Metaethical Revisionism:
A Response to Error Theory

Michael Prinzing
Word Count: 29,349

For Mom and Dad

Abstract

Error Theorists claim that there is a defect—an error—in our moral concepts. This error leads to their systematic reference failure, and the untruth of all substantive moral claims. I deny this and propose a Revisionist alternative. If the purported defect is present, then we should revise our moral concepts. Error Theorists claim that this is not possible, as the defect is a "non-negotiable" feature of the concepts. This final claim is, in my view, the most interesting part of their argument, and the focus of the thesis. Revisionist theories have been offered in a number of domains. Such views are sometimes claimed not to advance their respective debates because they succeed only in changing the subject. This objection is correct that revision is not always an option, but there are clearly some cases in which we can revise our concepts. A central task for the thesis is therefore to understand the conditions under which conceptual revision is possible. This means accounting for the phenomenon of conceptual "negotiability". If the view presented here is correct, it should be useful for Revisionists in any domain. Chapter 1 expounds the standard argument for Error Theory, clarifying three premises: *Conceptual*, *Ontological*, and *Non-Negotiable*. *Conceptual* and *Ontological* are granted for the sake of the argument. Chapter 2 offers my theory of conceptual negotiability. Chapter 3 applies that account, arguing against *Non-Negotiable*, the claim that moral concepts cannot be revised in the relevant way. Chapter 3 also discusses and defends Metaethical Revisionism as a plausible and appealing theoretical option. Finally, Chapter 4 offers my own (somewhat speculative) account of what is truly non-negotiable for moral concepts and considers some candidate metaethical revisions. The Conclusion presents a summary of the argument of the thesis (it might helpfully be read after the Introduction, and before Chapter 1).

Introduction

In the space of 30,000 words, a philosopher might do a couple of things. She might pick a very narrow topic, addressing some argumentative detail in great depth. Or, she might investigate an unexplored part of the theoretical landscape, blazing a new trail. For reasons of philosophical temperament, I've chosen the second, more ambitious (and, admittedly, riskier) option. This exploratory approach requires me to cover a lot of ground and prevents me from filling in all the details. I have attempted to address the most obvious concerns and questions. But there is certainly no room for it all. Nevertheless, I'm confident that the thesis performs two valuable services. It directs attention to an interesting, appealing, and previously unexplored theoretical option, and it sheds light on a popular, but remarkably under-explored general idea about concepts. It offers the beginnings, if not full details, of two views that merit serious consideration.

I open with a parable.[1] Ivan is a divine command theorist. He believes that moral facts are determined by God's decrees. However, Ivan comes to doubt God's existence, and eventually becomes an atheist. For a while, Ivan thinks that this means there are no moral facts. If there were any moral facts, they would be or entail facts about God's will. Since there is no such will, there are no such facts. "If God is dead…" he thinks. And yet, it is still quite important to Ivan how he and others behave. Ivan wants to continue discussing the topic. He also notices that the judgments and practices distinctive of morality could be grounded in some alternative, non-divine facts. Most of the moral claims he previously accepted could be understood, not as facts about God's will, but as facts about (say) ideal social arrangements with which no one could reasonably disagree,

[1] The idea for this parable comes from Vargas 2013.

the dictates of pure practical reason, or some such thing. Ultimately, Ivan concludes that his mistake wasn't in holding moral beliefs, but in thinking that moral facts depend on God.

Ivan is a Revisionist. He recognized a problem in his understanding of morality, and he responded not by abandoning his concepts, but by changing them. This was—I hope you'll agree—the appropriate thing to do. Moral Error Theorists claim that we are all in a position analogous to Ivan's. They argue that there is a problem with our moral concepts. However, they don't advocate a revision like Ivan's. The conceptual defect, they say, is not something morality can do without. It is an irremediable error.

An error theory makes two claims: (1) that a discourse has a certain commitment; and (2) that this commitment is problematic in such a way that the discourse is systematically in "error". Typically, (2) is cashed out in terms of the untruth of all substantive or atomic propositions within the discourse.[2] We saw an example already in Ivan's story. As an atheist, Ivan is a Theological Error Theorist. The atheist claims that theological discourse is committed to the existence of an omnipotent, omniscient, omnibenevolent creator of the universe. Since no such being exists, atheists claim, theological discourse is in error. Other discourses where an error theory might be appropriate are discourse about witches, phlogiston, unicorns, astrology, and so on.

In its most common contemporary form, Moral Error Theory (henceforth "Error Theory") claims that moral discourse is committed to the existence of a very particular kind of reasons (to be spelled out at length in Chapter 1), and that such reasons don't

[2] In the view I'm interested in, untruth results from reference failure. Reference failure might lead to false statements, or merely "untrue" statements. If atomic sentences (e.g., "*X* is wrong") that fail to refer are false, then embedding those propositions may make them true (e.g., "it is not the case that *X* is wrong"). I won't be concerned with the details of error theoretic semantics. (See Daly & Liggins 2009.)

exist.[3] Thus, substantive moral claims like, "Stealing is wrong", and "You should give to charity" are never true. This is because moral concepts like WRONG and SHOULD entail these non-existent reasons. Though not historically very popular, Error Theory seems to be on the rise.[4] Its opponents sometimes resist by arguing that the reasons purportedly entailed by moral concepts *do* exist.[5] The other standard response is to deny that moral concepts have the entailment that Error Theorists claim.[6] My aim is explore and develop an alternative to these standard views. Though this option has been noticed, it has never been explored in any real depth.[7] I am going to grant the Error Theorist both the metaphysical and conceptual claims for the sake of the argument. My central idea is that, even if moral concepts entail these special reasons, and even if the reasons don't exist, we *still* shouldn't be Error Theorists. Moral reality might turn out to be different from how we currently conceive of it. Thus, I propose that—if our understanding of morality is problematic in the way Error Theorists claim—we should revise it so that it's not. We should not be Error Theorists, but Metaethical Revisionists.

