

Morality: The Evolution of a Myth
Richard Joyce

Introduction to *Essays in Moral Skepticism* (Oxford University Press, 2016)

People will often express their strong moral convictions as claims of *knowledge*. Surely we know that making a sport of torturing innocents is evil, know that sympathy is morally better than spite, know that defrauding thousands of people of their retirement savings for no motive other than greed is morally deplorable? One would likely be confused by someone's claiming, seemingly sincerely, not to know such things. As when faced with somebody who honestly claims not to know whether kangaroos are animals, or claims not to know whether three is a number, one would probably be baffled as to what breakdown might lie behind such a fundamental epistemic flaw, and in all likelihood would feel unsure where even to begin correcting such a person. More than this, in the moral case we may not even feel that epistemic correction is quite the appropriate course. G. E. M. Anscombe (1958: 17) writes of someone who sincerely claims not to know that executing innocents is morally wrong: "I do not want to argue with him; he shows a corrupt mind."

Despite the importance of the idea of moral knowledge in human affairs, there is a very long philosophical tradition of doubting that any such thing exists. Sometimes this moral skepticism falls out of a more general epistemological skepticism; sometimes it is specifically *moral*. If we accept the mundane (though far from incontestable) view that to know that *p* involves (i) believing that *p* (ii) truly, and (iii) with justification, then moral skepticism is the disjunction of three theses:

- a) Noncognitivism: the denial that moral judgments express beliefs.
- b) Error theory: the acceptance that moral judgments express beliefs, but the denial that moral judgments are ever true.
- c) Justification skepticism: the acceptance that moral judgments express beliefs, but the denial that moral judgments are ever justified.

While both (b) and (c) are explicitly contraries of (a), they are not contraries of each other. One might endorse an error theory while maintaining that people are justified in their moral beliefs, or alternatively endorse an error theory while adding that all people's moral beliefs lack justification. Similarly, the claim that moral beliefs lack justification may combine with the view that they are all false, but is also consistent with the possibility that moral beliefs are not only true but objectively true. I add this last point about objectivity in order to draw attention to the fact that moral skepticism need not be construed as a form of moral anti-realism. If we take moral realism to be the view that moral discourse expresses beliefs that are sometimes true and, when true, are true in virtue of the obtaining of objective facts (under some specification of *objectivity*), then justification skepticism is compatible with a realist stance. Conversely, if one maintains that moral discourse expresses beliefs about some realm of *non-objective* facts—beliefs that are often both true and justified—then one will be a moral

anti-realist but not a moral skeptic.

It's a good thing for me that (b) and (c) are not contraries, since some time ago I wrote a book devoted largely to arguing for (b)—*The Myth of Morality* (2001)—which I followed up a few years later with a book devoted largely to arguing for (c)—*The Evolution of Morality* (2006). While I purposely gave these books titles that appear in tension (a decision that can be put down to nothing more than perverse philosophical misbehavior¹), in terms of content they were intended to be consistent with each other. *The Myth of Morality* argues that all moral claims are (though expressions of belief) untrue, and it does so using ordinary metaethical methods. *The Evolution of Morality* argues that all moral claims are (though expressions of belief) unjustified, and it does so using an argument that includes an empirical appeal to the evolutionary origins of human moral thinking (also known as an “evolutionary debunking argument”). The two books' central arguments do not interact much logically, and their respective skeptical conclusions are independent of each other—though, thankfully, compatible.

The claim that either of these conclusions implies the falsity of noncognitivism requires some qualification. As I have just presented it, noncognitivism is an entirely negative thesis. As a matter of fact, however, it is always (so far as I know) also put forward as a positive thesis: “Moral judgments do not express beliefs, but rather they linguistically function to do so-and-so” (where the openness of the “so-and-so” is what accounts for different forms of noncognitivism). It is natural to read this as asserting that “doing so-and-so” (where this is something other than expressing beliefs) is the *only* linguistic function of moral judgments. Let us call such a view “pure noncognitivism.” A weaker noncognitivist perspective places the “only” elsewhere in the sentence: “Moral judgments do not *only* express beliefs, but rather they [also] linguistically function to do so-and-so.” This softened view recognizes that moral judgments may have complex linguistic functions in a way that the pure view does not. While (b) and (c) imply the falsity of pure noncognitivism, they do not imply the falsity of weaker varieties. (This is discussed in Essay 1 of this collection.) In other words, one needs to make theoretical space for a hybrid view which claims that moral judgments both express beliefs *and* perform noncognitive function so-and-so. The beliefs in question may be true in virtue of objective facts (making for a kind of realist–noncognitivist mix), or may be never true (making for an error-theoretic–noncognitivist mix), or may be unjustified (making for a mix of noncognitivism and justification skepticism). If, therefore, moral skepticism is the disjunction of (a), (b), and (c), then the most extreme skeptical view available will combine elements of all three disjuncts: Moral judgments express beliefs but they also perform noncognitive functions, and the beliefs in question are both false and unjustified. I am inclined to accept this extreme view.

