

ETHICAL NATURALISM REVISITED*

LAWRENCE FOSTER
University of Massachusetts

Criticisms of ethical naturalism are numerous and varied. Yet, despite their number, they pay surprisingly little attention to the question: What are the theses of ethical naturalism? And despite their variety, many share the following important characteristics: either they attack one version of ethical naturalism and believe they have thereby refuted ethical naturalism, or they implicitly and correctly distinguish among several different versions of ethical naturalism but, once again, in allegedly showing that one version is mistaken, they infer that the other versions are also thereby refuted. Moore, I believe, takes the former approach; Hare, the latter.¹

What I wish to do in this paper is first to formulate three different and distinct versions of ethical naturalism. Secondly, I will illustrate the confused second line of attack by considering some aspects of Hare's criticism of ethical naturalism. A precise delineation of the three different versions of ethical naturalism clearly shows, *contra* Hare and other naturalistic critics, that certain criticisms, even if accepted as refutations of one version of ethical naturalism, are, in fact, irrelevant to the other versions; and that the falsity of one version of ethical naturalism does not entail the falsity of the other versions. Thirdly, I will examine some of Hare's criticisms of the three types of naturalistic theories and, in so doing, I will elicit some common yet mistaken assumptions about

* I am indebted to Lynn V. Foster and Herbert Heidelberger for their helpful comments on an earlier version of this paper.

¹ G. E. Moore, *Principia Ethica* (Cambridge, 1903), Ch. 1, and R. M. Hare, *The Language of Morals* (Oxford, 1952), Ch. 5. Hereafter *LM*.

criteria for entailments, definitions, and explications involved in such attacks.

Hare has been chosen for special consideration since his arguments, I believe, exemplify mistakes common among critics of ethical naturalism. Examination of his arguments helps to clarify debates concerning ethical naturalism and helps to show just what is, or is not, proved by a given line of criticism.

I

Although not usually distinguished, there are at least three important yet distinct theses which can accurately be construed as naturalistic ethical theses. (a) *The Definability Thesis*. This is the version of ethical naturalism most frequently criticized. This thesis holds that all ethical predicates are ultimately definable solely by non-ethical naturalistic predicates.² It is far from clear just which moral philosophers have held such a view, but nevertheless it is the position criticized by Moore and Hare among others. (b) *The Deducibility Thesis*. Loosely stated, this thesis asserts that some ethical statements (predicates) can be deduced from a set of statements (predicates) none of which is ethical, and all other ethical statements can be deduced from the previously entailed statements along with other nonethical statements.³ The deducibility thesis is the supposed object of Hume's attack in the notorious 'is-ought' passage in the *Treatise*. As we shall see, it is sometimes held that the falsity of the deducibility thesis follows from the falsity of the definability thesis. (c) *The Explication Thesis*. The contention here is that all ethical predicates can be explicated ultimately by non-ethical

² It is important to note that an ethical naturalist only has to hold that some *one* ethical predicate is *directly* definable by naturalistic predicates alone. He may then maintain that all other ethical predicates are indirectly definable naturalistically providing that they are definable by means of the directly definable ethical predicate plus other naturalistic predicates.

³ Following Hare, I will speak of statements or judgments as well as predicates or characteristics as entailing or entailed by other statements or judgments or predicates or characteristics.

naturalistic predicates alone. For some reason this thesis is generally either overlooked or not taken seriously by critics of ethical naturalism. This may be due to an unfamiliarity with explications or, as we shall see apparently in the case of Hare, to a serious misconception about the nature of explication.

If theses (a), (b), and (c) do represent legitimate alternative versions of ethical naturalism then, obviously, any criticism which purports to show that ethical naturalism is mistaken must be criticism of all three. Although Hare recognizes at least implicitly that there are these three distinct versions of ethical naturalism, he seriously misconstrues the relationship between (a) and (b)—apparently believing that the denial of (a) entails the denial of (b). Further, in rejecting theses (a) and (c), he accepts views concerning definitions and explications which are mistaken. We now turn to Hare's arguments in order to illustrate these points.

II

Hare speaks alternatively of ethical naturalism as the theory which construes the predicate 'good' as *logically dependent* on some group of characteristics of things (81), or construes 'good' as *entailed* by these characteristics (81), or seeks to make value judgments *derivable* from statements of fact (82), or construes 'good' as *having the same meaning* as some naturalistic group of predicates (82-83, 85), or construes 'good' as having such a group of characteristics as its *defining characteristics* (83, 85, 86).⁴ Obviously, not all of these versions of ethical naturalism are equivalent. Some represent thesis (a) and others, thesis (b). Thesis (c) appears later.