Revisionist theories have been proposed in a number of domains, including: race and gender, psychological states, truth, free will, and personal identity.[8] Such views are sometimes claimed not to advance their respective debates because they succeed only in

[3] These are not Normative Error Theorists, who claim that *all* normative discourse is in error. (Streumer 2013.) Error Theorists think that moral reasons are distinctively problematic.
[4] John Mackie, the original Error Theorist, initially had few followers. Things have changed in recent decades. Richard Joyce is the most famous contemporary Error Theorist, and will be my primary interlocutor. (Joyce 2001; 2006; 2011.) Others include: Christopher Cowie (2016), Chris Daly and David Liggins (2009), Richard Garner (1990), Matthew Lutz (2013), Joel Marks (2013), and Jonas Olson (2014).
[5] These often take the form of "partners in guilt" arguments. See Bedke 2010; Parfit 2011; Shafer-Landau 2009.
[6] This response is the one taken by effectively everyone who is not an Error Theorist or Non-Naturalist Realist.
[7] Joyce 2011, 519.
[8] Haslanger 2000; 2006 (Haslanger explicitly presents her view as a response to Racial Error Theory); Fodor 1987; Scharp 2013; Vargas 2013; Parfit 1986.

changing the subject. This objection is correct that revision is not always an option. Suppose someone claims that, if our concept of God entails the existence of an omnipotent, omniscient, omnibenevolent creator of the universe, and if no such being exists, then we should simply revise our understanding of God. We would be right to reject this kind of "Theological Revisionism". It's just too important, too central, to the theistic concept, GOD, that God is the omnipotent, omniscient, omnibenevolent creator of the universe. Talking about "God" in that revised way equivocates, inadmissibly changing the subject.[9] But, this objection doesn't hold in all cases. There are *many* concepts that have undergone revision. We have changed the way that we understand mass, momentum and atoms, as well as marriage, rape and race.

This leads one to wonder: In virtue of what may a Revisionist claim to have modified rather than replaced a concept? Error Theorists cash this idea out in terms of "negotiable" and "non-negotiable" features of concepts. Changing negotiable features preserves the original concept; changing non-negotiable features replaces it. For example, that God spoke to Moses on Mount Sinai is negotiable. We could still be talking about God even if we said that he did no such thing. That God is the creator of the universe is probably non-negotiable. We could no longer be talking about God if we denied it. According to Error Theorists, moral discourse cannot be cured of error because the error reflects a non-negotiable feature of the concepts. This is, in my view, the most interesting—if opaque—part of their argument.

[9] Of course, one might still be talking about a god in the lower case "g" sense. I'm thinking of the specifically theistic (Abrahamic) concept.

Though called by many names, the idea of non-negotiable conceptual features is fairly common in philosophy.[10] Nevertheless, it has somehow avoided serious exploration. To remedy this lacuna, I will offer a general theory of conceptual negotiability. An account like the one I offer provides a rubric for evaluating Revisionist proposals by allowing us to determine which features of a concept are negotiable. This tells us whether a proposed change would preserve the original concept, or simply replace it and change the subject. Thus, my project has utility well beyond this particular metaethical debate. It can provide guidance to Revisionist proposals in any domain.[11]

In short, the thesis will try to persuade readers of two things. One is a general view about conceptual negotiability and the conditions under which conceptual revision is possible. The other is a view about the possibility of revising moral concepts in particular.

[10] The idea, sometimes by this very name, appears in Diaz-Leon 2012; Ferrero 2009; Glasgow 2009; Haslanger 2000; Horgan & Timmons 2002; Joyce 2001, 2007, 2011; Loeb 2008; Lutz 2013; Parfit 2011; Shafer-Landau 2005; Vargas 2013.

[11] This aspect of the project falls into what Alexis Burgess and David Plunkett have termed "Conceptual Ethics" (2013a; 2013b). They use "ethics" quite loosely. Conceptual Ethics deals with questions about which concepts we should (in some loose sense) be using.

Chapter 1: The Argument for Error Theory

1 Introduction

The purpose of this chapter is to get clear on the argument for Error Theory by specifying three claims: *Conceptual, Ontological,* and *Non-Negotiable.* §2 explains *Conceptual.* §3 explains *Ontological,* and clarifies how the combination of *Conceptual* and *Ontological* implies that moral discourse is in error. §4 distinguishes between *Conceptual* and a claim about it, namely, *Non-Negotiable*. The former two claims have been the focus of existing debate. I will grant both for the sake of the argument. For lack of space, I will skip over the debate concerning *Ontological.* I spend more time with *Conceptual* since understanding this claim is crucial to understanding *Non-Negotiable.* This final claim has largely been ignored in existing debates. This is surprising, since it is the most interesting—and least clear—premise in the argument. My goal is to shed some light on it.

A final prefatory note: In this chapter (and the whole thesis really) I'm going to try to address Error Theorists on their own terms. They presuppose a variety of metaethical and metanormative views, each of which is subject to controversy in its own right. I'll point out some of these presuppositions along the way, but I won't bother to criticize or defend them. There are lots of issues to pick at; I'm interested specifically in the question of conceptual negotiability.

2 The Conceptual Claim

Error Theorists believe moral discourse is committed to a particular kind of reasons. These reasons have two important characteristics: categoricity and irreducibility.

