of norms despite indirect questions revealing a pattern like the MCD. But absent such evidence, indications of inconsistencies between how people explicitly categorize norms and the MCD puts pressure on Turiel and proponents of the MCD to explain why these inconsistencies exist. For instance, when Wright, Cullum, and Schwab (2008) explicitly asked to categorize a long list of topics and behaviors as moral or nonmoral, participants overwhelmingly agreed that some topics were nonmoral, such as music preferences and exercise, but many social behaviors and many seemingly paradigm instances of moral behavior, many of which Turiel and colleagues would undoubtedly classify as moral or be at pains to insist were not moral beliefs, did not even approach unanimity. For example, the proportions below represent just a handful of the proportion of participants who judged the item in question to be moral:

Abortion	76%
Cheating on exams/papers	69%
Honesty	66%
Euthanasia	36%
Incest	74%
Rape	90%
Euthanizing disabled children	77%

(Adapted from Wright, Cullum, & Schwab, 2008, p. 1465)

Participants frequently moralized issues that do not unambiguously conform to the MCD, either, including eating pets (51% judged this to be a moral issue), pornography (18%), gay marriage (25%), sexual promiscuity (21%), preferring women/minorities (21%), and believing in God (16%, p. 1465). Wright, Grandjean, and McWhite (2013) found similar results in a second study, only this time participants were asked to judge whether the item in question was moral or fell into one of several other categories (e.g., “social,” “personal,” “scientific fact,” etc.). Once again, there was little

consensus on whether a given item was moral, with the item most people agreed was moral (racial discrimination) achieving only 67% agreement.

Similar studies find significant cultural differences in what people regard as moral or nonmoral. Chinese participants were more likely to include “uncivilized” behavior in their notion of morality, compared to the Western emphasis on harm (Buchtel et al., 2015), and Levine et al., (2021) report similar variation in propensity to categorize issues as moral or nonmoral among Christian, Mormon, Jewish, and nonreligious participants. Arguably, many people may differ in which actions they regard as having harmful consequences (or are unjust or violate rights), in which case, variation between participants could be captured by differences in belief about whether an action causes harm, is unjust, or violates someone’s rights (Gray, Schein, & Ward, 2014). However, if this were so it would still raise serious methodological problems for the way the MCT is designed. Existing data using the MCT relies on researcher’s own *a priori* assumptions about whether a given act is prototypically moral or not, rather than the attitudes of participants themselves. Even if the MCD did reliably emerge among diverse populations, if it is inconsistent with the explicit distinctions people themselves make between moral and nonmoral norms, it is unclear whether this pattern has earned the empirical credibility to supersede folk classification and justifiably claim to refer to the *moral* domain.

The MCD also focuses exclusively on moral *transgressions*, but it is implausible that “the moral domain” consists exclusively of moral violations. People care about positive moral duties, supererogatory acts, moral character, the moral status of institutions, laws, and other abstractions that don’t reduce to specific transgressions, and other potentially moralized concerns which may fall within a broader conception of the moral domain the MCD appears to exclude. Thus, even if people did distinguish moral *transgressions* from conventional transgressions, it’s not obvious that the same distinction would extend to other qualities and acts that seem to fall within a plausible conception of what the moral domain is supposed to entail.

Given these challenges, it is unclear whether the MCT establishes that people reliably distinguish a domain of norms that we may justifiably regard as *moral*, since it is unclear whether prototypical moral norms uniformly exhibit the signature moral response pattern. On the contrary, it appears many children and adults (a) moralize issues unrelated to harm, justice, and rights (b) often regard prototypically conventional norms as serious, authority independent, and generalizable (c) often judge prototypically moral norms to lack one or more of these qualities (d) do not agree on what counts as a moral or nonmoral norm or issue and (e) moralize more than just transgressions. The MCD fails to pick out a distinctive cluster of characteristics that reliably distinguish moral norms from conventional norms. As a result, the MCD literature does not provide adequate evidence that there are distinctively moral norms characterized by (among other things) universality, which obviates the need for Stanford (2018) and others to account for a purportedly innate capacity for the universality of moral thought.

