

# *The Myth of Morality*

*Richard Joyce*  
*University of Sheffield*



**CAMBRIDGE**  
**UNIVERSITY PRESS**

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS  
The Edinburgh Building, Cambridge CB2 2RU, UK  
40 West 20th Street, New York, NY 10011-4211, USA  
477 Williamstown Road, Port Melbourne, VIC 3207, Australia  
Ruiz de Alarcón 13, 28014 Madrid, Spain  
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Richard Joyce 2001

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

*Typeface* Bembo 10/12 pt.    *System* L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> [TB]

*A catalogue record for this book is available from the British Library*

*Library of Congress Cataloguing in Publication Data*

Joyce, Richard, 1966–  
The myth of morality / Richard Joyce.  
p. cm.

Includes bibliographical references and index.

ISBN 0 521 80806 5

1. Ethics. I. Title.

BJ1012 .J69 2001

170–dc21

2001025740

ISBN 0 521 80806 5 hardback

# Contents

<i>Preface</i>	<i>page ix</i>
1 Error theory and motivation	1
2 Error theory and reasons	30
3 Practical instrumentalism	53
4 The relativity of reasons	80
5 Internal and external reasons	106
6 Morality and evolution	135
7 Fictionalism	175
8 Moral fictionalism	206
Epilogue: Debunking myths	232
<i>Select bibliography</i>	243
<i>Index</i>	247

# 1

## *Error theory and motivation*

### 1.0 FAULTY FRAMEWORKS

When European explorers first interacted with cultures of the South Pacific, they found the islanders employing an unfamiliar concept: a type of forbiddenness called “tapu.” Europeans developed this into the familiar English term “taboo,” but what we mean by “taboo” is quite unlike what the Polynesians meant. (It is to signal this difference that I have chosen the Maori word “tapu” over “taboo.”) It is not the case, for instance, that “tapu” may be translated into “morally forbidden,” with accompanying understanding that the Polynesians have different beliefs from Europeans concerning which actions are forbidden. “Tapu” centrally implicates a kind of uncleanness or pollution that may reside in objects, may pass to humans through contact, may be then transmitted to others like a contagion, and which may be canceled through certain ritual activities, usually involving washing. This is not a concept that we employ, though one may find something similar in ancient Roman and Greek texts.<sup>1</sup>

If one of the European explorers had a penchant for metaethics, what would he say about the Polynesians’ discourse? He would naturally take them to have a *defective* concept; no judgment of the form “ $\phi$  is tapu” is ever true (so long as “ $\phi$ ” names an actual action<sup>2</sup>) because there simply isn’t anything that’s tapu. Saying this implies nothing about how tolerant in

<sup>1</sup> The Roman term is “sacer” and the Greek “agos.” Cf. S. Freud, *Totem and Taboo* [1913] (London: Routledge & Kegan Paul, 1950), p. 18. Whatever ancient European equivalents there may have been to the Polynesian concept, they belonged to a bygone era by the time Maimonides (twelfth century) was trying to explain away the somewhat embarrassing references to “abominations” in Leviticus. I should say that my selection of “taboo” as an illustration is inspired by comments by Alasdair MacIntyre, *After Virtue* (Notre Dame, Ind.: University of Notre Dame Press, 1984), pp. 111–13.

<sup>2</sup> Throughout this book, the symbols “ $\phi$ ” and “ $\psi$ ” generally stand for actions. However, sometimes they stand in for verbs (“ $\phi$ ing is good,” “I want you to  $\phi$ ”) and sometimes they do the work of nouns (“The action  $\phi$ ”). I find this convenient and not noticeably jarring.

## *The myth of morality*

attitude the explorer would be of the Polynesians' discourse; his identifying their discourse as "defective" is consistent with recognizing that it serves them well, and choosing not to point out to them their error. It is also consistent with his electing to employ the concept in sincere assertions of the form " $\phi$  is tapu," but only when this is an anthropological judgment, elliptical for "For the islanders,  $\phi$  is considered tapu." It would be strange for him to make non-elliptical judgments of the form " $\phi$  is tapu" if he thought, as he naturally would, of the whole framework as mistaken. And in all of this the explorer would be quite correct: "tapu" is certainly not a term that I apply (non-elliptically), and the reason I don't is that reflection on the kind of "metaphysical uncleanness" that a literal application of the term presupposes leads to recognition that *nothing* is tapu. I treat the Polynesians' discourse – with all due cultural respect – as systematically mistaken.<sup>3</sup>

But how could it be that a discourse that is familiar to a group of perfectly intelligent people – one that they employ every day without running into any trouble or confusion – is so mistaken? After all, the users of the term unanimously apply it to certain types of action, unanimously withhold it from other actions, and perhaps even agree on a range of types of action which count as a "gray area." Doesn't all this amount to the predicate "... is tapu" having a non-empty extension? To see that the answer is "No" we might reflect again on the European explorer's own defective concept: *phlogiston* (we'll assume that his travels predated Lavoisier). The chemists Stahl, Priestley, *et al.*, were equally able to agree on the extension of their favored predicate. Indeed, they were able ostensibly to pick out paradigm examples of phlogiston: they could point to any flame and say "*There is the phlogiston escaping!*" And yet for all that they were failing to state truths, for there wasn't any phlogiston. Clearly, when speakers used the predicate "... is phlogiston" something more was going on than merely applying it to objects. What sentenced the predicate to emptiness, despite its ostensive paradigms, was that users of the term (considered collectively) thought and said certain things *about* phlogiston such as "It is that stuff stored in bodies," "It is that stuff that is released during combustion" and

<sup>3</sup> Cf. the anthropologist E. E. Evans-Pritchard, who, in his influential study of the Azande people, writes of their belief in witchcraft that they display "patterns of thought that attribute to phenomena supra-sensible qualities which...they do not possess"; that "witches, as the Azande conceive them, cannot exist". *Witchcraft, Oracles and Magic among the Azande* [1937] (Oxford: Clarendon Press, 1965), pp. 12, 63.