2.1 Categoricity

The *locus classicus* for Error Theory is John Mackie's *Ethics*. His central thesis in the first chapter of the book is that, "There are no objective values."[1] He later clarifies that what he's after is the "objectively prescriptive", "intrinsically action-guiding" character of morality. Elsewhere Mackie writes, "To say that [objective values] are intrinsically action-guiding is to say that the reasons that they give for doing or for not doing something are independent of that agent's desires or purposes".[2] In other words, Error Theorists claim that morality is committed to "categorical" reasons. This is the view that moral facts are reason-giving independently of the desires, values, interests, commitments, etc. of the agent. This is sometimes also called the "authority" of morality. It doesn't matter whether one's interests are best served by stealing; stealing is morally wrong, and therefore one has a reason (maybe even an overriding, or exclusionary reason) not to do it.[3]

These are to be contrasted with "hypothetical" reasons, which are action guiding conditionally upon the interests of the agent. For instance, I might have a reason to go to the kitchen: namely, that there are cookies there. But, this will only be a reason for me to go to the kitchen on the condition that I *want* a cookie. If you advise me to go the kitchen because there are cookies there, and I tell you that I have no interest in a cookie, then your natural response will be to retract your advice. In so doing, you are recognizing that

[1] Mackie 1977, 15.
[2] Mackie 1981, 115. Mackie has been accused of conflating or confusing reasons with motivational force (Parfit 2011, 451). However, this interpretation is questionable, and I shall assume that Mackie did distinguish between reasons and motivation. Phillips 2013 defends this interpretation.
[3] There are those that deny that moral concepts have any relationship with reasons at all. Error Theorists (and many more besides) assume that moral facts are, or at least purport to be, reason-giving.

the fact that there are cookies in the kitchen is only a reason for me to go there if I care. Reasons like these do not trouble Error Theorists.

Mackie also accepts that institutions can issue reasons to their participants. Chess is his favored example. Insofar as I am playing chess, I should move bishops diagonally. That is to say, I have a reason to move my bishops diagonally. But this reason too is contingent, this time upon my having bought into the "institution" of chess-playing. I have a reason to comply with the rules *insofar as I am playing chess*. But, institutions can apply to individuals without them "buying in." Philippa Foot gives a good example. According to the standards of etiquette, I shouldn't (I have a reason not to) belch at the dinner table. This will be true of me regardless of my interests. Suppose someone tells me that I shouldn't belch, and I respond by saying that I don't give a damn about being polite. This will not lead my interlocutor to retract her statement. It is *true* that I shouldn't belch *according to etiquette.* That is to say, the institution and its standards apply to me regardless of my desires, interests, commitments, etc.

The difference between the standards of etiquette and morality, according to Mackie, is that the "institutional reasons" of etiquette aren't authoritative unless one has bought in. As Foot writes, the standards of etiquette lack "automatic reason-giving force" and one might be right to deny that one had any reason to comply with its demands.[4] One can sensibly or rationally ignore the standards of etiquette. Its institutional reasons lack normative authority; they aren't fully real or rationally binding. Not so, it is said, for morality. Richard Joyce, a 21st century champion of Error Theory, illustrates this point dramatically. His example is Celadus, the unwilling gladiator. Having been kidnapped

[4] Foot 1972, 309.

and forced into the arena, Celadus' only hope for survival is to violate the rules of gladiator combat by throwing sand in his opponent's eyes. Surely, we think, he has a reason to throw the sand. Of course, Celadus is subject to the rules of combat—they *apply* to him. But, the requirements of gladiator combat are not real reasons for him. He should ignore them. In contrast, Joyce writes, "Morality is not just a list of *Dos* and *Don'ts* backed up with punishment. We think that a person is bound by those rules whether he accepts them or not—that the rules are, in some sense, *his* rules whether he accepts them or not."[5]

Obviously, as Stephen Finlay points out, "This claim is hardly unique to the error theorist; ever since Kant it has been an *idée fixe* in metaethics that an essential feature of morality is the 'categorical' as opposed to 'hypothetical' character of its requirements."[6] Indeed, a raft of non-Error Theorists can be cited saying much the same thing. According to David Wiggins, we cannot escape from moral reasons "by simply flying the skull and cross-bones and renouncing altogether the aim of belonging to the moral community."[7] Steven Darwall writes that, "Viewed from the inside, moral demands appear to present us with reasons for acting that are unconditional on our desires and, it seems, that override any considerations based in these."[8] Or, as Shafer-Landau puts it, "Genuine moral rules are those that impose demands on us regardless of our desires or interests. Their authority does not stem from any instrumental link to what we care about. We are

[5] Joyce 2001, 34.
[6] Finlay 2008, 349.
[7] Wiggins 1995, 298.
[8] Darwall 1997, 305.

morally bound to aid the vulnerable, or fulfill our promises, even if doing so is nothing to us."[9] So far, then, Error Theorists find plenty of agreement.