This collection traces out the broad strokes of my main metaethical preoccupations and the development of my views since publishing the two aforementioned books. It is divided, somewhat imperfectly, into three parts of four essays each. The essays of Part I (“Error Theory”) follow on most directly from *The Myth of Morality*; the essays of Part II

¹ The title of the introduction that you are currently reading may be seen as some sort of belated Hegelian resolution of the two (not that I flatter myself in thinking that anyone has been waiting with bated breath!)

(“Evolution and Debunking”) upgrade ideas presented in *The Evolution of Morality*; and Part III (“Projectivism and Fictionalism”) has the twin focal points of two related theses that were discussed in the two books but for which I felt there was more to be said.

I stand by the earlier skeptical spirit, but in some ways my views have shifted. Sometimes these are relatively minor adjustments, responses to criticisms, the straightening out of confusions (both mine and others’), or the application of old ideas to new areas. Two of the essays herein (4 and 6) articulate what are in principle quite major revisions in view: where the possibility of conceptual indeterminacy leaves matters that I once thought decidable in the skeptic’s favor potentially undecidable. In these situations I plump for a kind of ecumenical pluralism. Whether one sees this as renouncing the idealistic tendencies of youth, or as coming to appreciate a more nuanced metaethical Big Picture, is (in a phrase from David Lewis that echoes through the papers in question) “mainly a matter of temperament.”

In what follows of this Introduction I will present an overview of each section, though I won’t make an especial effort to describe every essay in turn; generally, they speak for themselves.

Error Theory

To take an error-theoretic stance toward a discourse is to maintain that the discourse consists of assertions that fail to be true. A paradigm familiar example is the atheist’s attitude toward religion. It seems reasonable to hold that most religious utterances are expressions of the speaker’s beliefs (for example, the belief that the gods care whether we keep or break our promises), and the atheist is confident that the world is not furnished with the objects/properties/relations necessary to render these beliefs true (for example, there exist no such caring gods).

A natural way of interpreting the atheist is as holding that religious concepts are reasonably well-defined but that the world contains nothing answering to these definitions, yet an atheist might also maintain that the fault lies with the religious concepts (*God, karma, sin*, and so on) being in some manner hopelessly confused (though still able to appear in assertions²). Similarly, moral error theorists might be moved by various kinds of arguments. One kind of error theorist will seek first to pin down the meaning of moral concepts (*obligation, evil, moral goodness*, and so on) and then argue that the world lacks any properties satisfying these concepts. Another kind of error theorist might admit bewilderment as to what these moral concepts are supposed to denote; she might say to moral believers: “Look, you can’t even seem to agree among yourselves as to what the central moral concepts denote, and, moreover, your disagreement is sufficiently deep-seated that I am left doubting whether these concepts in common usage even denote anything at all.” Just as there are both atheists who believe that the non-existence of gods is necessary and atheists who believe that it is contingent, so too should the label “moral error theorist” cover both those who maintain that moral properties exist at no possible world and those who maintain that their non-

² I realize that this parenthetical qualification raises some substantive questions; unfortunately I lack space to pursue them here.

existence is merely contingent.