“Soot is made up almost entirely of it,” and these concomitant statements are false.<sup>4</sup> It’s not that any competent user of the word “phlogiston” was disposed to make these statements – our Pacific explorer, for example, may have had only a rudimentary grasp of the theory, despite being considered perfectly competent with the term. But he would have been willing to defer to the firm opinions of the experts in chemistry of the day, and *they* would have said these things.

Let us say that the above three propositions concerning phlogiston were firmly held by the experts. Let us pretend, further, that these three propositions have a kind of “non-negotiable” status. What I mean by this is the following: Imagine that we were to encounter a population speaking a quite different language to our own, most of which we have translated and tested to our satisfaction, and we find that they have a concept that appears rather like our concept of *phlogiston* (say, it plays a central role in explaining combustion and calcification) – call their term “schmogiston” – but we also find that they don’t endorse one of the three propositions about schmogiston. If that would be sufficient for us to decide *not* to translate “schmogiston” into “phlogiston,” then the proposition in question must be a non-negotiable part of our concept *phlogiston*. It may not be that any *one* proposition is non-negotiable: perhaps we would be content with the translation if any two of the “schmogiston”-propositions were dissented from, but if the speakers dissented from all three (i.e., they said “No” to “Is schmogiston released during combustion?”, “Is schmogiston stored in bodies?”, and “Is soot made up of schmogiston?”) then we would resist the translation – we would conclude that they weren’t talking about *phlogiston* at all. In such a case we might call the disjunction of the three propositions “non-negotiable.”

This translation test gives us a way of conceptualizing what we mean by a “non-negotiable” proposition, though I don’t pretend that it gives us a widely usable decision procedure (involving, as it does, a complex counterfactual about when we would or wouldn’t accept a translation scheme). The point is to make sense of a distinction. On the one hand, we might have a discourse that centers on a predicate “. . . is P,” involving the assertion of a variety of propositions – “*a* is P,” “*b* is not-P,” “For any *x*, if *x*

<sup>4</sup> My rudimentary knowledge of phlogiston theory is derived from F. L. Holmes, *Lavoisier and the Chemistry of Life* (Madison, Wisc.: University of Wisconsin Press, 1987), and J. R. Partington and D. McKie, *Historical Studies on the Phlogiston Theory* (New York: Arno Press, 1981).

### *The myth of morality*

is P, then  $x$  is Q,” etc. – and when we discover that we’re *mistaken* about one or more of these things – e.g., we discover that some things that are P are not Q – we don’t decide that the whole “P discourse” has been a disastrous mistake; we simply change our minds about one aspect of it: we stop making the conditional claim and carry on much as before. On the other hand, there are some discourses regarding which the discovery that one or more of the things we’ve been assenting to is mistaken leads us to throw in the towel – to stop using the discourse altogether. The latter describes what happened in the phlogiston case: the discovery that we had been wrong in thinking that there is a stuff stored in combustible bodies and released during burning was sufficient for us to decide that there is no phlogiston at all. When Lavoisier gave us the concept *oxygen*, it wasn’t available for Stahl to say “Well, this stuff that Lavoisier is calling ‘oxygen’ just *is* what I’ve been calling ‘phlogiston’ all along – I was just mistaken about its being stored and released during combustion.” The belief that phlogiston is stored and released was a *non-negotiable* part of phlogiston discourse – the falsity of this belief was sufficient to sink the whole theory.

Now we can see how a smooth-running, useful and familiar discourse, apparently with clear paradigms and foils, could be systematically flawed. The users of the target predicate (or the experts to whom most users firmly defer) assent to a number of non-negotiable propositions – propositions which would play a determinative role in deciding whether or not a translation goes through – and a critical number of these non-negotiable propositions are, in fact, false. This might be how our explorer-cum-metaethicist conceives of the concept *tapu*. If the Polynesians had merely used “tapu” as a kind of strong proscription, and thought, say, that public nudity is *not* tapu but burying the dead is, then (*ceteris paribus*) this would not have prevented the explorer from translating “tapu” into “morally forbidden” while ascribing to the Polynesians some different beliefs about which actions are morally forbidden. But given the kind of robust metaphysics surrounding the notion of tapu – centrally involving supernatural and magical forces – no obvious translation (along with the ascription of different beliefs) was available. The explorer doesn’t just attribute to the Polynesians a set of false beliefs – he attributes to them a faulty *framework*. (I don’t intend this to sound culturally critical – the eighteenth-century European is certainly no better off with his concept *phlogiston*, and nor, I will argue, are we with our familiar moral concepts.)

The terminology introduced by John Mackie to describe this situation is that the European explorer holds an *error theory* regarding the historical Polynesians' "tapu discourse," just as we now hold an error theory with respect to phlogiston theory (for shorthand we can say that we are "error theorists about phlogiston").<sup>5</sup> We don't hold an error theory about *any* discourse involving the term "phlogiston," of course. People continue to talk about phlogiston long after Lavoisier's discoveries – saying things like "Georg Stahl believed in phlogiston," "Phlogiston doesn't exist" – and *that* phlogiston discourse is just fine. What we don't do is assert judgments of the form "*a* is phlogiston" (or make assertions that imply it). It is only a discourse that made such assertions, such as the one existing through the seventeenth century, regarding which we are error theorists.

An error theory, as we have seen, involves two steps of argumentation. First, it involves ascertaining just what a term *means*. I have tried to explicate this in terms of "non-negotiability," which in turn I understood in terms of a translation test (but there may be other, and better, ways of understanding the notion). So, in artificially simple terms, the first step gives us something roughly of the form "For any *x*, *Fx* if and only if *Px* and *Qx* and *Rx*." We can call this step *conceptual*. The second step is to ascertain whether the following is true: "There exists an *x*, such that *Px* and *Qx* and *Rx*." If not, then there is nothing that satisfies "... is *F*"<sup>6</sup> Call this step *ontological* or *substantive*. The concept of *phlogiston* – with its commitment to a stuff that is stored in bodies and released during combustion – and the concept of *tapu* – with its commitment to a kind of contagious pollution – do not pass the test.

<sup>5</sup> J. L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin Books, 1977). All textual page references to Mackie are to this work. See also his "A Refutation of Morals," *Australasian Journal of Philosophy* 24 (1946), pp. 77–90.