2.2 Irreducibility

John Mackie and Richard Joyce—certainly the most well known Error Theorists—initially cashed out *Conceptual* purely in terms of categoricity. In more recent work, however, Joyce writes: "Part of [morality's] authority can be captured by noting that morality is categorical rather than hypothetical, but that doesn't seem enough."[10] Yet, he has a hard time explaining what more there is, describing it as "practical oomph", "a deeply mysterious kind of force", and "bindingness".[11] He writes:

> [The experience of moral normativity is] a primitive sort of feeling/thought which resists analysis, decomposition, explication, or naturalistic demystification… [I am] prepared to accept that it may turn out that this oomph can never be adequately analyzed, that it is a kind of magical and indescribable quality.[12]

Here Joyce sounds very much like a Non-Naturalist. Non-Naturalists claim that that moral reasons are irreducible, and the concept is unanalyzable. They cannot be explicated in naturalistic non-normative terms. Reasons are considerations that count in favor of actions. But, they "count in favor" in that they *provide a reason* for that action. The most well-known Non-Naturalist, Derek Parfit, sounds a bit like an Error Theorist when he writes: "If some normative concept, claim, or fact cannot be defined or restated in non-normative terms, we can call it irreducibly normative… Normativity is either an

[9] Shafer-Landau 2005, 107.
[10] Joyce MS, 7.
[11] Joyce 2006, 60-63; 2011, 525.
[12] Joyce 2008, 559-561.

illusion, or it involves irreducibly normative truths."[13] (Of course, Parfit thinks normativity, including moral normativity, is not an illusion because he thinks he can make sense of irreducibly normative reasons.)

Jonas Olson, a more recent addition to the Error Theorist camp, articulates *Conceptual* in precisely these terms. According to Olson, moral facts, he writes, "are or entail facts that count in favor of or require certain courses of behavior, where the favoring relation is irreducibly normative."[14]

Here we run into some tricky exegetical issues. Olson writes that focusing on irreducibility yields the "best articulation" of *Conceptual*, which implies that he sees it as modifying rather than adding to Mackie's and Joyce's claims about categoricity. However, there is an important difference between the two. Categorical reasons—considerations that are reasons for all agents independently of their contingent desires, interests, commitments, etc.—need not be irreducibly normative. Moreover, irreducible reasons might be hypothetical in form. A number of metaethicists claim to be able to account for reducible categorical reasons.[15] When Olson suggests that his version of *Conceptual* is a better "articulation" of the claim about categorical reasons, he must be thinking that these accounts are doomed to fail. In other words, he claims that moral facts entail categorical reasons, and in order for that to be so, those reasons must be irreducible.

The irreducibility of moral reasons is the source of the putative ontological liability. What Error Theorists "emphatically will not accept is any naturalist attempting

[13] Parfit 2011, 266-267; my emphasis. Parfit thinks that *all* reasons are irreducible; Error Theorists don't. Moral reasons are supposed to be special. I'll assume that Error Theorists are right.
[14] Olson 2014, 118.
[15] For instance, Schroeder 2007 and Markovits 2014.

to sidestep the challenge [of explaining moral normativity] by claiming that there is nothing especially unusual about the practicality of morality that requires any special explanation."[16] On this point, Error Theorists again agree with Non-Naturalists. For example, according to William FitzPatrick, the moral wrongness of an act is a property that gives all (moral) agents reason not to perform it. And, he writes, a "commitment to facts such as [this] constitutes a theoretical commitment to non-natural facts."[17] Non-natural facts, he clarifies, "are facts that cannot be cashed out in empirical terms, as by appeal to facts of psychology or biology, or to complex facts constructed entirely from such facts".[18] According to Error Theorists and (some) Non-Naturalists, we cannot explain the normativity morality purports to have without appealing to facts of this kind.

We are now finally in a position to give a precise formulation of the Error Theorist's conceptual claim:

> ***Conceptual****:* Moral facts are or entail facts about reasons, where those reasons are categorical and irreducibly normative.

Let's call categorical, irreducibly normative reasons "CINRs" for short. *Conceptual* comes with many theoretical commitments, let's take stock by summarizing them. Error Theorists claim that moral concepts like WRONG or GOOD entail facts about reasons. This is analogous to the way that the concept MARE entails both FEMALE and HORSE.[19] The reasons entailed by moral concepts are categorical—i.e., they apply regardless of one's interests, values, commitments, etc. They are also not reducible—e.g., to desires, values,

[16] Joyce 2008, 261.
[17] FitzPatrick 2008, 185. Not all Non-Naturalists agree. (Parfit 2011, 479.)
[18] *Ibid,* 184 note 63.
[19] Conceptual entailments are typically transparent to competent users of the concepts. Surely anyone competent with MARE will realize that it entails FEMALE and HORSE. But Error Theorists don't expect competent users of moral concepts to accept *Conceptual* (see §2.3). This is one reason to be suspicious of *Conceptual*, but I won't press it.

or the constitutive standards of reasoning. Finally, in order to make sense of such reasons, one must postulate the existence of non-natural normative facts.

2.3 Evidence for Conceptual

Error Theorists don't claim that ordinary people would recognize and endorse *Conceptual*. (Imagine asking the Tesco clerk whether he thinks moral facts are or entail facts about CINRs.) Instead, they assert that *Conceptual* is what "best explains and makes sense of the dominant majority of moral practices."[20] *The ways in which we use* moral concepts indicate that we, in some tacit way, take moral facts to entail facts about CINRs. This is a "conceptual commitment". Moral discourse commits us to CINRs in the way that inductive inferences commit one to principles of induction. If I come to the belief that my chair will support my weight on the basis of the fact that it has always supported my weight, then I am, in some sense, committed to the claim that the principles of induction are rational or justified inference rules. The same is thought to apply in the moral case. The fact that ordinary people might not recognize or accept *Conceptual* doesn't mean that they aren't committed to it.

Error Theorists defend *Conceptual* by appeal to intuitions about ordinary moral practice: how "we" speak, and think about morality. Joyce writes:

> Consider, again, the comparison between [Celadus] the unwilling gladiator, and the hurtful Gyges. One rejects and violates the rules of gladiatorial combat, one rejects and violates the rules of morality. Yet we do not think them on a par. We invest the moral judgment with an extra authority.[21]

[20] Joyce 2001, 98.
[21] Joyce 2001, 45.