The principal alternatives to the error-theoretic view are noncognitivism, moral realism, and (for want of a better title) moral non-objectivism. The last is the view that our moral discourse succeeds in referring to moral properties of a non-objective nature.³ (Think, by analogy, of what it takes for something to be *illegal*. It is certainly a fact that it is illegal to drive on the right-hand side of the road in New Zealand, but it is a fact constituted by our collective decisions/beliefs/practices—it is not, in the relevant manner, an *objective* fact.) The error-theoretic view can be defined in terms of what it agrees with and what it rejects from these alternatives. The error theorist agrees with the realist and the non-objectivist (and not with the pure noncognitivist) that moral discourse functions to express our beliefs about the moral status of various aspects of the world. A standard argument for the error theory (defended by John Mackie and myself) agrees with certain moral realist views (and not with non-objectivist views) about what moral facts would have to be like in order for our moral judgments to be true. Mackie, for example, thinks that a sort of Kantian/Moorean moral realism gets things basically correct at the conceptual level. But the error theorist parts company from the realist (and joins company with the noncognitivist and the non-objectivist) in maintaining that the realistic conception of morality asks too much of the world; there is nothing answering to the Kantian/Moorean conception of moral facts, for example.

Defining the error-theoretic position in this fashion allows us to see that it contains no unique or outrageous sub-thesis; at each step, taken in isolation, one should expect to find many non-error-theorists nodding enthusiastically and offering their own arguments in support. Indeed, the error theorist could get by without developing any novel argument of her own, simply by purloining all her arguments from other metaethical positions. What the error theorist does is combine these arguments and sub-theses in a manner that leads to a radical skepticism that many find unpalatable and threatening (and, perhaps, wishing to retract their supportive arguments!).

In *The Myth of Morality* I raised some fairly blunt considerations against noncognitivism. In Essay 1 of this collection, “Expressivism, Motivation Internalism, and Hume,” I revisit the issue more conscientiously, advocating a weak form of noncognitivism but remaining firm against pure noncognitivism, and careful to clarify how this concession steals no wind from the error theorist’s sails. (I also offer an interpretation of Hume along these lines, though it has a decided “for-what-it’s-worth” air.)⁴ It is not, however, the error theorist’s take on the dispute between the noncognitivist and the cognitivist on which I wish to focus here, but rather the error theorist’s take on the dispute between the realist and the non-objectivist.

³ Earlier in this Introduction, “moral realism” was defined as the conjunction of three theses: belief, truth, and objectivity. Moral non-objectivism is the endorsement of the first two and the denial of the third. Since the *objectivity/non-objectivity* distinction is notoriously difficult to articulate (see Essay 10), this taxonomy is a slippery one.

⁴ I’m no Hume scholar. My “serious” attempts at history of philosophy came early in my career—“Early Stoicism and akrasia” (*Phronesis* 1995) and “Cartesian memory” (*Journal of the History of Philosophy* 1997)—efforts aroused by wonderful teachers at Princeton: John Cooper and Margaret Wilson. I was quite tempted to write a PhD dissertation under Margaret’s supervision on early modern conceptions of secondary qualities, but at the eleventh hour veered into metaethics (with Gilbert Harman).

Earlier I spoke of a kind of error theorist who tries to pin down the meaning of moral concepts (*obligation, evil, moral goodness*, and so on), and I subsequently referred to a “standard argument” for the error theory being one that agrees with certain realists regarding the conceptual commitments of moral discourse. Before proceeding, let me stress that endorsing such a view is not a necessary feature of the error-theoretic position; it is, rather, *a strategy for arguing for* that position. There are other possible grounds for becoming a moral error theorist. One might, for example, come to accept the moral error theory through becoming thoroughly disillusioned with all other metaethical positions. Or alternatively consider, for example, a non-objectivist view with the simple structure “Moral goodness = Nness” (where “Nness” denotes some naturalistic non-objective property). The standard error-theoretic complaint against such a theory is that Nness lacks the distinctive practical authority with which moral properties are essentially imbued, and therefore the non-objectivist’s equation can be rejected by appeal to Leibniz’s law. But instead the error theorist might embrace the non-objectivist’s equation yet argue that it nevertheless leads to an error theory because “Nness” fails to denote any actually instantiated property. I explore this alternative (non-standard) strategy for the error-theoretic conclusion in Essay 3 of this collection, “The Accidental Error Theorist.” This non-standard argument has limits, of course, for it can hardly be argued that “Nness” *must* suffer from this failure. For those many occasions where the non-objectivist’s “Nness” succeeds in picking out an actually instantiated property, the error theorist must return to the standard strategy of rejecting the reasonableness of the equation.