Although Hare's initial arguments against ethical naturalism are not always clearly directed against one version of

⁴ L. W. Sumner has noted these different interpretations in his "Hare's Arguments against Ethical Naturalism", *The Journal of Philosophy*, LXIV, 1967, pp. 779-791.

ethical naturalism rather than another, presumably they are meant to show that since the predicate 'good' does not mean the same as any set of naturalistic predicates, the predicate 'good' cannot be defined by such predicates. Hence, the predicate 'good' is not entailed by any set of naturalistic predicates.

For the moment, let us assume that Hare has shown that 'good' does not have the same meaning as any set of naturalistic predicates and further that 'good' cannot be defined by such predicates. Does it follow that 'good' is not logically dependent upon or entailed by any such predicates? Does the denial of naturalistic thesis (a) entail the denial of (b)? Hare's answer is clearly in the affirmative.⁵

Hare's position on the connection between (a) and (b) can be brought out clearly by considering the distinction he draws between the predicates 'rectangular' and 'good'. Hare asks "whether there is any characteristic or group of characteristics which is related to the characteristic of being good in the same way as the angle-measurements of figures are related to their rectangularity."⁶ In the latter case, Hare believes, the characteristics are so related that it is *logically* impossible for two objects to be exactly alike in all respects including the measurements of their angles yet one be rectangular and the other not. The logical impossibility here is grounded in the fact that a statement which asserts that an object has these angle-measurement characteristics *entails* that the object is rectangular. This entailment, in turn, is based on the fact that "'rectangular' means 'rectilinear and having all its angles of a certain size, namely ninety degrees'",⁷ and thus

⁵ This is obvious in *LM*, Ch. 5. Also, in the beginning of Ch. 6. Hare writes "The argument of the preceding chapter establishes that 'good', being a word used for commending, is not to be defined in terms of a set of characteristics whose names are not used for commending. This does not mean that there is *no relation* between what have been called 'good-making' characteristics and 'good'; it means only that this relation is not one of entailment." p. 94.

⁶ *LM*, p. 82. The argument that follows appears in *LM*, pp. 81-83.

⁷ *Ibid.* In regard to this argument Hare remarks that "the impossibility that we are speaking of is a logical one which depends upon the meaning of the word 'rectangular'." *LM*, p. 83.

upon the definition of the former predicate by the latter ones.

A similar argument, Hare contends, is not forthcoming for the predicate 'good'. Although 'good' is supervenient upon certain naturalistic characteristics of an object, these characteristics do not entail that an object is good. The reason for this nonentailment is that 'good' does not have the same meaning as any such set of naturalistic characteristics. Hence, 'good' cannot be defined by any set of naturalistic predicates.

Since it is not always clear how Hare's conclusions follow from his premises, we need to lay out his argument in a more explicit manner—supplying suppressed premises when necessary. In the 'rectangular-rectilinear' case, the requisite entailment apparently is established by the following argument:

- (A) (1) For any two sets of predicates 'P' and 'Q'
 if 'P' defines 'Q' then 'P' entails 'Q'.
 (2) The predicate 'rectilinear and having all
 its angles of a certain size, namely ninety
 degrees' defines the predicate 'rectangular'.
 ∴ (3) The predicate 'rectilinear and having all
 its angles of a certain size, namely ninety
 degrees', entails the predicate 'rectangular'.

It is logically impossible, then, for an object to be rectilinear and have all its angles be ninety degrees and yet not be rectangular.

Now this argument is surely valid. And presumably Hare's long argument against the definability of 'good' by naturalistic predicates is meant to show in an analogous way that no similar entailment relation exists between 'good' and any set of naturalistic predicates. Hence, the reason for rejecting naturalism (b).

But Hare is surely wrong here. For the thesis that 'good' cannot be so defined does not, in conjunction with the premises of (A), yield the conclusion that 'good' cannot be entailed

by any set of naturalistic predicates. Thus, the following argument is invalid:

- (B) (1) For any two sets of predicates 'P' and 'Q'
if 'P' defines 'Q' then 'P' entails 'Q'.
(2') No set of naturalistic predicates defines
the predicate 'good'.
∴ (3') No set of naturalistic predicates entails
the predicate 'good'.