<sup>6</sup> This way of representing the problem is known as giving a "Ramsey sentence" of a term. See F. Ramsey, "Theories," in D. H. Mellor (ed.), *Philosophical Papers* (Cambridge: Cambridge University Press, 1990), pp. 112–36. See also M. Smith, *The Moral Problem* (Oxford: Blackwell, 1994); D. Lewis, "How to Define Theoretical Terms," *Journal of Philosophy* 67 (1970), pp. 427–46, and *idem*, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, supplementary volume 72 (1989), pp. 113–37. Those familiar with Lewis's work will know that satisfying the Ramsey sentence does not require finding something that "exactly fits," but something "close enough" will often suffice. The way that I have presented the matter, however, propositions are "weighted" (in a desirably vague manner) *before* the Ramsey sentence is constructed; i.e., a vagueness is intended to be respected in the procedure whereby we establish "non-negotiability." The end result, I take it, is the same. That is, there is no difference between (i) putting forward the Ramsey sentence " $\exists x (Fx \ \& \ Gx \ \& \ Hx)$ " and claiming that the sentence is satisfied so long as two of the three conjuncts are satisfied by some object, and (ii) putting forward the sentence " $\exists x$  (Two out of three of the following: { *Fx*, *Gx*, *Hx* })."



## *The myth of morality*

### 1.1 THE SEMANTICS OF AN ERROR THEORY

Before proceeding, we shall consider what might be said, semantically, about such an erroneous discourse. It has been claimed that an error theory is the view that all the judgments that comprise the discourse are false.<sup>7</sup> This seems unlikely. For a start, within, say, phlogiston discourse – even that employed by Stahl – there would be a smattering of true claims: he may assert things like “Priestly also believes in phlogiston,” “If we were to burn X and phlogiston were to escape, then X would get slightly heavier” (sometimes phlogiston was considered lighter than air). One might try to define “discourse” more carefully, so as to rule out these embedded claims – claiming, perhaps, that they are not *central* to the discourse – but I don’t know how that might be done in a systematic way, and I see no reason why such a claim might not be a central one.

A different worry would be that some of the claims might best be considered neither true nor false, especially if we take on board certain views from philosophy of language. Peter Strawson argued that an utterance of “The present king of France is wise” is neither true nor false (if uttered in the present), due to the referential failure of the subject-term of the sentence.<sup>8</sup> Earlier I had Stahl making claims of the form “*a* is phlogiston,” but this was rather artificial – surely he also made numerous claims of the form “Phlogiston is F.” It would appear then, that if Strawson is correct, the latter kind of judgment ought to be considered neither true nor false. We can take this even further. Frank Ramsey argued that in a sentence of the form “*a* is F,” which element is the subject and which element the predicate is entirely arbitrary.<sup>9</sup> For any such sentence we may nominalize the predicate (provide a name for the property) and make *it* the subject of the sentence, and thereby express the same proposition. So “Socrates is wise” becomes “Wisdom is had by Socrates”; “*a* is phlogiston” becomes (less elegantly) “Phlogistonness is had by *a*”; “Mary is next to John” becomes (I suppose) “The relational property of *being next to* is had by the pair <Mary, John>.” If we combine Strawson and Ramsey’s views, we get

<sup>7</sup> For example, by G. Sayre-McCord, “The Many Moral Realisms,” in *The Spindel Conference: The Southern Journal of Philosophy*, supplementary volume 24 (1986), pp. 1–22.

<sup>8</sup> P. F. Strawson, “On Referring,” in A. Flew (ed.), *Essays in Conceptual Analysis* (London: Macmillan & Co. Ltd., 1956), pp. 21–52.

<sup>9</sup> F. P. Ramsey, “Universals” [1926], in D. H. Mellor (ed.), *Philosophical Papers*, pp. 8–30. For a much more recent discussion of the same idea, see A. Oliver, “The Metaphysics of Properties,” *Mind* 105 (1996), pp. 61–8. For defense of Ramsey see L. Nemirow, “No Argument Against Ramsey,” *Analysis* 39 (1979), pp. 201–09, and M. C. Bradley, “Geach and Strawson on Negating Names,” *Philosophical Quarterly* 36 (1986), pp. 16–28.

the interesting result that if “Fness” fails to refer, then an assertion of the form “*a* is F” ought to be considered neither true nor false.

Just how far this takes us is difficult to say without embarking on a detailed course of metaphysics, for it is not clear what an abstract singular term like “Fness” refers to – what *kind* of thing is a property? I will not advance an answer to this question, but will indulge in a couple of suggestive comments. Nominalizing predicates may smack of Platonistic tendencies, but this appraisal would be unfair. Whatever account we give of *satisfying a predicate* – however metaphysically austere our preference – we can give a matching account of property names. Quine, for example, understands having properties in terms of class membership; the nominalized predicate would then be a name for the class.<sup>10</sup> Referential failure for the class name would require the non-existence of the class, but, since classes are usually thought of as abstract entities, it is hard to know what this would amount to.

One thing it might amount to is this: if the predicate “. . . is F” has an empty extension across all possible worlds, then “Fness” fails to refer. Typically, concepts that we think of as defective will not satisfy this criterion. For example, the natural thing to say about “. . . is phlogiston” is that it has an empty extension in the actual world, but has a non-empty extension in other possible worlds: phlogiston theory is false, but only contingently so. In other words, there *is* a property which “phlogistonness” denotes, it is just that nothing in the actual world has this property. It is possible, however, that a predicate might suffer a more serious kind of defect: if it were in some manner self-contradictory, or if it entailed a strong modal claim which turned out to be false, then we might conclude that its extension is empty in all possible worlds. (Later I will discuss concrete cases for which this might be argued.) I am suggesting, though not arguing, that this may be sufficient for the conclusion not merely that nothing has the property in question, but that there simply is no property at all.

Whether we accept the latter unusual view is a matter of how we choose to theorize about properties, which in turn is dependent on weighing the theoretical costs and benefits of various contending positions, and none of this is attempted here, bar one comment. An obvious rejoinder from the Quinean is that property names *do* succeed in referring even when they have empty extensions over all possible worlds – they refer to the null set.