One can consider the standard strategy either generally or specifically. Speaking generally, the strategy identifies some thesis to which moral discourse is committed and then argues that the thesis is false. Thus this argumentative strategy faces two kinds of opponent: those who reject the error theorist’s conceptual step engage in a semantic dispute, while those who reject the latter step disagree about what features the world contains. The semantic dispute is, I think, the trickier to prosecute, for it is challenging to know how best to articulate the difference between a discourse being *committed* to some thesis (in the sense that dropping that thesis would amount to changing the subject) and a discourse being such that people sometimes/often/always have some false beliefs concerning it. Considering the strategy in general terms leaves open what the specific problematic thesis (or theses) might be. Speaking more specifically, perhaps moral discourse is committed to a problematic notion of desert, or a problematic notion of autonomy, or a problematic notion of personhood, or a problematic epistemology. The specific version of the argument that has dominated discussion, however, is that moral discourse is committed to a problematic notion of objective practical authority.

I don’t mind confessing that I have never really nailed the conceptual step of this argument to my own satisfaction; but, on the other hand, I have never found the efforts of those opposed to the step terribly persuasive either. Part of the challenge is to render the idea of “objective practical authority” in a sufficiently clear manner, and part of the challenge is to establish that moral discourse is committed to such a thing. Regarding the former, although in *The Myth of Morality* I gave it my best shot—hypothesizing that this authority might be understood by reference to certain kinds of practical reasons—even then I felt that it may be asking too much of the error theorist to provide this much specificity. After all, error theorists may worry that there is something utterly mysterious about the kind of authority with which

moral properties are essentially imbued; they may consider morality to be something like a pseudoscience, and the concepts (or pseudo-concepts) employed by a pseudoscience often defy clarification. By analogy: As an atheist I do not believe that anything is literally *sacred*, yet I don't suppose I could do an especially good job of articulating precisely what it means for something to be sacred; it is, rather, the very obscure and nebulous quality of the concept that encourages my disbelief. The fact that the atheist may be unable to draw a very precise bead on concepts like *sacred*, *God*, or *heaven* should not count against the reasonableness of his atheism; it hardly seems incumbent upon *him* to give definition to these ideas! That said, of course the atheist needs to have *some* idea of the content of these concepts, or else he could not object to someone who tries to reassure him that "sacred" means nothing more than *salubrious*, that "God" just means *love*, and that "heaven" is a word that denotes *Tahiti*. (And how can one reasonably doubt the existence of salubrity, love, and Tahiti?) The atheist needs to be sufficiently conceptually *au fait* to protest that these religious concepts are used to denote something other than such innocuous entities. In a similar way, when faced with a moral naturalist who proposes to identify moral properties with some kind of innocuous naturalistic property—the maximization of happiness, say—the error theorist will likely object that this property lacks the kind of "normative oomph" that permeates our moral discourse. Why, it might be asked, should we care about the maximization of happiness any more than the maximization of some other mental state, such as surprise?⁵ Yet (the error theorist may continue) moral properties are those about whom the failure to care counts as a transgression; this is in fact the *whole point* of having a moral discourse. The error theorist's defense here relies on identifying a conceptual commitment of morality, though not necessarily a precisely defined one.

There has been some work in recent years trying to ascertain the conceptual commitments of moral discourse via empirical methods. (See Goodwin and Darley 2008, 2012; Sarkissian et al. 2011; Uttich et al. 2014.) A typical experimental design is to have test subjects assess various kinds of normative claims (such as "Wearing pajamas and a bathrobe to a seminar is wrong behavior") as true or false or a matter of opinion. Of course, even if all subjects are adamant that moral norms have quality Q it would not follow that Q is an indispensable conceptual commitment of moral discourse, but one might be tempted to conclude that such a result would at least show that *not-Q* is not such a commitment. This temptation should be treated with care. What students tick on a questionnaire need not reflect their real moral commitments; these emerge only in the testing ground of actual practice. Such experiments are also prone to over-interpretation or misinterpretation. For example, all the experiments that I have seen along these lines seem to target the extent to which subjects make pronouncements in line with a *relativistic* metaethical view, yet the conclusion drawn over and over again is that this has some bearing on the subjects' attitudes toward moral *objectivity*. But any metaethicist worth his or her salt will tell you that the opposite of relativism is *absolutism*, not objectivism. Relativism-versus-absolutism does not figure in the