We can say of Celadus that he *ought-according-to-the-rules-of-gladiatorial-combat* not throw sand, though he really has no reason not to. Contrast this with what we would say of Gyges. We would not say that he *ought-according-to-the-rules-of-morality* not kill. But if he stands to gain something important from killing then that's what (*all-things-considered*) he should do.

> That's *not* how we think of morality. Someone who reasoned in such a way might be accused of fundamentally misunderstanding what we mean by 'morally ought'. … [I]n the moral case, we are not content to admit that our claim that there is a reason to refrain from killing is merely a permissible way of speaking from a perspective that endorses the dictates of morality. We are still left with a desire to say something more—to imbue the moral imperative with a greater authoritative force. [22]

Again and again, it's about what "we" do, what "we" would say or think. The assertion is that "our" moral thought and language only makes sense if moral facts are taken to entail CINRs.

The relevant "we" is not just philosophers. *Conceptual* is a claim about the concepts that being employing by ordinary people on the street. "[A]nyone is mistaken" Joyce claims, "who dismisses categorical [moral reasons] as a bizarre extravagance endorsed only by Kant and his followers."[23] According to Mackie, *Conceptual* is a part of the "main tradition of European moral philosophy", including: Plato, Kant, and Sidgwick.[24] But, crucially, it is not just a philosophical curiosity: "It has also a firm basis in ordinary thought, and even in the meanings of moral terms…"[25] Ordinary people, in their ordinary moral thinking and speaking, are committed to the truth of *Conceptual*. In

[22] *Ibid*, 36-41.
[23] Joyce 2001, 43.
[24] Mackie 1977, 30.
[25] *Ibid,* 31-33.

this way *Conceptual* is actually an empirical claim. It's a claim about what is happening in the minds of human beings going about their ordinary lives.

This means that verifying *Conceptual* requires empirical investigation into the way that ordinary people think and talk about morality. Good old-fashioned conceptual analysis and armchair speculation are useful places to start. But, if we really want to know whether the ordinary concepts entail CINRs, then we need to get our hands dirty and do some actual empirical investigation. I'm certainly not advocating that anyone explicitly quiz ordinary people on their metaethical views. This would not yield any interesting results. I don't think that ordinary people have explicit metaethical beliefs. The idea is simply that implicit metaethical commitments can be indirectly inferred from ordinary practices and "folk" intuitions. Now, I don't want to get bogged down in the raging methodological debate over experimental philosophy ("X-Phi").[26] But, let me make one point in defense of empirical investigation into folk metaethics.

Antti Kauppinen argues that the methods of X-Phi are not helpful for conceptual analysis because they fail to capture the right kind of intuitions. We are interested in the purely semantic (i.e., not pragmatically influenced) and un-confused intuitions of competent users of concepts. "Running a poll", Kauppinen writes, "provides no shortcut in this business of reaching a better conceptual self-understanding."[27] This is an important cautionary note for the study of folk metaethics. We must make efforts to avoid distorting pragmatic influences, and to ensure the conceptual competence of participants. And, indeed, direct polling is likely *not* the best approach for teasing out the implicit commitments of moral practice. This is because questionnaires tend to reveal only what

[26] Pölzler 2014 provides a good overview of potential concerns about the relevance of empirical research to metaethics.
[27] Kauppinen 2007, 113; see also Weinberg and Alexander 2014.

we might call faulty "first-pass" articulations of concepts. For instance, ask Anna what she understands birds to be, and she may say "feathered, flying animals". Remind her of penguins and ostriches, and she will be inclined to modify her account. We should not understand this amendment of her "first-pass" as a change in her concept. Instead we should recognize that the precise contours of her concept are not immediately accessible to her.[28] This represents a potential pitfall for any studies that rely too heavily on polls and questionnaires. They tend to elicit only these first-pass kinds of answers, which may be unhelpful or even positively misleading.[29]

But, these criticisms do not invalidate the enterprise of folk metaethics as a whole. In order to avoid the faulty first-pass problem, studies could be designed to elicit more considered judgments.[30] One possibility would be to set up an online introductory metaethics class, which focused on moral normativity. Students from around the world could be presented with the various views by their proponents, compare the merits of each, discuss with other students, and weigh their intuitions. Along the way, these intuitions would be reported in the form of class assignments. A simpler version of this experiment might forgo the class structure and aim instead to create a kind of Socratic dialogue between participants. Prompting from precisely phrased questions in the context of a dialogue might be able to elicit considered judgments without falling into explicit questioning. Alternatively, participants might be presented with a vignette, asked what the characters have reason to do, and then questioned by an experimenter. The kinds of examples given to us by Error Theorists (e.g., Gyges and Celadus) would make for good

[28] This is an important point, which we'll return to in Chapter 2.
[29] This objection also seems to apply to the appeals to armchair intuitions about what "we" would say. How we expect others to respond likely reflects primarily a first-pass understanding of the concepts.
[30] Aguiar et al. 2014 discusses the possibilities for utilizing a kind of competence/performance distinction by testing the behavioral implications of conceptual competence.

test cases. If Joyce is right, Gyges' flouting of moral norms should elicit a very different kind of response from Celadus' flouting of gladiatorial norms.[31] If the participants persist in claiming that characters like Gyges have reasons not to violate moral norms even after it is made clear that doing so is in their interest, this would be evidence that the participants are committed to categorical moral reasons. Needless to say, framing the cases and questions properly would require serious work.