E1.2 The MCD is not evidence of realism

I speculate in the main text about various reasons philosophers have not engaged with or conducted research on folk metaethics. I omitted the possibility that research on the moral/conventional distinction (MCD) has already confirmed that ordinary people are moral realists. The intuition that moral norms differ from nonmoral norms will probably seem compelling to most of us. Consider what *telling white lies* and *committing genocide* have in common. Very little, it would seem. But the moment we contrast them with *wearing crocs to a wedding* or *showing up to a meeting in our pajamas*, their relationship is obvious: despite the vast gulf in severity, white lies and genocide are both *moral* issues. Questionable fashion choices are not. One way to put this is that genocide and white lies fall within the *moral domain*, a distinct category of norms that we readily distinguish from nonmoral norms.

As discussed in the previous section, this intuitive distinction is supported by a substantial body of research pioneered by Turiel and others demonstrating that children¹²⁸ and adults reliably attribute a distinct cluster of properties to moral norms that they do not attribute to other norms (Huebner et al., 2010; Nucci, 1982; Nucci & Nucci, 1982; Nucci & Turiel, 1979; 1993; Nucci, Turiel, & Encarnacion-Gawrych, 1983; Smetana, 1981; Smetana & Braeges, 1990; Smetana, Kelly, & Twentyman, 1984; Smetana et al., 1999; Turiel, 1979; 1983). This research purportedly shows that people regard moral norms as (1) *authority independent* (the rule cannot be overturned by authorities), (2) *universal* (i.e. they apply to everyone, including people outside the participant's community), and (3) *justified in terms of harm, justice, and rights*, while conventional norms (such as table manners) display the inverse of these characteristics (i.e., authority dependence, non-generalizable, and not justified by appeal to harm, justice, or rights).

This moral/conventional distinction (MCD) has fueled claims that humans possess an innate, pancultural capacity for distinctively moral cognition (Dwyer, 1999, 2004; Joyce, 2006) characterized by the experience of moral norms as *objective moral truths* (Levy, 2005; Stanford, 2018). This use of *objective* means more or less what I mean by *realism*, in that *realism* is the view that there are "*objective*" (*stance-independent*) *moral facts*. Whatever Turiel's own interpretation of his findings, others have taken the MCD as evidence of folk realism. Nucci (2001) appeals to features of the MCD to argue that, from an early age, children "understand that it is objectively wrong to hurt others" (p. 86). Nichols and Folds-Bennett (2003) express a similar view, claiming that "the child's capacity to distinguish morality from convention shows that children regard moral violations as objectively wrong" (p. 23), and Kelly et al. (2007) claim that among the "core ideas that researchers in this tradition have advanced about moral rules" is that they have "an objective, prescriptive force" that is "not dependent on the authority

¹²⁸ As young as age two or three (Royzman, Leeman, & Baron, 2009; Smetana & Braeges, 1990).

of any individual or institution” (p. 118; for more examples see e.g., Colebrook & Sarkissian, 2018; Heiphetz & Young, 2017; Stanford, 2018).

Evidence of the MCD has also inspired more recent efforts to directly investigate folk realism. Both sets of findings purportedly demonstrate that children are realists, and that despite some increase in antirealism as people mature, adolescents and adults for the most part continue to be realists about the moral domain but not other evaluative domains, such as aesthetic and taste preferences (Beebe et al., 2015; Beebe & Sackris, 2016; Heiphetz & Young, 2017; Nichols & Folds-Bennett, 2003; Schmidt, Gonzalez-Cabrera, & Tomasello, 2017; Wainryb et al., 2004). Both lines of evidence have in turn prompted efforts to explain why this unique mode of cognition evolved, or at least why it is widespread (Dwyer, 1999; 2004; Dwyer, Huebner, & Hauser, 2010; Joyce, 2006, Ch. 4; Stanford, 2018). Jebari and Huebner (2018) summarize a recent example from Stanford (2018):

Stanford argues that human ultrasociality is possible because we treat moral obligations as part of an externally imposed moral order, which applies equally to all; because the experience of moral motivation feels *objective*, it automatically generates the demand that others be similarly motivated, and in populations of like-minded individuals, this yields correlated interactions that are less likely to be exploited. (p. 22, emphasis mine)

Yet Stanford’s claims rely in part on evidence of the universality of the MCD and the claim that it supports folk realism. Unfortunately, the MCD does not support Stanford or anyone else’s claims of a pancultural tendency towards moral realism. As a result, there is no need to put forward elaborate accounts of how and why a capacity for distinctively moral cognition evolved, or to explain why we regard moral norms as objective.