⁵ I should quite like to see someone defend *thaumastitic utilitarianism*: the thesis that one is obligated to maximize surprise. I suspect that the silliness of the enterprise would cast some light on forms of utilitarianism that are taken seriously, like hedonic utilitarianism.

criteria distinguishing moral realism from non-realism.⁶ The proposal that moral discourse is imbued with a kind of objective practical authority (for want of a better phrase) is entirely consistent with a relativistic moral discourse.

The seeming impasse between the error theorist and her critics over what is and is not a conceptual commitment of moral discourse should put one in a diagnostic mood. How can thoughtful and intelligent people disagree over the content of their own concepts? I have become increasingly sympathetic to the idea that the line between a discourse having a faulty conceptual commitment and a discourse being such that users tend to have false beliefs about its subject matter is an extremely blurry one—not just in an epistemic sense, but in the sense that there often is no fact of the matter. The moral concepts are indeterminate beasts, available for different equally legitimate but non-equivalent precisifications. Some such precisifications may find something answering to them in the world, while others of the same concept may not. In other words, certain forms of moral naturalism may be permissible to maintain, but the moral error theory may also be permissible to maintain. This viewpoint, and the resulting metaethical pluralism, is the subject of Essay 4 of this collection. In this Essay I also make a plea for the pragmatic value of the error-theoretic position; I do this to provide a counterweight to what I suspect will be the widespread assumption that this kind of Scottish verdict does not really encourage pluralism but rather plays into the hands of the moral naturalist. Here we get to some really rather deep issues about what the point of philosophy is supposed to be: Should a theory that vindicates the vernacular be preferred to one that does not? I am yet to encounter an argument that convinces me that a positive answer to this question is mandatory; rather, I feel more inclined to side with the bleak romanticism expressed by Edward Gorey: “My mission in life is to make everybody as uneasy as possible. I think we should all be as uneasy as possible, because that’s what the world is like.” Well, maybe not as uneasy *as possible*, but I certainly dislike the image of philosophy as a tool for providing a soothing background voice reporting the world to be as we believe it to be, moral facts and all.

Evolution and Debunking

The atheist accuses the vast majority of human beings of embracing a doxastic error of enormous proportions. Yet if asked “Where does this error come from?” I don’t think the atheist’s position is particularly undermined if she admits that she has no idea beyond, perhaps, having a somewhat pessimistic view of humans as silly and gullible creatures. In the same way, I don’t judge it incumbent on the moral error theorist to offer a theory of why nearly all humans have fallen into the mistaken ways of moral thinking. Yet it also seems reasonable to claim that both the atheist and the moral error theorist would strengthen their positions somewhat if each could provide a plausible hypothesis concerning how such systematic errors might arise in human thought. In *The Myth of Morality* (chapter 6) I turned to evolution as a plausible explanation of human moral thinking, observing that if Darwinian selection has wired the human brain for moral judgment, it is because moral judgment

⁶ See Joyce 2007.

enhanced our ancestors' reproductive fitness (relative to competitors) in a way that might be accomplished even by encouraging false beliefs.

While developing these thoughts I became intrigued by another possibility: that one might argue for a moral error theory *on the basis of* these evolutionary considerations. My next book, *The Evolution of Morality*, was intended to investigate this argument comprehensively, but in the end a great many other interesting things arose and only the final chapter was devoted to the debunking argument (as it has become known). By the time I wrote this final chapter it had also become clear to me that the error-theoretic conclusion is beyond the reach of the argument; the correct skeptical conclusion is of an epistemological nature: that all moral judgments are *unjustified*.⁷ Thus *The Evolution of Morality* does not argue for moral anti-realism at all.