Around a dozen folk metaethics studies have been conducted in the past decade or so. Several early studies were claimed to have relevance to Error Theory, and even cited Mackie in support of this contention.[32] However, the experimenters misunderstood the relevant philosophical claims.[33] Instead of addressing folk assumptions about the categoricity and irreducibility of moral reasons, these studies investigated their assumptions about the universality of morality's application (i.e., Relativism vs. Absolutism). Nearly all of the subsequent studies on folk metaethics have continued to explore this same issue. These studies are interesting, but do not explore what we are really interested in.[34] Perhaps sometime soon we will have better evidence with which to assess *Conceptual.*[35] In the meantime I'll grant the premise for the sake of the argument.

[31] Extant data suggest that it may make a big difference which kind of moral infraction is being considered. It would therefore be important to contrast the flouting of gladiatorial norms with the flouting of a variety of different kinds of moral norms (e.g. norms of harm, fairness, and honesty). It would also be important to vary the degree of the severity of moral infractions.
[32] Goodwin & Darley 2008 and 2012.
[33] Fraser 2013 demonstrates this quite clearly.
[34] They are also subject to a number of particular methodological difficulties, which I lack the space to discuss. See Pölzler 2015.
[35] There has been one study on the categoricity of moral reasons. However, it is unpublished and has yet to undergo peer review. See Wright MS.

3 The Ontological Claim and "The Error"

Now that we are clearer on *Conceptual*, we can dispense with *Ontological* fairly easily.

Ontological: The reasons described in *Conceptual* do not exist.

In particular, *Ontological* claims that there are no non-natural facts of the kind needed for CINRs.

The Error Theorists I'm concerned with are not Normative Error Theorists, who claim that *all* normative discourse is in error. Error Theorists think that there are *some* reasons—it's just that there are only hypothetical and/or reducible reasons. There are a variety of naturalistic metanormative theories available for Error Theorists. Joyce, for instance, adopts an Internalist view (*á la* Bernard Williams) on which an agent's reasons are reducible to facts about her interests and values, and about what would promote those ends.[36]

It should be obvious how the combination of *Conceptual* and *Ontological* leads to the charge of error. If moral facts entail CINRs, and if no such reasons exist, then there is something defective about moral concepts. This defect, Error Theorists claim, means that moral concepts have null extension, and that no propositions employing them are ever non-trivially true.[37] This would be a very serious problem for moral discourse.

Some metaethicists resist this conclusion by denying *Ontological*. As we saw, Non-Naturalist Realists accept *Conceptual*. However, they think that sense can be made of

[36] Joyce 2001, Chapters 4-5; Williams 1981.

[37] See Introduction note 2. Careful readers will have noticed that this conclusion is a little quick. Conceptual defects do not necessarily entail reference failure or first-order untruth. As will become clear, this point is crucial to my own view.

CINRs.[38] Thus, there is no cause for concern. Of course, this is a very contentious claim. Many metaethicists, and not just Error Theorists, find *Ontological* very plausible. We should have something more to say about reasons, they think. Irreducible normativity, and the non-natural facts it brings in tow, feel too much like magic. Unfortunately, I lack the space to explore the topic in depth. In any case, this is well-travelled terrain, and I'm interested in another part of the argument. This premise I will grant *gratis.*

Other metaethicists resist Error Theory by denying *Conceptual*. One way of doing this is to accept that moral facts are categorically reason-giving, but deny that these reasons are irreducibly normative. We saw already that some metaethicists claim to be able to account for reducible categorical reasons. Some go even further, denying the categoricity of moral reasons.[39] Foot once argued that moral standards are really institutional standards on a par with etiquette. Bernard Williams claimed to "doubt very much" whether categorical reasons are what "the ordinary moral consciousness wants from the *ought* of moral obligation".[40] Peter Railton advocates a version of "stark raving realism", which nevertheless "does not yield moral imperatives that are categorical in the sense of providing a reason for action to all rational agents regardless of their contingent desires."[41] He does admit, however, that the denial of *Conceptual* is "troubling". He writes:

> Thus, while *it certainly is a limitation* of the argument made here that it does not yield a conception of moral imperatives as categorical, that may be a limitation we can live with and still accord morality the scope and dignity

[38] Enoch 2011; Parfit 2011; Shafer-Landau 2009.
[39] Foot 1972; Williams 1981; Brink 1984; Railton 1986; Street 2008.
[40] Williams 1981, 122-123.
[41] Railton 1986, 165, 201.

> it traditionally has enjoyed. Moreover, it may be a limitation we must live with.[42]

We will return to ideas along these lines in Chapter 4, where I consider the options available to Metaethical Revisionists.

So where does this leave us? We've now seen the first two premises in the argument for Error Theory: *Conceptual* and *Ontological*. These have been at the center of attention in existing debates over Error Theory. We briefly discussed the argument for *Conceptual* and the kind of evidence I think we would need to evaluate it. Unfortunately, there is very little data to go on. Nevertheless, *Conceptual* was granted. For lack of space, I passed over the disputes concerning *Ontological*. Error Theorists may have this premise too. This is because I'm really interested in the third premise, which is a claim about *Conceptual*.

4 The Neglected Claim: *Non-Negotiable*

The argument for Error Theory includes another premise, one that has gone largely unnoticed:

> ***Non-Negotiable***: The reasons described in *Conceptual* are a non-negotiable feature of moral concepts.

It is strange that this premise has received so little attention. In my view, it is the most interesting—if extremely opaque—part of the argument.

