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 uncleanliness 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.[1]

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 "ϕ is tapu" is ever true (so long as "ϕ" names an actual action[2]) because there simply isn't anything that's tapu. Saying this implies nothing about how tolerant in

[1] 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.

[2] Throughout this book, the symbols "ϕ" and "ψ" generally stand for actions. However, sometimes they stand in for verbs ("ϕing is good," "I want you to ϕ") and sometimes they do the work of nouns ("The action ϕ"). I find this convenient and not noticeably jarring.

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 "ϕ is tapu," but only when this is an anthropological judgment, elliptical for "For the islanders, ϕ is considered tapu." It would be strange for him to make non-elliptical judgments of the form "ϕ 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 uncleanliness" 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.[3]

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 ostensively 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

[3] 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.[4] 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*

[4] 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).

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").[5] 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."[6] 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.

[5] 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.

[6] 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})."

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.[7] 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.[8] 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.[9] 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

[7] 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.

[8] P. F. Strawson, "On Referring," in A. Flew (ed.), *Essays in Conceptual Analysis* (London: Macmillan & Co. Ltd., 1956), pp. 21–52.

[9] 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.[10] 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.

[10] 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).

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.[11]

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,

[11] 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.

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 "ϕ is tapu" she is doing nothing more than evincing her disapproval; she is really saying something equivalent to "ϕ: 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 ϕ; do so as well!"[12] 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

[12] 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).

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 "ϕ is good" what the sentence *means* is "ϕ: 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."[13] 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 ϕ!" then why don't we just *say* "Do ϕ!"? 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."[14] This objection states that utterances like "Hurray!" and "Do ϕ!" do not behave *logically* like their supposed counterparts of the indicative mood. You cannot sensibly put "ϕ: Boo!" into the antecedent slot of a conditional (whereas you can plug in "ϕ is tapu"); nor could it appear as the minor premise of a valid piece of *modus ponens* reasoning

[13] A. J. Ayer, *Language, Truth and Logic* [1936] (Harmondsworth: Penguin, 1971), p. 110.
[14] See P. T. Geach, "Ascriptivism," *Philosophical Review* 69 (1960), pp. 221–5; *idem*, "Assertion," *Philosophical Review* 74 (1965), pp. 449–65.