

PHILOSOPHY: THE BIG QUESTIONS



# ETHICS:

## *The Big Questions*

EDITED BY JAMES P. STERBA

 Blackwell  
Publishing

---

## 4 How to Derive “Ought” from “Is”

---

*John R. Searle*

I

It is often said that one cannot derive an ‘ought’ from an ‘is’. This thesis, which comes from a famous passage in Hume’s *Treatise*, while not as clear as it might be, is at least clear in broad outline: there is a class of statements of fact which is logically distinct from a class of statements of value. No set of statements of fact by themselves entails any statement of value. Put in more contemporary terminology, no set of *descriptive* statements can entail an *evaluative* statement without the addition of at least one evaluative premise. To believe otherwise is to commit what has been called the naturalistic fallacy.

From *Philosophical Review*, vol. 73 (1964), pp. 43–58.  
Reprinted with permission.

I shall attempt to demonstrate a counter-example to this thesis.<sup>1</sup> It is not of course to be supposed that a single counter-example can refute a philosophical thesis, but in the present instance if we can present a plausible counter-example and can in addition give some account or explanation of how and why it is a counter-example, and if we can further offer a theory to back up our counter-example – a theory which will generate an indefinite number of counter-examples – we may at the very least cast considerable light on the original thesis; and possibly, if we can do all these things, we may even incline ourselves to the view that the scope of that thesis was more restricted than we had originally supposed. A counter-example must proceed by taking a statement or statements which any proponent of the thesis would grant were purely factual or ‘descriptive’ (they need not ac-

tually contain the word 'is') and show how they are logically related to a statement which a proponent of the thesis would regard as clearly 'evaluative'. (In the present instance it will contain an 'ought'.)<sup>2</sup>

Consider the following series of statements:

1. Jones uttered the words 'I hereby promise to pay you, Smith, five dollars.'
2. Jones promised to pay Smith five dollars.
3. Jones placed himself under (undertook) an obligation to pay Smith five dollars.
4. Jones is under an obligation to pay Smith five dollars.
5. Jones ought to pay Smith five dollars.

I shall argue concerning this list that the relation between any statement and its successor, while not in every case one of 'entailment', is none the less not just a contingent relation; and the additional statements necessary to make the relationship one of entailment do not need to involve any evaluative statements, moral principles, or anything of the sort.

Let us begin. How is (1) related to (2)? In certain circumstances, uttering the words in quotation marks in (1) is the act of making a promise. And it is a part of or a consequence of the meaning of the words in (1) that in those circumstances uttering them is promising. 'I hereby promise' is a paradigm device in English for performing the act described in (2), promising.

Let us state this fact about English usage in the form of an extra premise:

- (1a) Under certain conditions *C* anyone who utters the words (sentence) 'I hereby promise to pay you, Smith, five dollars' promises to pay Smith five dollars.

What sorts of things are involved under the rubric 'conditions *C*'? What is involved will be all those conditions, those states of affairs, which are necessary and sufficient conditions for the utterance of the words (sentence) to constitute the successful performance of the act of promising. The conditions will include such things as that the speaker is in the presence of the hearer Smith,

they are both conscious, both speakers of English, speaking seriously. The speaker knows what he is doing, is not under the influence of drugs, not hypnotized or acting in a play, not telling a joke or reporting an event, and so forth. This list will no doubt be somewhat indefinite because the boundaries of the concept of a promise, like the boundaries of most concepts in a natural language, are a bit loose.<sup>3</sup> But one thing is clear; however loose the boundaries may be, and however difficult it may be to decide marginal cases, the conditions under which a man who utters 'I hereby promise' can correctly be said to have made a promise are straightforwardly empirical conditions.

So let us add as an extra premise the empirical assumption that these conditions obtain.

- (1b) Conditions *C* obtain.

From (1), (1a) and (1b) we derive (2). The argument is of the form: If *C* then (if *U* then *P*): *C* for conditions, *U* for utterance, *P* for promise. Adding the premises *U* and *C* to this hypothetical we derive (2). And as far as I can see, no moral premises are lurking in the logical woodpile. More needs to be said about the relation of (1) to (2), but I reserve that for later.