Unfortunately, two features of the MCD, *universality* and *authority independence*, are frequently cited as evidence that people are moral realists (e.g., Colebrook & Sarkissian, 2018; Kelly et al., 2007; Nichols, 2004; Nichols & Folds-Bennett, 2003; Nucci, 2001; Stanford, 2018).

Neither attribute entails nor even strongly suggests realism. Recall that realism is the claim that moral truths are made true by certain stance-independent facts about the world, rather than the

attitudes, values, preferences, or beliefs of different individuals, nor consensus among social groups (Shafer-Landau, 2003).¹²⁹ As Goodwin and Darley (2008) note, universality does not entail realism because “the question of whether ethical standards should apply to all cultures is a question about the *scope* of ethical standards, and is independent of the question of whether such standards and beliefs are objectively or subjectively true” (p. 1341). Realism concerns the source of moral truth: whether such standards are made true by facts that aren’t reducible to any particular standard of evaluation, i.e., they’re *stance independent*. In principle, subjectivists, cultural relativists, and noncognitivists could judge that others ought to conform to their values, even if they acknowledge that those values are not objectively true. There is no compelling reason to think antirealists must believe that their moral standards are only locally applicable, and ought not be universalized. They *may* think this, or they *may* not. Whether or not the antirealist does so is a contingent matter and cannot be assumed *a priori*. Contrary to popular misconception, cultural relativism does not require that one must be tolerant of other cultures or refrain from seeking to impose one’s moral values on other societies (Bush, 2016).

More generally, antirealist positions do not commit antirealists to any particular normative moral standards. Subjectivism, noncognitivist, error theory, *do not* entail that one’s moral standards are only locally applicable, e.g., that they only apply to oneself or members of your culture. Even error theorists, who deny that there are moral facts at all, can and often do believe that everyone ought to abide by the same moral standards, where *ought* is understood to convey their nonrealist moral position. Who one believes their moral standards ought to apply to simply has *nothing to do* with whether those moral norms are stance-independently true. Since universality is consistent with

¹²⁹ This should not be mistaken for the claim that people’s attitudes and values are morally irrelevant. Attitudes and other mental states can play a role in determining the moral status of an act even for objectivism. For instance, if it were an objective moral fact that one ought to treat others as they wish to be treated, how any given person wishes to be treated would depend on their preferences. However, the moral rule itself would not be true or false based on whether people believed or preferred it to be true.

antirealism, belief in the universality of moral norms cannot be used to infer objectivism. I'll cite myself as an example. I'm a moral antirealist. I don't believe there are any stance-independent moral facts. Nevertheless, I think that *nobody* should torture babies for fun, and that *everyone* should (generally) be honest and compassionate. I just don't think there are stance-independent normative facts that provide people with external reasons to comply with these moral standards.

It's also possible that judgments about the universality of a norm indicate inferences about what norms are likely to emerge in other communities. Since conventional norms can be at least somewhat arbitrary¹³⁰, there is little reason to expect precisely the same set of conventions to arise everywhere. Yet the universality of suffering and aversion to injustice could lead people to infer that other communities would have an interest in enforcing the same "moral" rules. If so, universality would not indicate objectivism, but a recognition that non-arbitrary norms are likely to arise in similar contexts. For example, if a child affirms that they think it would be wrong for children in distant times or places to hit each other or divide treats unequally, this would not necessarily indicate that they think these actions are *objectively* wrong; it could simply involve the recognition that other children tend to share the same subjective preferences, preferences which could very well be true on account of stance-*dependent* facts about shared psychological characteristics.¹³¹

For comparison, we would expect people in other societies to share similar food preferences, e.g., a preference for meat and sweet foods over dirt and feces, yet this would not demonstrate we are "gastronomic realists" that believe facts about which food is good or bad are true independent of our

¹³⁰ There may be non-arbitrary elements of social conventions, e.g., style of dress may be both a matter of convention and promote comfort or survival in particular climates.