Since (B) is invalid, it is difficult to understand why Hare believes that the falsity of the definability thesis establishes the falsity of the deducibility thesis and why the 'rectangular-rectilinear' example is considered relevant to this belief. In order to be fair to Hare, we shall assume that he is not guilty of such a glaring logical flaw as committing the fallacy of denying the antecedent. But we also need to assume what is obvious, namely, that Hare does believe that his arguments against naturalistic definitions of 'good' are relevant to the denial of the deducibility thesis. Given these two assumptions, we are led to the view that Hare's argument against naturalism (b) is based not on premise (1), but rather on its converse. Hare's argument then becomes:

- (C) (1') For any two sets of predicates 'P' and 'Q',
if 'P' entails 'Q' then 'P' defines 'Q'.
(2') No set of naturalistic predicates defines
the predicate 'good'.
∴ (3') No set of naturalistic predicates entails
the predicate 'good'.

Now argument (C) unlike (B) is, in fact, valid. But if (C) is Hare's argument, then new problems arise. First, it is now mysterious why Hare believes that the 'rectangular-rectilinear' example is at all relevant to his attack on the deducibility thesis of ethical naturalism. For, conjoining (2) ("The predicate 'rectilinear and having all its angles of a certain size,

namely ninety degrees' defines the predicate 'rectangular'") with (1'), rather than with premise (1), does not yield the conclusion that the predicate 'rectilinear and having all its angles of a certain size, namely ninety degrees' entails the predicate 'rectangular'. And thus, such an argument cannot be used to establish Hare's claim that it is logically impossible for an object to be rectilinear and have all its angles equal ninety degrees and yet not be rectangular. Further, although (C) is valid, it cannot be used to establish the falsity of the deducibility thesis. The reason now is that premise (1') is clearly false.

Counterexamples to (1') are not hard to find. In Hare's sense of entailment the predicate 'is a bachelor' entails 'is a male', yet the former predicate obviously does not define the latter one. Again, take any predicate 'P', it entails 'P \vee Q', yet surely the former predicate does not define the latter one. Finally, any contradictory predicate 'P \cdot \sim P' entails any predicate whatever. But once again, no corresponding definition ensues. The first two examples, but perhaps not the last one, are clear cut cases of valid entailments according to Hare's sense of entailment.⁸

Hare's argument against naturalism (b) which is based on the denial of naturalism (a) is, then, either invalid or unsound. By itself, the denial of (a) does not entail the denial of (b). What is required for the entailment is an additional premise which happens to be false.

To do full justice to Hare's arguments against (b) we need to consider one final argument which is suggested by some of his remarks but which does not employ a premise explicitly denying the definability thesis. Hare writes "that what is wrong with naturalist theories is that they leave out the prescriptive or commendatory element in value-judgments by seeking to make them derivable from statements of fact."⁹

⁸ See *LM*, p. 25. In order not to introduce extra complexities, I accept throughout this paper Hare's curious criterion of entailment.

⁹ *LM*, p. 82.

Now in the language of predicates rather than of judgments what is suggested by this remark and others is that Hare is appealing to the following premise:

- (1'') For any two sets of predicates 'P' and 'Q' if 'P' entails 'Q' then it is not the case that 'Q' has some function that 'P' does not have.

Further, Hare believes he has shown that the predicate 'good' commends and that

- (2'') No set of naturalistic predicates has the same functions as the predicate 'good'.

And (1'') and (2'') entail

- (3') No set of naturalistic predicates entails the predicate 'good'.

The problem with naturalism, then, is that it derives value judgments from factual judgments and in so doing, it takes away an important function, that of commending, from value judgments.¹⁰

This new argument, however, fares no better than the

¹⁰ Hare has a variant of this argument. Reconstructed, it goes something like this:

- (1''') If a set of naturalistic predicates 'P' entailed a set of value predicated 'Q' then we could not commend an object to which 'P' applies by applying the set of predicates 'Q'.
(2''') We sometimes do commend objects to which 'P' applies by applying the value predicate 'good' to them.
(3') No set of naturalistic predicates 'P' entails the value predicate 'good'.
(See *LM*, Ch. 5; also *LM*, p. 145.)