What is the relation between (2) and (3)? I take it that promising is, by definition, an act of placing oneself under an obligation. No analysis of the concept of promising will be complete which does not include the feature of the promiser placing himself under or undertaking or accepting or recognizing an obligation to the promisee, to perform some future course of action, normally for the benefit of the promisee. One may be tempted to think that promising can be analysed in terms of creating expectations in one's hearers, or some such, but a little reflection will show that the crucial distinction between statements of intention on the one hand and promises on the other lies in the nature and degree of commitment or obligation undertaken in promising.

I am therefore inclined to say that (2) entails (3) straight off, but I can have no objection if anyone wishes to add – for the purpose of formal neatness – the tautological premise:

- (2a) All promises are acts of placing oneself under (undertaking) an obligation to do the thing promised.

How is (3) related to (4)? If one has placed oneself under an obligation, then, other things being equal, one is under an obligation. That I take it also is a tautology. Of course it is possible for all sorts of things to happen which will release one from obligations one has undertaken and hence the need for the *ceteris paribus* rider. To get an entailment between (3) and (4) we therefore need a qualifying statement to the effect that:

- (3a) Other things are equal.

Formalists, as in the move from (2) to (3), may wish to add the tautological premise:

- (3b) All those who place themselves under an obligation are, other things being equal, under an obligation.

The move from (3) to (4) is thus of the same form as the move from (1) to (2): If *E* then (if *PUO* then *UO*): *E* for other things are equal, *PUO* for place under obligation and *UO* for under obligation. Adding the two premises *E* and *PUO* we derive *UO*.

Is (3a), the *ceteris paribus* clause, a concealed evaluative premise? It certainly looks as if it might be, especially in the formulation I have given it, but I think we can show that, though questions about whether other things are equal frequently involve evaluative considerations, it is not logically necessary that they should in every case. I shall postpone discussion of this until after the next step.

What is the relation between (4) and (5)? Analogous to the tautology which explicates the relation of (3) and (4) there is here the tautology that, other things being equal, one ought to do what one is under an obligation to do. And here, just as in the previous case, we need some premise of the form:

- (4a) Other things are equal.

We need the *ceteris paribus* clause to eliminate the possibility that something extraneous to the relation of 'obligation' to 'ought' might interfere.<sup>4</sup> Here, as in the previous two steps, we eliminate the appearance of enthymeme by pointing out that the apparently suppressed premise is tautological and hence, though formally neat, it is redundant. If, however, we wish to state it formally, this argument is of the same form as the move from (3) to (4): If *E* then (if *UO* then *O*); *E* for other things are equal, *UO* for under obligation, *O* for ought. Adding the premises *E* and *UO* we derive *O*.

Now a word about the phrase 'other things being equal' and how it functions in my attempted derivation. This topic and the closely related topic of defeasibility are extremely difficult and I shall not try to do more than justify my claim that the satisfaction of the condition does not necessarily involve anything evaluative. The force of the expression 'other things being equal' in the present instance is roughly this. Unless we have some reason (that is, unless we are actually prepared to give some reason) for supposing the obligation is void (step 4) or the agent ought not to keep the promise (step 5), then the obligation holds and he ought to keep the promise. It is not part of the force of the phrase 'other things being equal' that in order to satisfy it we need to establish a universal negative proposition to the effect that no reason could ever be given by anyone for supposing the agent is not under an obligation or ought not to keep the promise. That would be impossible and would render the phrase useless. It is sufficient to satisfy the condition that no reason to the contrary can in fact be given.

If a reason is given for supposing the obligation is void or that the promiser ought not to keep the promise, then characteristically a situation calling for evaluation arises. Suppose, for example, we consider a promised act wrong, but we grant that the promiser did undertake an obligation. Ought he to keep the promise? There is no established procedure for objectively deciding such cases in advance, and an evaluation (if that is really the right word) is in order. But unless we have some reason to the contrary, the

*ceteris paribus* condition is satisfied, no evaluation is necessary, and the question whether he ought to do it is settled by saying 'he promised.' It is always an open possibility that we may have to make an evaluation in order to derive 'he ought' from 'he promised', for we may have to evaluate a counter-argument. But an evaluation is not logically necessary in every case, for there may as a matter of fact be no counter-arguments. I am therefore inclined to think that there is nothing necessarily evaluative about the *ceteris paribus* condition, even though deciding whether it is satisfied will frequently involve evaluations.