¹³¹ Children may also be reluctant to state that a moral rule would not apply in other times and places due to a desire to appear to be "good," well-behaved children. If so, demand effects may inflate the apparent strength of the MCD. Haidt et al. (1993) found intriguing hints of this possibility, observing that:

"In pilot testing, we found that many children judged violations of social conventions, such as eating with one's hands, to be universally and unalterably wrong, across many wordings of the probe questions. We feared that some children thought they were being tested and were seeking to demonstrate that they were 'good' children by condemning all violations in the strongest possible terms." (p. 618)

stance or preference toward food (Loeb, 2003). In other words, merely expecting or presuming that other people would have similar standards, values, rules, or preferences does not entail that you think those rules reflect stance-independent normative or evaluative facts. In short, inferences about the universality of norms or evaluative attitudes are consistent with both realist and antirealist stances; they are *not* identical with realism and nor even good evidence of realism.

If universality does not establish objectivity, what about authority independence? This is a far better candidate, as it suggests that moral norms cannot be arbitrarily dispensed with on a whim, which could imply that moral norms are derived from some external source. Yet authority independence is also not great of realism because believing that moral norms can or cannot be changed by authorities are both consistent with realism *and* antirealism.

First, realism is compatible with believing (at least some) moral norms are authority-dependent in certain respects. For instance, a realist could believe that certain authority figures are in a position to make determinations about whether an action is permissible or impermissible, e.g., people could believe that they have an objective moral obligation to abide by the decrees of religious figures, the rulings of judges, or the moral rules set by the elders of the community. In such cases, these people would believe in a form of lower-order authority-dependence, whereby authority figures are endowed with the ability to determine or change moral rules.¹³² If so, they could also believe there is a higher-order objective fact of the matter that authorities have this ability (and perhaps objective facts about who is and isn't an authority). However, questions about whether authorities can change rules do not disambiguate a metaethical interpretation of authority-independence from the lower-order, normative form of authority independence I've described. As such, responses to questions about authority independence do not clearly and ambiguously prompt metaethical interpretations, but could instead

¹³² Perhaps with some constraints, e.g., they can make moral rules about who can get married, but not whether an incestuous marriage is acceptable.

prompt normative interpretations. If so, questions about authority independence may not be valid measures of higher-order (metaethical) conceptions of authority independence that could potentially serve as evidence of folk realism.

At the same time, antirealists may not endorse authority-dependence. One of the primary alternatives to realism is *relativism*, which treats the truth of moral claims as contingent on the standards of cultures (*cultural relativism*) or individual subjective values (*subjectivism*) (Bush, 2016). Both realism and relativism are compatible with denying that moral norms are dependent on authorities. For subjectivists, whether an action is moral depends on individual attitudes, not the judgments of authorities. Unless a subjectivist regards an authority figure as being in a position to do so, that figure simply lacks the ability to alter moral rules by fiat. Granted, it is *possible* for a subjectivist to outsource their moral position to authority figures by believing e.g., “My personal moral position is that I ought to abide by the moral judgments of authorities,” but they *need not* think this. And most people probably *don’t* think this. So it’d be odd to presume that someone who denies authority independence isn’t a realist, which is what one would have to do if one is to take agreement with authority independence to be evidence of realism.

Likewise, there is no reason why cultural relativists must think moral rules are authority dependent. For cultural relativists, moral rules are true or false relative to the standards of different cultures. It may simply not be part of a culture’s standards that authority figures can alter moral rules on a whim. Instead, they could believe that their moral standards cannot be changed, or that if they can be changed, that individuals lack the authority to do so. Perhaps changes to moral rules require cultural consensus, or novel exegesis of sacred texts, or a democratic referendum. It could be anything. And even if a person feels that *nothing* can change the moral rules, this still does not entail realism. An antirealist is under no obligation to regard their moral position as negotiable or subject to change based on social consensus or the public decree of authorities. One of the most persistent and egregious