We can think of *Conceptual* and *Non-Negotiable* as elements of weaker and stronger versions of Error Theory respectively. Someone who maintains *Ontological* and *Conceptual* but not *Non-Negotiable* holds a "Weak" Error Theory. This is the view that moral

[42] *Ibid*, 203.

discourse is only contingently in error. On my reading, Mackie himself was a Weak Error Theorist. Mackie clearly thought that ordinary moralists are in error; but didn't obviously think that this error is fundamental to morality as such. After establishing an Error Theory in the first chapter of *Ethics*, he proceeds to spend the rest of the book laying out first-order moral principles. This would be puzzling if he believed the error to be conceptually necessary for morality. One way of reading the subsequent chapters is as Mackie's idea of how error-free ethics might look. This may even be the idea behind his subtitle "*Inventing* Right and Wrong"—i.e., creating anew. This reading is further supported by a passage from his earlier book, *Problems from Locke*: "[*C*]*onceptual reform*, rather than mere analysis of our present concepts, is, I believe, needed for ethics. I hope to discuss this topic in another book".[43]

On the other hand, someone who maintains *Ontological, Conceptual* and *Non-Negotiable* holds a "Strong" Error Theory. On this view, the error is not a contingent feature of moral discourse. Moral discourse is necessarily and irremediably in error. Contemporary Error Theorists are all of this strong variety. (Unless otherwise indicated, "Error Theorist" means "Strong Error Theorist".)

Quite a few philosophers have relied on the idea of non-negotiable conceptual features.[44] None, however, have explored the idea in any real depth, and it's not clear exactly how we should understand conceptual negotiability. I attempt to remedy this omission in Chapter 2. For now, however, all we need do is get clear on what Error Theorists have said on the matter, and on the argument they offer for *Non-Negotiable*. We

[43] Mackie 1976, 196 note 27; my emphasis. Though he may have been a Revisionist in some sense, Mackie is not a Metaethical Revisionist in *my* sense. This is because he holds that all substantive moral propositions are false, even if moral propositions could be true at some point in the future. See Chapter 3 §3 for more on the distinction.

[44] Introduction, note 10.

may be fairly brief, as my ultimate aim is to challenge this understanding of conceptual negotiability.

The thought is, roughly, if concept *C* is typically conveyed with term *T*, and someone uses *T* to convey a concept lacking a non-negotiable feature of *C*, then one is not actually employing *C*. One has equivocated or changed the subject. One way of thinking about it is in terms of cores and peripheries. Concepts have more and less central elements. The more central some feature of *C,* the less negotiable that feature is to *C.*[45] Peripheral elements of a concept can be eliminated without threatening to eliminate the concept itself, but core elements cannot. For instance, that witches have warts and ride on brooms is a peripheral, negotiable feature of WITCH. If you denied this I wouldn't think that you were in any way confused about the concept. On the other hand, that witches have supernatural powers is central and non-negotiable. Someone without supernatural powers just couldn't be a witch. If you denied this, you would be employing a different concept.

Shafer-Landau takes a similar view, though he puts it in terms of a concept's role in a theory:

> Conceptually non-negotiable propositions are those that are embedded within theories, and have a special role to play therein. They are conceptually necessary conditions that are such that their falsity undermines the entire theoretical edifice that has been raised on their foundations. To be non-negotiable, then, is to play the role of a conceptual linchpin.[46]

Again, the idea is that some features of concepts are more essential than others. Some changes threaten to eliminate the concepts, while others do not. According to Error

[45] This makes it sound like negotiability comes in degrees, which is not something Error Theorists discuss. It does seem like a natural view, however.
[46] Shafer-Landau 2005, 108.

Theorists, CINRs are so central to moral concepts, that if we eliminate them from our thought and language, we eliminate moral concepts themselves. This claim is obviously much stronger than *Conceptual*, according to which the commitment to CINRs could be merely a contingent feature of our moral concepts.

Non-Negotiable implies that removing the error from our moral discourse would leave us with mere "schmoral" discourse.[47] Of course, schmoral discourse may be interesting and perhaps even appealing in its own right. But, it's certainly not the *real thing*. It's not immediately clear why this would be a problem. What difference would it make whether our discourse is moral or schmoral? It makes a difference, the Error Theorist claims, because if we made moral discourse like other unproblematic discourses we would "strip the discourse of its very purpose".[48] "[T]he whole point of having a moral discourse is to prescribe and condemn various actions with categorical force."[49] Joyce writes:

> If moral concept *M* contains erroneous claim *E*, and transparently eliminating *E* would leave concept *M**, then can we *use M** in all or most of the ways that we heretofore used *M*? If not, then we have strong ground for thinking that *M essentially* presupposes *E*. My claim, then, is that a normative discourse for which categorical imperatives were transparently institutional (and desire-transcendent reasons transparently institutional) could not be used in many of the ways we use moral discourse, and therefore would not *be* a moral discourse.[50]

In other words, to correct the error in moral discourse, we would need some set of concepts that did not entail CINRs, but could still be put to all (or perhaps *nearly* all) the

[47] Joyce 2007; see also Blackburn 1993.
[48] Joyce 2001, x.
[49] *Ibid,* 157.
[50] Joyce 2011, 530. Given this use-centered understanding of conceptual negotiability, it's surprising that Joyce doesn't ever bother to rebut Quasi-Realists. The ambition of the Quasi-Realist program is to show that we can support a "realist-seeming" moral discourse (i.e., preserve the ways we use the concepts) without any realist ontological commitments. See Blackburn 1993.

same uses as those that do. This, it is claimed, cannot be done. Concepts which do not entail CINRs are simply too "wimpified" to be mistaken for the real thing.[51] Once again, the Error Theorist is in agreement with Non-Naturalist Realists.[52] FitzPatrick makes effectively the same argument, calling any alternative to his Non-Naturalism a "watered down substitute" for the real thing.[53] Realism, he claims, is impoverished by the "retreat to metaphysically safe ground"; it gives up on what is attractive about the view in the first place. It is these supposedly attractive features that the Error Theorist thinks cannot be made sense of. One way of thinking about Error Theory, then, is that it's Non-Naturalist Realism for skeptics.[54]

In Chapter 2 I shall argue that use is not the right criterion for conceptual negotiability. For the time being, however, it's important to see that, on the Error Theorist's account, supporting *Non-Negotiable* requires an accurate and detailed description of what we do with moral discourse.