At this time Sharon Street was writing an influential paper, "A Darwinian dilemma for realist theories of value," whose argument starts with the same premise as mine: that the human tendency to assess the world in moral terms is a biological adaptation. (Neither of us, of course, purport to be in a position to assert this moral nativist thesis with great confidence; our arguments are conditional.) From this similar starting point, though, our arguments diverge importantly. Street's target is moral realism; she uses evolutionary considerations to cast doubt on the existence of objective moral facts. Since, however, she is willing to accept a constructivist metaethical view—according to which moral facts have a non-objective status—she is no error theorist. The reason I mention Street's argument at this point is that it appears to me that because our arguments came out at much the same time there has been a tendency to lump them together. A cottage industry focused on "evolutionary debunking arguments of morality" has sprouted (most of it critical), and while much of it is very worthwhile, some of it is also, in my opinion, based on fundamental misunderstandings. I take some blame for this, since my attempt to make the case in *The Evolution of Morality* is flawed and unclear in various ways; I was still squinting to discern the structure of my own argument. Essays 7 and 8 of this collection represent my attempts to develop the argument more cleanly. (The ordering of this pair of essays reflects the sequence in which they were written, but whereas Essay 8 was published in 2013, the project for which Essay 7 was originally commissioned was delayed and remains uncertain; hence Essay 7 is published here for the first time.)

As I mentioned, the debunking argument is conditional: It relies on an empirical premise concerning the evolution of morality which is yet to be established. Before we come to the debunking argument, then, it makes sense to examine this nativist hypothesis carefully. This is done in Essays 5 and 6 of this collection. Here metaethics is put mostly to one side, and instead I adopt my role as a philosopher of biology—albeit, I'll be the first to admit, as something of a Sunday painter.

One of the flaws of *The Evolution of Morality* is that I had not yet come to appreciate fully the difficulties surrounding the key notion of *innateness*; my subsequent papers on the

⁷ In a manner very much bringing to mind Bertrand Russell's comparison of theft and honest toil, I decided that one might nevertheless call this epistemological conclusion a version of "error theory" (Joyce 2006: 223). This was, I now see, a foolish stipulation, and I discourage anyone from adopting this usage.

topic are more cognizant of this. Essay 5, “The Origins of Moral Judgment,” focuses on the question of how we are to distinguish traits that are adaptations from those that are byproducts; Essay 6, “The Many Moral Nativisms,” looks at different meanings of “innateness.” Both papers are to some extent motivated by desire to respond to a moral anti-nativist movement that emerged in the years after my book. It’s not so much that I think that the anti-nativists (or “spandrel theorists” as I sometimes call them) are definitely mistaken; I am more interested in diagnosing the conceptual framework of the disagreement and straightening out misunderstandings.

Something to which all should agree is that there is no point in arguing over whether some trait is or is not innate (regardless of what notion of *innateness* is under discussion) unless we have a reasonable grasp of what the trait is that we are arguing about. In the present case, that trait is the capacity to make moral judgments. So: What is it to make a moral judgment? At one time I was keen to argue for a particular kind of answer, but years of encountering critics with very different views have led me to suspect that there might be no fixed answer; the matter is simply indeterminate in various ways. This suspicion, obviously, matches that voiced in Essay 4. There the worry was that our moral concepts are indeterminate, so that the judgment “ ϕ is morally prohibited” (say) may be reasonably considered true or reasonably considered false, depending on how one precisifies the idea of *moral prohibition*: Understood one way it picks out an actual property; understood another way it fails to do so. Here the worry is that what it takes to make a moral judgment suffers from indeterminacy, so that the sentence “S judges that ϕ is morally prohibited” (say) may be reasonably considered true or reasonably considered false, depending on how one precisifies the idea of making this kind of moral judgment. The node of indeterminacy examined in Essay 6 is that one understanding of *moral judgment* constructs the phenomenon entirely out of noncognitive building blocks, whereas another requires of it more conceptual sophistication. The possibility then arises that understanding the phenomenon one way may pick out a capacity that is innate, but understanding it another way may pick out a capacity that is not innate; hence, the debate between the moral nativist and the anti-nativist may be undecidable.⁸

Projectivism and Fictionalism

A question that has always divided philosophers (if one may be so crass as to speak of such a thing) is how close the world really is to how it seems to us. Those of a skeptical temperament, who lean toward thinking that it may not be very close at all, have always faced the further question of how one should respond, in practical and psychological terms, to this human condition. An ancient response—as ancient, at least, as Pyrrho and his followers—is that it should not make too much difference at all; we can and should carry on living in accordance with appearances. Projectivism is a way of making sense of the skeptical answer to the first question, and fictionalism is a way of making sense of the skeptical answer to the

⁸ Please note the modal qualifications in my expression. In neither Essay 4 nor Essay 6 am I asserting that there *is* this indeterminacy; I am merely exploring the possibility in a sympathetic mood.

second question. Both theories rely on a similar distinction being drawn between how the world appears to us (/is experienced by us) and our critical understanding of what is really going on. Both theories also may have general applicability or can be restricted to the moral realm.