<sup>10</sup> See W. V. Quine, *Mathematical Logic* (Cambridge, Mass.: Harvard University Press, 1951); *idem*, *Methods of Logic* (London: Routledge & Kegan Paul, 1952); *idem*, *Theories and Things* (Cambridge, Mass.: Harvard University Press, 1981).

### *The myth of morality*

A consequence of this is that all such property names would refer to the same entity, so that “Round squareness is purplish yellowness” would be a necessary (and *a priori*) truth. This is such a counter-intuitive result that it must be classed as a theoretical cost.<sup>11</sup>

The above comments are not put forward with any rigor, and are intended primarily to undermine the claim that an error theory holds that the judgments of a discourse are *all false*. Putting aside the complex question of property names, the same point may be made employing only Strawson’s familiar (though by no means uncontested) views. We can conceive of a discourse revolving around a normal singular term, like “Elizabeth I,” and if we were to discover (bizarrely) that our Tudor history has been the subject of a monstrous hoax, and in fact the name “Elizabeth I” fails to denote anybody, then a Strawsonian would conclude that large tracts of our “Tudor discourse” are neither true nor false. (This might be what we choose to say about the failure of ancient Greek polytheistic discourse – with all those empty names like “Zeus,” “Aphrodite,” etc.) This conclusion would be properly classed as an *error theory*.

To some readers, this may seem like a surprising taxonomy. The view that our moral judgments are neither true nor false is often equated with the metaethical position known as “noncognitivism,” but the noncognitivist and error theoretic positions are distinct. However, I prefer to understand noncognitivism not in terms of truth values, but in terms of assertion. Assertion is not a semantic category; it is, rather, a purpose to which a sentence may be put: one and the same sentence may on some occasions be asserted, on other occasions not asserted. The question then is not whether “*a* is F” is an assertion, but whether it is typically used assertorically. The noncognitivist says “No”: the sentence “*a* is F” is typically used to express approval, or as a disguised command.

A moral cognitivist will, by contrast, hold that sentences of the form under discussion are usually used assertorically. But this is not to say that the cognitivist holds that moral sentences are usually either true or false, for (some have argued) there can be assertions that are neither. Strawsonian presupposition failure is one example. According to some views,

<sup>11</sup> An insistence that such terms refer to the null set might be accused of being a philosophically motivated attempt to provide a term with a referent at all costs (a “shadowy entity” as Quine called it in *Methods of Logic*, p. 198) – a strategy widely, though not universally, rejected for empty ordinary singular terms like “Zeus,” “the present king of France,” etc. For proponents of the *null individual*, see R. Carnap, *Meaning and Necessity* (Chicago: University of Chicago Press, 1956), pp. 36–8; R. M. Martin, “Of Time and the Null Individual,” *Journal of Philosophy* 62 (1965), pp. 723–36.

### *Error theory and motivation*

the assignment of certain vague predicates to “gray area” objects will also result in assertions that are neither true nor false. The difference is brought out by imagining a conversation in which one person utters “The present king of France is wise” and her companion responds “Say that again.” A Strawsonian would hold that neither utterance is true or false, but it would be an odd view that held that the former utterance is not asserted (and an odd view that held that the latter utterance *is* asserted). We might say that the former utterance was “in the market for truth,” whereas “Say that again,” being a command, is never in that market, and is therefore automatically neither true nor false.

An error theory, then, may be characterized as the position that holds that a discourse typically is used in an assertoric manner, but those assertions by and large fail to state truths. (These qualifications of vagueness should not cause concern; to expect more precision than this would be unrealistic.) This is clearly the correct stance to take towards phlogiston discourse. The view that seventeenth-century speakers typically spoke without assertoric force when they uttered sentences of the form “*a* is phlogiston” may be rejected. And such judgments were not true. (Presumably they were simply false, though we’ve left open the door for an argument to the conclusion that they were neither true nor false.) However, when it comes to our other model – “tapu discourse” – noncognitivism raises its seasoned head.

#### 1.2 NONCOGNITIVISM

A noncognitivist of the classic stripe might claim that when a Polynesian utters the sentence “ $\phi$  is tapu” she is doing nothing more than evincing her disapproval; she is really saying something equivalent to “ $\phi$ : boo!” Charles Stevenson claimed something more complex (about “morally bad” rather than “tapu”) – that the utterer is both asserting something about herself *and* issuing a command: “I disapprove of  $\phi$ ; do so as well!”<sup>12</sup> If either version is correct, the error theoretic stance dissolves: regardless of what kind of properties there are or are not in the world, the speaker is not reporting them – and *a fortiori* is not mistakenly reporting them. (I’m putting aside the self-describing element of Stevenson’s account, since one is hardly usually going to be in error regarding oneself.) If one employs a faulty theory – astrology, say – but withholds assertoric force from the propositions in

<sup>12</sup> C. L. Stevenson, “The Emotive Meaning of Ethical Terms,” *Mind* 46 (1937) pp. 14–31; *idem*, *Ethics and Language* (New Haven: Yale University Press, 1944).

*The myth of morality*

question – for example, if one says “As a Cancer, I’m inclined towards domestic pursuits and sentimentality” as part of telling a story, or a joke – then (despite the falsity of the *sentence*) one has not made a mistake. Should there be a discourse comprised of such utterances, the error theoretic stance would be inappropriate.