5 What Do We Do with Moral Discourse?

The obvious question before us is: What is moral discourse actually used to do? While this is a patently empirical question, once again, armchair reflection has stood in for actual empirical investigation. For instance, Joyce suggests we use moral discourse to "justify deserved punishment", to "undergird the emotion of guilt", and to "act as a bulwark against a range of motivational infirmities".[55] Elsewhere he suggests that moral

[51] Joyce MS.
[52] Not all Non-Naturalist Realists accept *Non-Negotiable*. (Shafer-Landau 2005.)
[53] FitzPatrick 2008, 163, 192.
[54] David Enoch recognizes Error Theorists as the Non-Naturalist's "kindred spirits" and "most respectable opponent[s]" (2011, 81, 121).
[55] Joyce 2007, 66.

discourse is an evolutionary adaptation to counteract weakness of the will.[56] However, despite his own armchair speculations, Joyce admits:

> [The answer to this question] will almost certainly be extremely complex, and is, moreover, largely an empirical business. It is extraordinary how rarely this matter has been squarely faced, and deplorable that on those occasions that are exceptions, vague intuitions from the armchair have, more often than not, been thought to suffice.[57]

Note that an empirical assessment of this claim doesn't require any commitment to the methodologies of X-Phi.[58] This is because it wouldn't necessarily involve gathering data on the intuitions of ordinary moralists. Instead, it would require investigating ordinary moral discourse out "in the field", as it were, of real human life.[59] What we are looking for is features of our moral discourse that would have to be abandoned if we adopted an understanding of morality on which moral facts did not entail CINRs. We might begin by exploring the extent to which, and the ways in which, moral discourse is actually put to the uses that Joyce mentions. (Of course, it is extremely unclear what it would mean to "justify deserved punishment" on Joyce's view.) For instance, in ordinary social contexts when people attempt to evoke guilt in offenders by engaging them in moral discourse, do they do it in a way that would be compromised by a denial of *Conceptual*? Or would, perhaps, sympathy with the victims or the censure of the community be sufficient to produce such feelings?

---

[56] Joyce 2006.
[57] Joyce 2007, 66. It's not obvious whether "deplorable" is intended to be self or other deprecating.
[58] It could be thought to correspond to one *very* loose understanding of X-Phi. "[W]hen philosophical arguments invoke or make assumptions about empirical matters, those assumptions should be assessed according to the best natural and social scientific evidence available… Thus, the term, 'experimental moral philosophy' might be replaced by a phrase like '*empirically well-informed moral philosophy.'"* (Alfano & Loeb 2014) I take it that this *very thin* view is not controversial.
[59] This is not to say that controlled psychological experiments wouldn't also be useful. For instance, it would be interesting to see how the folk respond to arguments against the existence of CINRs.

Traditional anthropological, psychological, and sociological research would be of importance here. Jonathan Haidt's book, *The Righteous Mind,* and Part III of Steven Hitlin and Stephen Vaisey's *Handbook to the Sociology of Morality* are good examples of what I have in mind.[60] However, the investigation need not be confined to traditional methods. The Internet and Big Data analytics would be enormously powerful tools. Data aggregators like Google have enormous reserves of data on social interactions that can be put to use: news reports, TV shows, movies, books, social media, and more. From these masses of data we could pick out distinctively moral claims and interactions, compile and analyze them, and produce a catalogue of the various uses to which we put moral discourse. The catalogue could then be scoured for uses that would no longer be available if *Conceptual* were explicitly denied. Of course, this would be a very serious undertaking. But, who said empirical work was easy? The investigation would surely benefit from greater input by philosophers. As any philosopher who has read much psychological or sociological literature knows, it is sometimes painfully clear how little of *our* literature the students of these disciplines read. A philosopher's understanding of the pertinent philosophical questions, and training with nitty-gritty distinctions would be an invaluable contribution.

To be clear, this section was intended specifically to address the evidence (or lack thereof) needed to defend the Error Theorist's argument for *Non-Negotiable.* As it stands, there is next to none. Joyce concedes as much. "[U]ntil the jury delivers its verdict on this empirical matter, the fundamental metaethical disagreement between the moral error

[60] Haidt 2012; Hitlin & Vaisey 2010.

theorist and the moral success theorist… *remains at a stalemate.*"[61] While I certainly agree that the ways in which a concept is used are important for determining which features are negotiable, on my view, this is not the whole story. In the next chapter I will explore conceptual negotiability in much greater depth and offer an alternative account. The true assessment of *Non-Negotiable* must wait until Chapter 3.

6 Conclusion

This chapter expounded the argument for Error Theory. We discussed *Conceptual,* the claim that moral facts are or entail facts about CINRs. Accounting for these reasons requires, according to Error Theorists, positing the existence of non-natural facts. The second premise, *Ontological*, states that there are no CINRs (because there are no non-natural facts). These two premises have been the center of attention in debates over Error Theory. They are certainly questionable, but I've granted them both. I'm interested in the third premise, *Non-Negotiable*, which states that *Conceptual* is a non-negotiable feature of moral concepts. While the first two premises are quite familiar, the third has hardly been touched. It is still very unclear what it means to be conceptually (non-)negotiable, much less whether *Conceptual* is a non-negotiable feature of moral concepts. And so, with the groundwork laid, I turn to these questions in the following chapters.

[61] Joyce 2007, 66.