But suppose I am wrong about this: would that salvage the belief in an unbridgeable logical gulf between 'is' and 'ought'? I think not, for we can always rewrite my steps (4) and (5) so that they include the *ceteris paribus* clause as part of the conclusion. Thus from our premises we would then have derived 'Other things being equal Jones ought to pay Smith five dollars', and that would still be sufficient to refute the tradition, for we would still have shown a relation of entailment between descriptive and evaluative statements. It was not the fact that extenuating circumstances can void obligations that drove philosophers to the naturalistic fallacy; it was rather a theory of language, as we shall see later on.

We have thus derived (in as strict a sense of 'derive' as natural languages admit of) an 'ought' from an 'is'. And the extra premises which were needed to make the derivation work were in no case moral or evaluative in nature. They consisted of empirical assumptions, tautologies and descriptions of word usage. It must be pointed out also that the 'ought' is a 'categorical' not a 'hypothetical' ought. (5) does not say that Jones ought to pay up if he wants such and such. It says he ought to pay up, period. Note also that the steps of the derivation are carried on in the third person. We are not concluding 'I ought' from 'I said "I promise"', but 'he ought' from 'he said "I promise"'.

The proof unfolds the connection between the utterance of certain words and the speech act of promising and then in turn unfolds promising into obligation and moves from obligation to

'ought'. The step from (1) to (2) is radically different from the others and requires special comment. In (1) we construe 'I hereby promise . . .' as an English phrase having a certain meaning. It is a consequence of that meaning that the utterance of that phrase under certain conditions is the act of promising. Thus by presenting the quoted expressions in (1) and by describing their use in (1a) we have as it were already invoked the institution of promising. We might have started with an even more ground-floor premise than (1) by saying:

- (1b) Jones uttered the phonetic sequence:  
ai'hirbai'pramis'tapei'yu'smiθ'faiν'dal  
ərz/

We would then have needed extra empirical premises stating that this phonetic sequence was associated in certain ways with certain meaningful units relative to certain dialects.

The moves from (2) to (5) are relatively easy. We rely on definitional connections between 'promise', 'obligate', and 'ought', and the only problem which arises is that obligations can be overridden or removed in a variety of ways and we need to take account of that fact. We solve our difficulty by adding further premises to the effect that there are no contrary considerations, that other things are equal.

## II

In this section I intend to discuss three possible objections to the derivation.

### *First objection*

Since the first premise is descriptive and the conclusion evaluative, there must be a concealed evaluative premise in the description of the conditions in (2b).

So far, this argument merely begs the question by assuming the logical gulf between descriptive and evaluative which the derivation is designed to challenge. To make the objection stick, the defender of the distinction would have



to show how exactly (2b) must contain an evaluative premise and what sort of premise it might be. Uttering certain words in certain conditions just *is* promising and the description of these conditions needs no evaluative element. The essential thing is that in the transition from (1) to (2) we move from the specification of a certain utterance of words to the specification of a certain speech act. The move is achieved because the speech act is a conventional act; and the utterance of words, according to the conventions, constitutes the performance of just that speech act.

A variant of this first objection is to say: all you have shown is that 'promise' is an evaluative, not a descriptive, concept. But this objection again begs the question and in the end will prove disastrous to the original distinction between descriptive and evaluative. For that a man uttered certain words and that these words have the meaning they do are surely objective facts. And if the statement of these two objective facts plus a description of the conditions of the utterance is sufficient to entail the statement (2) which the objector alleges to be an evaluative statement (Jones promised to pay Smith five dollars), then an evaluative conclusion is derived from descriptive premises without even going through steps (3), (4) and (5).

### *Second objection*

Ultimately the derivation rests on the principle that one ought to keep one's promises and that is a moral principle, hence evaluative.