Noncognitivism is often naively presented in terms of “When people say X all they are really saying is Y.” This relation of “all they are really saying,” “all they really mean,” is quite puzzling. There are two ways of understanding the relation: as a semantic or as a pragmatic relation. Early noncognitivists, it would appear, read it as a semantic relation. When a person says “ $\phi$  is good” what the sentence *means* is “ $\phi$ : hurray!” (or whatever). In a much-quoted passage, A. J. Ayer claims that a moral judgment like “Stealing is wrong” lacks “factual meaning.” If I utter it, I “express no proposition which can be either true or false. It is as if I had written “Stealing money!!” – where the shape and thickness of the exclamation marks show, by a suitable convention, that a special sort of moral disapproval is the feeling which is being expressed. It is clear that there is nothing said which can be true or false.”<sup>13</sup> This is, on the face of it, an odd claim. Why would we clothe our emotive expressions in the form of sentences generally used to report facts, when we have at our disposal a perfectly good means for expressing them without going to the trouble? If all we’re saying is “Do  $\phi$ !” then why don’t we just *say* “Do  $\phi$ !”? The fact is, if someone participating in a serious moral discussion chose to express herself explicitly in the “uncooked” manner – imagine a member of a hospital ethics committee expressing her judgments as a series of “Hurray!”s and grunts of disapproval – we would be appalled. This is quite telling against the noncognitivist: it is implausible that two types of sentence could *mean the same* if we would treat discourse conducted in terms of one as sober and serious, and reject discourse conducted in terms of the other not merely as inappropriate, but as utterly mystifying. This kind of semantic noncognitivism, furthermore, is notoriously subject to a powerful criticism known as the “Frege–Geach problem.”<sup>14</sup> This objection states that utterances like “Hurray!” and “Do  $\phi$ !” do not behave *logically* like their supposed counterparts of the indicative mood. You cannot sensibly put “ $\phi$ : Boo!” into the antecedent slot of a conditional (whereas you can plug in “ $\phi$  is tapu”); nor could it appear as the minor premise of a valid piece of *modus ponens* reasoning

<sup>13</sup> A. J. Ayer, *Language, Truth and Logic* [1936] (Harmondsworth: Penguin, 1971), p. 110.

<sup>14</sup> See P. T. Geach, “Ascriptivism,” *Philosophical Review* 69 (1960), pp. 221–5; *idem*, “Assertion,” *Philosophical Review* 74 (1965), pp. 449–65.

(since validity is defined in terms of the *truth* of the premises guaranteeing the truth of the conclusion).

A noncognitivist fares better if he proposes the relation of “what they really mean” as a pragmatic one. This is how the theory was presented above, and is, apparently, how Stevenson understands things when he claims that the “major *use* [of ethical judgments] is not to indicate facts, but to create an influence.”<sup>15</sup> This need not be a claim about meaning, but a claim about how we employ our moral language (thus “what they really mean” is roughly equivalent to “what they really intend”). One advantage of this version is that the noncognitivist can at least point to areas of our non-moral vocabulary for which noncognitivism is a highly plausible option. A useful case to think about is that presented by John Austin.<sup>16</sup> According to Austin (and I’ve never found reason to doubt it), someone who utters the sentence “I name this ship *The Beagle*,” when in the orthodox circumstances involving cheering crowds, a bottle of champagne, etc., is not asserting anything, despite the indicative mood. She is not *describing*, or *reporting* the fact that she names the ship – the uttering of the sentence *is* the naming of the ship. Another example is that of an actor: someone playing the part of Hamlet on the stage would at some point utter “The air bites shrewdly,” but would not be asserting this fact. A third example is sarcasm: if one were to utter the sentence “That dinner party was fun” in a tone dripping with sarcasm, one would not thereby be asserting that the event was fun.

Now the noncognitivist might present her position along similar lines: although we frequently render our moral judgments in the indicative mood, we are (generally) not asserting them; rather, we are expressing emotions, issuing commands, etc. Such a noncognitivist could claim immunity from the Frege–Geach problem. If it were pointed out to Austin that the following is a valid instance of *modus ponens*, it would hardly cause concern for his theory of performatives:

1. I name this ship *The Beagle*
2. If I name this ship *The Beagle*, then I must have the authority to do so
3. Therefore, I must have the authority to do so

The fact that the sentence “I name this ship *The Beagle*” is usually, or even always, used in a non-assertoric manner does not mean that it cannot

<sup>15</sup> Stevenson, “Emotive Meaning,” p. 18 (italicization altered).

<sup>16</sup> J. L. Austin, *How to do Things with Words* (Oxford: Oxford University Press, 1962).

### *The myth of morality*

function perfectly legitimately in logically complex contexts; it does not follow that the sentence *means* anything other than what it appears to mean.

So the noncognitivist who claims that moral judgments (or tapu judgments) are not assertions can at least locate some partners in innocence. But despite this it is highly implausible that moral discourse is non-assertoric. Let us say that to assert that *p* is to express the belief that *p*.<sup>17</sup> It does not follow that the speaker need have the belief: a liar may express a belief that she doesn't have (a lie being a species of assertion). The pertinent relation of "expression" here denotes an expectation of what the utterance accomplishes, which is determined by a set of linguistic conventions. For example, if the utterance is preceded by "Once upon a time," then convention stipulates that the speaker may not believe what follows, and is not putting it forward for others to believe. (Similarly if it is uttered in an overtly sarcastic manner.) Linguistic conventions are not maintained through mind-reading – they are taught, learned, and communicated – and we should therefore be able to determine whether a fragment of language is assertoric.<sup>18</sup> What is required is an investigation of the ways moral language is used, in order to determine whether it bears any features indicative of the withdrawal of assertoric force.

Peter Glassen, in his 1959 paper, argued that whether an utterance is an assertion depends upon the intentions of the speaker.<sup>19</sup> I do not think this is quite correct, since a person may misunderstand the linguistic conventions to such an extent that despite a sincere intention to assert something, she fails to. (I once saw a comedy sketch in which an unfortunate person was doomed to utter everything in a sarcastic tone of voice.) Nevertheless, Glassen's way of arguing against noncognitivism is along the right lines. He asks "What would count as evidence of a person's intentions when he uses moral language?" – and he answers that since we are fallible with respect to reporting our own intentions, the best one could do is look at how a person does in fact use moral judgments. "We must observe the way he utters them, what else he says in relation to them, how others interpret them, and so on; in short, we must observe the characteristic

<sup>17</sup> The account of "assertoric force" appealed to throughout this book is intended to be that put forward by Austin, *How to do Things with Words*, and developed by J. R. Searle, *Speech Acts* (Cambridge: Cambridge University Press, 1969). Assertion is discussed in greater length in Chapter 7, below.

<sup>18</sup> By "linguistic conventions" here I do not mean to include the grammatical features of language, for which, of course, there is excellent evidence of their being innate.