Moral projectivism is the view that our emotional life creates and colors our moral experience. One's seeing an act of violence as *wrong* (say) is not the result of successfully tracking the presence of wrongness, but is rather the result of having an emotion like disapproval which plays an active role in constructing how that action seems to us. The moral skeptic need not endorse projectivism, but doing so can be useful for the skeptic inasmuch as it can provide an explanation for how this systematic mistake in human thinking comes about. The skeptic can even argue that humans have been designed by natural selection to perform this kind of projective error—that the mechanisms underlying the process were adaptive for our ancestors—thus moral projectivism sits quite comfortably with the moral nativist program.

Part of my interest in writing about this topic was born of a frustration with a widespread view of moral projectivism holding it to be the exclusive province of the noncognitivist. I recall as a graduate student feeling anxious that I must have weird or mistaken ideas about what projectivism is; I heard people speaking of “projectivist semantics” and couldn't work out what this might mean. Around 2009 I wrote two papers that tried to straighten things out, at least to my own satisfaction; both attempt to clarify the relation between moral projectivism and various positions in metaethics. Essay 9, “Patterns of Objectification,” takes its title from a phrase used by Mackie (who prefers “objectification” to “projectivism”), and investigates what role projectivism plays (or could play or should play) in his argument for a moral error theory. (In this respect, Essay 9 could easily plug into Part I of this collection.) Essay 10, “Is Moral Projectivism Empirically Tractable?,” teases apart the various sub-theses of different potential versions of projectivism. Here I am keen to push past the metaphors (“spreading,” “gilding,” “projecting”) to get at some literal theses which can then be (in principle) tested for truth. I continue to think that moral projectivism is a theory ripe for experimental investigation. This testable core of moral projectivism is, however, metaethically neutral.

Moral fictionalism is a proposal for how the error theorist might carry on.⁹ There is no need to eliminate morality entirely from our thoughts and language (the fictionalist declares); we can maintain its use as a kind of functional fiction. The fictionalist doesn't propose that we maintain morality as a set of *beliefs* and *assertions* (for the advice “Carry on believing something that you believe to be false” is likely to prove problematic in various obvious ways), but rather in a fashion reminiscent of a kind of highly-played make-believe. Like all pieces of advice, the reasonableness of the fictionalist's proposal depends on the outcome of a cost-benefit analysis. Making a make-believe of morality is not, of course, going to produce the same costs and benefits as sincerely believing morality. Believing that ϕ is morally

⁹ Moral fictionalism can also be considered as a proposal for how a justification skeptic might carry on—though the way that the statement of the theory is worded would have to be adjusted in various ways. Think by analogy of religious fictionalism: The proposal might be offered both to atheists and agnostics. (Here “agnosticism” is used in the popular way to denote a position of doxastic indecision.)

obligatory is likely to strengthen one's motivation to ϕ more robustly than make-believing that ϕ ing is obligatory—so on the assumption that ϕ ing is beneficial, moral belief is better than moral make-believe. But this is not the relevant comparison, since for the skeptic the option of belief is gone. Rather, we must compare the fictionalist proposal with that of the eliminativist, who counsels that we drop moral thinking and moral talk altogether. So long as thinking and talking of ϕ ing as morally obligatory (even in a fictionalist manner) *in some way* enhances one's motivation to perform the beneficial action (with no countervailing cost), then the fictionalist has made her case. (And it need not be good advice for *all* the people *all* the time; that is not a requirement we ever put on something's counting as good advice!)