I don't know whether 'one ought to keep one's promises' is a 'moral' principle, but whether or not it is, it is also tautological; for it is nothing more than a derivation from the two tautologies:

All promises are (create, are undertakings of, are acceptances of) obligations,

and

One ought to keep (fulfil) one's obligations.

What needs to be explained is why so many philosophers have failed to see the tautological character of this principle. Three things I think have concealed its character from them.

The first is a failure to distinguish external questions about the institution of promising from internal questions asked within the framework of an institution. The questions 'Why do we have such an institution as promising?' and 'Ought we to have such institutionalized forms of obligation as promising?' are external questions asked about and not within the institution of promising. And the question 'Ought one to keep one's promises?' can be confused with or can be taken as (and I think has often been taken as) an external question roughly expressible as 'Ought one to accept the institution of promising?' But taken literally, as an internal question, as a question about promises and not about the institution of promising, the question 'Ought one to keep one's promises?' is as empty as the question 'Are triangles three-sided?' To recognize something as a promise is to grant that, other things being equal, it ought to be kept.

A second fact which has clouded the issue is this. There are many situations, both real and imaginable, where one ought not to keep a promise, where the obligation to keep a promise is overridden by some further considerations, and it was for this reason that we needed those clumsy *ceteris paribus* clauses in our derivation. But the fact that obligations can be overridden does not show that there were no obligations in the first place. On the contrary. And these original obligations are all that is needed to make the proof work.

Yet a third factor is the following. Many philosophers still fail to realize the full force of saying that 'I hereby promise' is a performative expression. In uttering it one performs but does not describe the act of promising. Once promising is seen as a speech act of a kind different from describing, then it is easier to see that one of the features of the act is the undertaking of an obligation. But if one thinks the utterance of 'I promise' or 'I hereby promise' is a peculiar kind of description – for example, of one's mental state – then the relation between promising and obligation is going to seem very mysterious.

*Third objection*

The derivation uses only a factual or inverted-commas sense of the evaluative terms employed. For example, an anthropologist observing the behaviour and attitudes of the Anglo-Saxons might well go through these derivations, but nothing evaluative would be included. Thus step (2) is equivalent to 'He did what they call promising' and step (5) to 'According to them he ought to pay Smith five dollars.' But since all of the steps (2) to (5) are in *oratio obliqua*, and hence disguised statements of fact, the fact-value distinction remains unaffected.

This objection fails to damage the derivation, for what it says is only that the steps *can* be reconstrued as in *oratio obliqua*, that we can construe them as a series of external statements, that we can construct a parallel (or at any rate related) proof about reported speech. But what I am arguing is that, taken quite literally, without any *oratio obliqua* additions or interpretations, the derivation is valid. That one can construct a similar argument which would fail to refute the fact-value distinction does not show that this proof fails to refute it. Indeed it is irrelevant.

**Notes**

Earlier versions of this paper were read before the Stanford Philosophy Colloquium and the Pacific Di-

vision of the American Philosophical Association. I am indebted to many people for helpful comments and criticisms, especially Hans Herzberger, Arnold Kaufmann, Benson Mates, A. I. Melden and Dagmar Searle.

- 1 In its modern version. I shall not be concerned with Hume's treatment of the problem.
- 2 If this enterprise succeeds, we shall have bridged the gap between 'evaluative' and 'descriptive' and consequently have demonstrated a weakness in this very terminology. At present, however, my strategy is to play along with the terminology, pretending that the notions of evaluative and descriptive are fairly clear. At the end of the paper I shall state in what respects I think they embody a muddle.
- 3 In addition the concept of a promise is a member of a class of concepts which suffer from looseness of a peculiar kind, viz. defeasibility. Cf. H. L. A. Hart, 'The Ascription of Responsibility and Rights', *Logic and Language*, first series, ed. A. Flew (Oxford, 1951).
- 4 The *ceteris paribus* clause in this step excludes somewhat different sorts of cases from those excluded in the previous step. In general we say, 'He undertook an obligation, but none the less he is not (now) under an obligation when the obligation has been *removed*, e.g. if the promisee says, 'I release you from your obligation.' But we say, 'He is under an obligation, but none the less ought not to fulfil it' in cases where the obligation is *overridden* by some other consideration, e.g. a prior obligation.