<sup>19</sup> P. Glassen, "The Cognitivity of Moral Judgments," *Mind* 68 (1959), pp. 57–72.

features of moral discourse, and see how they compare with the characteristic features of discourse already known or, at any rate, believed to be cognitive” (pp. 61-2). Glassen’s point is that if all the evidence suggests that we intend to use our moral language in an assertoric manner, then all the evidence suggests that our moral language *is* assertoric, for assertion is entirely a matter of our intentions. The evidence that Glassen assembles I would employ to a slightly different end: as confirmation that the linguistic conventions that govern moral discourse are those of assertions. Here is Glassen’s list (which we can feel confident in assuming holds for historical Polynesian “tapu discourse,” too):

1. They (moral utterances) are expressed in the indicative mood
2. They can be transformed into interrogative sentences
3. They appear embedded in propositional attitude contexts
4. They are considered true or false, correct or mistaken
5. They are considered to have an impersonal, objective character
6. The putative moral predicates can be transformed into abstract singular terms (e.g., “goodness”), suggesting they are intended to pick out properties
7. They are subject to debate which bears all the hallmarks of factual disagreement

We can add to this list the two related characteristics highlighted by Peter Geach.<sup>20</sup>

8. They appear in logically complex contexts (e.g., as the antecedents of conditionals)
9. They appear as premises in arguments considered valid

The noncognitivist Rudolf Carnap wrote “Most philosophers have been deceived [by syntactic structure] into thinking that a value statement is really an assertive proposition . . . But actually a value statement is nothing else than a command in a misleading grammatical form . . . It does not assert anything.”<sup>21</sup> Given Glassen’s evidence, Carnap’s claim that *philosophers* have been misled into thinking moral utterances are assertoric is surely too weak – rather, it would seem that we are *all* misled. But it simply will not do for the noncognitivist to claim that we are all misled or mistaken in

<sup>20</sup> P. Geach “Assertion.” See also C. Wellman, “Emotivism and Ethical Objectivity,” *American Philosophical Quarterly* 5 (1968), pp. 90–9.

<sup>21</sup> R. Carnap, *Philosophy and Logical Syntax* (London: Kegan Paul, Trench, Trubner & Co. Ltd., 1935), pp. 24–5.



*The myth of morality*

participating in the above practices, for it is exactly our participation in these practices that provides the best evidence as to the truth of the matter.

There is one kind of *non-assertoric* discourse that bears many of the above features, and that is fictive discourse. This is so because fictive discourse by its very nature mimics ordinary discourse – it is the job of make-believe to copy. But fictive discourse is still, in a wider context, distinguishable from assertoric discourse, for we are disposed to “step out” of make-believe when pressed. If someone says in the appropriate tone of voice “But you don’t *really* believe in Sherlock Holmes, do you?”, the story-teller (despite having just uttered a series of indicative, logically complex sentences involving “Holmes”) answers “No, of course not.” We find no such widespread tendency concerning moral discourse.

In the absence of an explanation of *why* we would have a non-assertoric discourse bearing all the hallmarks of an assertoric one – in the absence of an explanation of why, if we already have perfectly good linguistic devices for expressing commands (imperatives) we should choose to cloak them systematically in indicative form – we must assume that if something walks and talks like a bunch of assertions, it’s highly likely that it *is* a bunch of assertions. Stevenson attempted to provide such an explanation when he claimed that moral language is largely a manipulative device: “When you tell a man that he oughtn’t to steal . . . [you are attempting] to get him to disapprove of it. Your ethical judgment has a quasi-imperative force which, operating through suggestion, and intensified by your tone of voice, readily permits you to begin to influence, to modify, his interests.”<sup>22</sup> I find Stevenson’s explanation, however, curiously self-undermining.

If I want James to  $\phi$ , and I am going to attempt to satisfy this desire using language, I have a choice of how to proceed. I might say to James “Do  $\phi$ !” Or I might say “I very much want you to  $\phi$ .” Or I might say “You must  $\phi$ ;  $\phi$ ing is obligatory.” Stevenson says that moral language (the third option) evolved because it is the most effective. Whence its extra rhetorical force? Presumably, the thought of an action being *required* carries more influence than a mere order or statement of desire. By merely barking an order to James, I can expect a request for a reason, and that reason may be only the statement that I have a desire to see  $\phi$ ing done. But just telling James that I want to see him  $\phi$  wears its “escapability” on its sleeve: if James has no interest in satisfying my desires, then he has been provided with no reason to  $\phi$ . If, however, I say that  $\phi$ ing is *morally required*, then it would

<sup>22</sup> Stevenson, “Emotive Meaning,” p. 19 (italicization altered).

seem that I have provided James with a reason: morality demands  $\phi$ ing, and James, like everyone else, is bound by the prescriptions of morality. If I can promulgate to my fellows that there is a set of rules which binds us all, then I have created a framework within which I can express my “will to power” most effectively – all I need do is persuasively claim of any action which I desire to see done that the set of rules demands it.

Suppose this Stevensonian picture, with its manipulative view of moral interaction laid bare, were correct. Then, when I say “ $\phi$ ing is obligatory” what is the function of my utterance? We can agree with Stevenson that the utterance is a tool by which I hope to see my desires satisfied. But how does the tool work? It’s a tool which makes reference to certain properties had by certain actions (even potential actions): “obligatoriness.” It is in virtue of making reference to these properties that the utterance has more rhetorical influence than “I want you to  $\phi$ ” or “I approve of  $\phi$ ing; do so as well.” If I am clear-headed about my manipulative behavior, then I do not really believe that there is any such property – I am making reference to it merely in the hope that my audience is gullible. In such a case I am lying; I am trying to get my audience to form beliefs that I don’t have. Chances are, few of us are so scheming in our expressions of will to power; rather, we have been subject to the manipulative behaviors of others, and thus have bought into the whole “must-be-doneness” framework. So when I really want to see  $\phi$ ing done, I may well *believe* that it instantiates this property; thus my utterance is no lie, it’s an expression of a belief. But either way – whether we are clear-headed about our manipulative ways or not – moral utterances turn out to be *assertions*. The fact that I say something in order to satisfy a desire to see James  $\phi$  does not make that utterance a command, any more than my saying to a student “Descartes was French” is a command in virtue of the fact that ultimately I hope to influence the student to write true things rather than false things in the exam.