But how can mere pretense have a sufficient impact on one's motivations as to affect significant costs and benefits? The quick answer to this question (as to so many others) is that humans are strange creatures. And one of our stranger quirks is our interest in fictions. It is easy to imagine intelligent creatures for whom engaging with characters who never existed and narratives that never happened holds no attraction whatsoever. But that's not us. Stories and images affect our moods, emotions, and motivations. The error theorist who immerses herself in a moral fiction takes advantage of these peculiar aspects of her own psychology. She is likely acutely familiar with classifying her social environment in moral terms, so she continues the habit of bringing moral concepts to bear on practical problems, allowing moral emotions like disgust, anger, and guilt to wash through her. All going according to plan, this moral fiction doesn't encourage her to do anything that she would not upon reflection choose to do anyhow on non-moral grounds. But moral thinking has some advantages over careful non-moral thinking, in that it can be fast and frugal, less prone to self-sabotaging rationalizations, able to banish practical calculation from the decision procedure when the very act of calculating is suboptimal. If the nativist is correct, then moral thinking is a well-honed tool that suits our psychological configuration; we are comfortable with its contours.

Essay 11, "Moral Fictionalism," develops the case I made in *The Myth of Morality*—struggling (as is always the case when I discuss this topic) to make a weird theory seem a bit less weird. (Some responses to critics of moral fictionalism also appear toward the end of Essay 2 of this collection.) I'm not sure whether my tentative advocacy of moral fictionalism over the years has won many converts, but in a sense this is how it should be, since the cost-benefit analysis upon which the theory rests involves so many unknown variables and counterfactuals that anyone who claims with confidence to believe that moral fictionalism is *correct* has probably missed the point.

The final essay of this collection, "Psychological Fictionalism, and the Threat of Fictionalist Suicide," takes what I have learned from thinking about moral fictionalism and applies it to another potential error-theoretic view: concerning the entities of folk psychology. It becomes quickly apparent that the psychological fictionalist faces some special problems that do not trouble other forms of the theory, and this paper attempts a fix. I should say that I'm not particularly inclined to doubt the existence of such things as beliefs and desires, though nor am I willing to declare that such doubt is misplaced. It is good to know, though, that even if one were to embrace such doubt, metaethical theories such as the error theory and fictionalism could still be identified, expressed, and advocated.

In preparing this collection I have resisted the temptation to mend any content which I now

judge incorrect or at least think could be better expressed. (If I started down that road, where would I stop?) An exception is the addition of the first footnote to Essay 11, where I could not let my earlier self's claims go unchallenged. And on one occasion I have corrected a reference to the optative mood to the cohortative mood. (What was I thinking?!) Apart from that, I have restricted myself to fixing typos, updating citations, and imposing a uniform spelling, punctuation, and formatting structure. I apologize for the occasional repetitions found in this collection; it is in my nature as a philosopher to go over the same ground frequently, and in this way gradually and incrementally make progress (or so the hope goes). The papers herein were produced while I was at the Australian National University, at the University of Sydney (which included a sabbatical near Périgueux), and then at Victoria University of Wellington. My thanks to the publishers of the various books and journals from which these papers are drawn.

Mitimiti, New Zealand
January 2015

References

- Anscombe, G. E. M. 1958. "Modern moral philosophy." *Philosophy* 33: 1-19.
- Goodwin, G. and Darley, J. 2007. "The psychology of meta-ethics: Exploring objectivism." *Cognition* 106: 1339-1366.
- Goodwin, G. and Darley, J. 2012. "Why are some moral beliefs perceived to be more objective than others?" *Journal of Experimental Social Psychology* 48: 250-256.
- Joyce, R. 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Joyce, R. 2007. "Moral anti-realism." In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. <<http://plato.stanford.edu/archives/sum2015/entries/moral-anti-realism/>>
- Sarkissian, H., Park, J., Tien, D., Wright, J. C., and Knobe, J. 2011. "Folk moral relativism." *Mind & Language* 26: 482-505.
- Street, S. 2006. "A Darwinian dilemma for realist theories of value." *Philosophical Studies* 127: 109-166.
- Uttich, K., Tsai, G., and Lombrozo, T. 2014. "Exploring metaethical commitments: Moral objectivity and moral progress." In *Advances in Experimental Moral Psychology*, edited by H. Sarkissian and J. C. Wright: 188-208. London: Bloomsbury Academic.