mistakes researchers make in inferring realism from authority-independence is the failure to recognize that one's first-order moral beliefs may be rigid and non-negotiable *regardless* of whether that person is a realist or antirealist. I submit myself as an example: I am an antirealist. I don't believe there are any moral facts. And yet my response to questions used to establish the belief that moral norms are authority-independent are exactly the same as the realist: should people in other communities be able to change the moral rules so that hitting people or stealing their belongings is okay? No. If a teacher or authority figure says it would be okay to break these rules, or that these rules no longer exist, would that be okay? No. Do I think everyone, everywhere, should follow the same moral rules? Yes. I'm not unusual in this regard. Most moral antirealists have strong moral standards and values, just like realists do. There's no reason to think they're especially likely to be indifferent to people committing atrocities overseas or to think it would be morally permissible to kill their grandmother and bake her into a pie if Congress passed a bill declaring that they could.

Relativists could also believe morality is authority-dependent, but that the figure *in the situation the participant was presented with* lacked the requisite authority to change the rule. Thus, many studies using the moral-conventional task couldn't detect genuine authority-dependence even if it were present. Finally, there is no reason to think that noncognitivists must be committed to the notion that authorities can change moral rules. The belief that moral norms express nonpropositional attitudes simply does not entail authority-dependence.

In short, realism and antirealism are both conceptually consistent with authority-dependence and authority-independence, and their consistency with each is neither strained nor implausible. If there is an association between authority-independence and realism, it is at best only a contingent one, i.e., it could be that antirealism correlates with authority-independence, even if neither conceptually entails the other or should be presumed on *a priori* grounds to suggest the other. As a result, measures of authority-independence cannot distinguish realists from antirealists without independent empirical

evidence of a contingent connection of some form. This is because, if the belief that moral norms were authority independent indicated a commitment to realism, we should be able to distinguish moral realists from moral antirealists by how they respond to the MCT. Not only is there no evidence of such a connection, what little evidence there is suggests this is *not* the case. Nichols (2004) found that realists and antirealists distinguished between moral and conventional norms in the same way, which is hard to reconcile with claims that the MCD is evidence of folk realism.¹³³ Even if there is a correlation between belief in authority-independence and realism, it is implausible that they form such a tight connection that measures of the former can serve as proxies for the latter. The persistence of the belief that the MCD is evidence of realism is, in light of this, rather puzzling. Many researchers may have simply taken universality or authority independence to conceptually entail realism. This is definitely *not* the case. Still others may have presumed a contingent connection between features of the MCD and realism. Yet there is no obvious reason why universality or authority-independence should entail or even be an especially strong indication of a commitment to realism. I'm an antirealist, and I endorse both universalism and authority-independence: I think murder is *bad* and that there should be laws against it. I think it's bad and should be illegal regardless of whether it occurs in my country, a distant country, or an alien planet. And I don't think authority figures should have carte blanche to override rules against assaulting other people. Speaking on behalf of antirealists more generally, *almost all of us think this*.¹³⁴ The MCD is just as transparent and natural a distinction for me to draw as it is to most members of WEIRD populations, even for those of us who are cognizant of

¹³³ This is, of course, assuming that the measures used to classify participants as realists or antirealists were themselves valid. Perhaps they were not. This may be cold comfort for anyone treating the MCD as evidence of realism since methodological reasons for doubting the validity of direct measures of folk metaethical belief might also apply to the MCT.

¹³⁴ This *ought* to be obvious to anyone who spends a minute reflecting on what it would mean if only realists endorsed universality and authority-independence. Are those who make this error operating under the assumption that antirealists like me are *indifferent* to moral rules in faraway places? Or that we think that if judges and law enforcement declared it okay to chop our neighbors up with a hatchet that it would be fine to do so? It is so absurd on the face of it that I can only imagine that they haven't really adequately reflected on the implications of the MCD being diagnostic of realism. More generally, critics of antirealism often seem to underestimate the resources antirealists have available for accommodating conventional moral beliefs and attitudes.

being antirealists. Recognizing the MCD is simply and straightforwardly *not* an indicator of realism. It is time we retire this misconception.