In light of these problems for the noncognitivist, I will proceed under the natural assumption that our moral language is used largely in an assertoric manner. Noncognitivism is implausible as a description of our own moral language, just as it is implausible as an account of serious judgments of the form “ $\phi$  is tapu.” Of course, there is much moral language that is clearly not assertoric (“Don’t do that!” etc.), but it bears a vital relation to the assertoric portion: if one were not willing to assert “ $\phi$ ing is morally forbidden” one would not be willing to press the moral injunction “Don’t  $\phi$ !” Were the assertoric language shown to be hopelessly flawed – based on a mistake

### *The myth of morality*

about the nature of agency, or the nature of the world – then the imperatival portion of the language would not remain unscathed.<sup>23</sup>

Now we know what an error theory in general looks like, we can turn to the case that interests us: the possibility that it is the appropriate stance to take towards our own familiar moral discourse. In this section I have argued that moral discourse is assertoric; it remains to be argued that these assertions are *untrue*. The argument is most usefully approached via a discussion of Mackie’s original version.

#### 1.3 MACKIE’S ERROR THEORY

John Mackie’s argument for a moral error theory embodies two steps. First he attempts to establish a conceptual relation – that is, he looks for what a moral use of the predicate “. . . is good” *means*. He then embarks on the substantive step, showing that the meaning in question is not satisfied by the world. Concerning the former, Mackie thinks that all uses of the word “good” boil down to “*such as to satisfy the requirements of the kind in question.*” The “requirements in question” could involve the use to which we put cars (allowing us to speak of “a good car”), the end of winning a game of chess (“a good move”), or the fulfilling of a social role (“a good quarterback”), etc. These are all, in one way or another, requirements which *we* impose upon the world. When we use “good” with moral strength, however, we advert to requirements which are “just there” – in the nature of things. All non-moral uses of “good” involve requirements for which there is, roughly speaking, a “requirer”; but when we up-the-ante to a moral “good,” we are implicitly referring to requirements for which there is no requirer – laws for which there is no law-maker. Non-moral uses of “good” are what we might call “subjectively prescriptive” (they are prescriptive ultimately in virtue of our desires, intentions, beliefs, etc.), but moral uses of “good” are “objectively prescriptive.” That’s the conceptual step. The substantive step of Mackie’s argument is to argue that there *are no* “objective prescriptions”: the universe, without our impositions upon it, simply does not make requirements. Thus judgments of the form “ $\phi$  is morally good” are never true (when  $\phi$  takes an actual value).

One may attempt to block the two-step operation at either stage. One class of critics agrees that “objective prescriptions” are completely bizarre, but they deny that our moral discourse ever commits us to anything so

<sup>23</sup> I discuss noncognitivism in more detail in “Noncognitivism, Motivation, and Assertion” (forthcoming).

strange. Another class is inclined to agree that our moral discourse does embody a commitment to objective prescriptions, but denies that they are particularly odd – properly understood, the universe does make requirements of us. Strategically, it is available to the error theorist to play off the two types of critic against each other. That critic who finds “objective prescriptivity” sustainable will generally also want to argue that we *are* committed to it (for it would be an unusual view that held that objective prescriptions are defensible but the truth of nothing we say requires their defense). From this critic’s latter arguments the error theorist may draw adventitious support against her other type of opponent.

Sketched in the above terms, Mackie’s notion of “objective prescriptivity” is too blunt for a proper argument to be conducted; in what follows I shall attempt to nail down what it might mean in more precise terms. Whether my claims ultimately are convincing as illuminations of Mackie is not important. What matters is that I utilize the same form of argument: first to find some thesis T to which our ordinary moral discourse is committed (a “non-negotiable” element), then to argue that T is false. The latter step promises to be the more straightforward: the annals of philosophy are strewn with arguments exposing faulty theses. This is not to say it is easy – but at least we know the nature of the sport. But to make a case that a discourse is “committed to some thesis” is an altogether more elusive game. In §1.0 I suggested a way of conceptualizing the issue – in terms of when we would or wouldn’t accept a translation – but this was not intended as a decision procedure.

I will examine two broad interpretations of “objective prescriptivity.” The first occupies the remainder of this chapter. The reader ought to be warned that I do not take this argument to be very convincing, and so the chapter ends on rather an unsatisfactory note. The intention is, first, to pursue an argument which is interesting even if not altogether persuasive, and, second, to gain insight into how an argument for a moral error theory might operate. Perhaps it is best if the rest of this chapter is seen as a warm-up exercise for a much stronger argument – the second interpretation of “objective prescriptivity” – which will occupy later chapters.

#### 1.4 INTERNALISM ABOUT MOTIVATION

There is a thesis which I will call “internalism about motivation” which has been thought (i) to be a non-negotiable commitment of moral discourse, and (ii) to be false.

## *The myth of morality*

MI: It is necessary and *a priori* that any agent who judges that one of his available actions is morally obligatory will have some (defeasible) motivation to perform that action.

Advocates of (i) tend to reject (ii), and proponents of (ii) tend to reject (i), and thus most followers of either (i) or (ii) avoid an error theory. David Brink appears to interpret Mackie's error theory as consisting of the endorsement of (i) and (ii).<sup>24</sup> Understandably so: reading *Ethics: Inventing Right and Wrong*, it certainly seems as if the flaws of moral discourse have *something* to do with motivation. "Objective prescriptivity" is compared with Platonism, whereby knowing that something is good "will not merely tell men what to do but will ensure that they do it" (*Ethics*, p. 23), and will provide the knower "with both a direction and an overriding motive" (p. 40). The argument that Brink finds in Mackie presumably goes like this:

"Mackie's Argument":

1. MI is false
2. Morality is committed to MI
- ∴ 3. Morality is flawed (i.e., a moral error theory)

Brink goes on to argue that MI is false. But in the context of reading Mackie as above, while trying to resist the conclusion, this seems an odd strategic move. Perhaps Brink is arguing as follows:

"Brink's Argument":

1. MI is false
- ∴ 2. Morality was never committed to MI in the first place
- ∴ 3. "Mackie's Argument" is unsound

However, the move from premise (1) to (2) in "Brink's Argument" is simply question-begging against the moral error theorist.

At any rate, if MI is so fantastic, it is curious that it finds so many staunch defenders. Hume wrote: "Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular."<sup>25</sup> In a similar vein, W. Frankena writes that "it would seem paradoxical if one were to say 'X is good' or 'Y is right' but be absolutely indifferent to its being

<sup>24</sup> D. Brink, "Moral Realism and the Skeptical Arguments from Disagreement and Queerness," *Australasian Journal of Philosophy* 62 (1984), pp. 111–25; *idem*, "Externalist Moral Realism," *Southern Journal of Philosophy*, supplementary volume 24 (1986), pp. 23–41.

<sup>25</sup> D. Hume, *A Treatise of Human Nature* [1739], book 3, part 1, section 1, ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1978), p. 457. The noncognitivist conclusion that Hume immediately draws is that the "rules of morality, therefore, are not conclusions of our reason."

done or sought by himself or anyone else. If he were indifferent in this way, we would take him to mean that it is generally regarded as good or right, but that he did not so regard it himself.”<sup>26</sup> The latter quote, focusing on the possibility or impossibility of a type of agent – one who makes sincere moral judgments but is left motivationally inert – draws the battle lines for debating MI: if such an agent is possible, then MI is false; if not, then it is true. I agree with Brink that MI is false, and will present my reasons in a moment. I must admit, however, to being somewhat half-hearted about the task, since I have doubts about the second step of the argument: that our moral discourse is committed to MI.

Debate has focused on the *amoral* agent: a stipulated form of amorality that consists precisely of making sincere moral judgments but having no motivation. We are invited to imagine a thoroughly depressed person – utterly unmoved yet still making moral judgments.<sup>27</sup> The motivation internalist will try to deny the case: perhaps the person is not really making a moral judgment, but rather saying something “in quote marks.” “Although I know that fulfilling my promises is correct, I just feel unmoved” becomes “Although I know that most people think that my fulfilling my promises is correct, I just feel unmoved.” The latter claim is not a moral judgment; it is a non-moral judgment about what other people’s moral judgments are. This is a useful rejoinder for the motivation internalist. It is a response that can be used again and again because we are considering an agent who *ex hypothesi* is motivationally inert, and therefore our only grounds for holding that she makes a moral judgment is that she *says so*; but given people’s notorious unreliability at reporting their own states, the evidence for the occurrence of a sincere moral judgment is always going to be vulnerable to reinterpretation. It would be better if we could locate an agent for whom there is some feature that is *explained* by her having made a sincere moral judgment. The best contender for the role of counter-example to MI, therefore, is not the amoral agent but the thoroughly *evil* agent – the agent whose moral judgments do not leave him motivationally cold, but which provide the reason for his diabolical actions. (This is not to deny that depressed agents may well be counter-examples to MI, it is just that it is difficult to establish the fact.<sup>28</sup>)

<sup>26</sup> W. K. Frankena, *Ethics* (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1973), p. 100.

<sup>27</sup> See Brink, “Moral Realism”; Smith, *Moral Problem*; R. D. Milo, “Moral Indifference,” *The Monist* 64 (1981), pp. 373–93.

<sup>28</sup> See M. Stocker, “Desiring the Bad: An Essay in Moral Psychology,” *Journal of Philosophy* 76 (1979), pp. 738–53.

## The myth of morality

### 1.5 PURE EVIL

Few cases from history and literature of what we would usually call “evil” will satisfy this criterion. One of the internalist’s standard rejoinders – that the agent is akratic, or acts badly so as to attain some end (and therefore is not doing evil *for evil’s sake*), or has just rejected the whole realm of morality – will probably be highly plausible and not *ad hoc*. Occasionally, though, we run into an evil character who is really interested in morality itself, and with whom we must credit genuine moral judgments in order to explain his or her behavior. Some of the villains from the Marquis de Sade’s work, for example, are not just interested in hedonism and sadism – they appear to be self-consciously pursuing whatever they consider to be *bad*.<sup>29</sup> If they judged excessive, sadistic hedonism to be morally acceptable, then we naturally think of them as ceasing to pursue it. Shakespeare’s despicable Aaron, from *Titus Andronicus*, goes to his death with the words “If one good deed in all my life I did, I do repent it from my very soul.” Or consider the following from Edgar Allan Poe:

Who has not, a hundred times, found himself committing a vile or a silly action, for no other reason than because he knows he should *not*? Have we not a perpetual inclination, in the teeth of our best judgment, to violate that which is *Law*, merely because we understand it to be such? This spirit of perverseness, I say, came to my overthrow. It was this unfathomable longing of the soul *to vex itself* – to offer violence to its own nature – to do wrong for the wrong’s sake only – that urged me to continue and finally to consummate the injury I had inflicted upon the unoffending brute.<sup>30</sup>

The upholder of MI will have to deny that Poe’s character, taken at face value, is possible. But this denial, if pursued, becomes implausible. Before proceeding though, let me make a brief aside to head off some potential misunderstanding. The “big picture” that we are considering is that moral discourse is committed to MI, and MI is false (hence, an error theory). However, we are now attacking MI, and our means for doing so is to

<sup>29</sup> Marquis de Sade, *The Complete Justine, Philosophy in the Bedroom, and other Writings*, trans. and ed. R. Seaver and A. Wainhouse (New York: Grove Press, 1965); *The Misfortunes of Virtue, and other Early Tales*, trans. D. Coward (Oxford: Oxford University Press, 1992). My discussion of de Sade is greatly indebted to G. Rosen’s “Internalism and Common Sense: A Philosophical Lesson from *Philosophie dans le boudoir*” (unpublished manuscript, 1989). Though Rosen draws conclusions from the philosophical lesson very different from my own, the major arguments that I employ here, and the example of Eugenie, are his.

<sup>30</sup> Edgar Allan Poe, “The Black Cat” [1843], *Selected Writings* (Harmondsworth: Penguin, 1985), p. 322.