One of the main problems with this argument lies in premise (1'''). In defense of (1'''), Hare incorrectly assumes that the entailment thesis implies the definability thesis; for he believes that if 'P' defines 'Q', then we could not commend an object to which 'P' applies by applying 'Q'. But, he reasons, since we sometimes do commend objects to which 'P' applies by calling them 'Q', 'P' cannot define 'Q'.

previous ones, for it is undermined by (1''). Whether one set of predicates entails another surely is independent of whether the entailing predicates have all the same functions as the entailed ones. Once again, 'P' entails ' $P \vee Q$ ' but the latter, being a disjunction, may function in certain contexts to express indecision while the former may not. And since any predicate whatever, including a value predicate, is entailed by ' $P \sim P$ ', the entailed predicate obviously may function differently from ' $P \sim P$ '. So, even if some value predicates were entailed by a set of naturalistic predicates, these value predicates need not, as a consequence, have only those same functions as the naturalistic predicates. And even if Hare were correct in his belief that there is a difference in function between value and naturalistic predicates, such a difference does not provide a reason for rejecting naturalism (b). Although sameness of function may be a necessary condition for sameness of meaning, and sameness of meaning a necessary condition for an acceptable definition, neither sameness of function nor sameness of meaning is a necessary condition for valid entailment.

III

Although much can be said in regard to the deducibility thesis apart from the definability thesis, I shall consider it further only in brief. The previous remarks raise an important problem for critics of naturalistic thesis (b). They cannot appeal either to the alleged falsity of the definability thesis or to the alleged differences in function between naturalistic and ethical predicates in order to reject the deducibility thesis. Further, although the rejection may be based upon an appeal to some standard accepted set of deductive rules of inference, this option is not open to Hare; for he accepts a much more informal notion of entailment. And we can now show that given his notion of entailment some value or ethical predicates are entailed by some naturalistic predicates.

Recall Hare's notion of entailment: "A sentence P entails

a sentence Q if and only if the fact that a person assents to P but dissents from Q is a sufficient criterion for saying that he has misunderstood one or other of the sentences."¹¹ Hare might argue that our previous examples of entailment relations between sets of predicates which were neither interdefinable nor performed all the same functions could not be so interpreted that some naturalistic predicates entailed some value predicates. Thus, for example, if ' P ' is a naturalistic predicate and ' Q ' a value predicate, then, although ' P ' entails ' $P \vee Q$ ', Hare might deny that any such disjunctive predicate is a value predicate. However, given Hare's notion of entailment, other examples can be found of naturalistic predicates entailing value predicates.

Hare admits that the predicates 'good' and compound predicates of which 'good' is a component do have descriptive meaning.¹² And further, he admits that 'good' and its compounds are supervenient upon this descriptive meaning. Thus, he claims the descriptive meaning of 'good strawberry' is, 'sweet, juicy, firm, red, large, and ripe strawberry'. But given this descriptive meaning, the fact that a person assents to the statement that a given strawberry is sweet, juicy, firm, red, large, and ripe, but dissents from the statement that it is a good strawberry, is a sufficient criterion for saying that he has misunderstood one or other of the statements. And hence, by Hare's own criterion, the former factual statement entails the latter value statement.

It is no reply to this argument to claim that a person may not accept the stated characteristics of a strawberry as entailing the value predicate. For, given Hare's notion of entailment, this fact clearly is irrelevant to whether the entailment holds. Hare's rejection of such entailments appears to be based principally upon the belief that the deducibility thesis implies the definability thesis. But enough has been said about this belief.

¹¹ *LM*, p. 25.

¹² *LM*, pp. 117-118.

The cited entailment is in conflict with one final position maintained by Hare. He holds that value statements, but not factual statements, entail imperatives. In light of our arguments, however, Hare must reject this view either by denying that value statements entail imperatives or by denying that factual statements do not entail imperatives. If neither of these options is accepted, then Hare must either strengthen his notion of entailment in order to prevent factual statements from entailing value statements, or he must give up the transitivity of entailment for his notion of entailment. In any case, naturalistic thesis (b) withstands Hare's attack.

IV

Up to now, we have been working on the assumption that naturalistic thesis (a) is false. To many this is not a very controversial assumption. The reasons for accepting this belief, however, are not objects of such wide agreement. And despite the popularity of this view, arguments in defense of it are frequently absent or, if given, employ criteria for acceptable definition which, when made explicit, can be shown to be unacceptable. A case in point is Hare's argument against (a).

Hare's attack on naturalism (a) employs several suppressed premises. Basically Hare's strategy is to show there is some context in which 'good' functions differently from any set of naturalistic predicates. This difference in function is then assumed to constitute a difference in meaning. And this difference in meaning provides the grounds for rejecting the definability thesis.

At least two important questions relevant to thesis (a) are raised by this kind of argument. Is it true that sameness of function in *all* contexts is a necessary condition for sameness of meaning? If so, is sameness of meaning a necessary condition for the definition of one set of predicates by another set? Since my intuitions on predicate synonymy are somewhat deficient, I am willing to accept Hare's affirmative

answer to the first question. But given this condition for predicate synonymy, is it correct to hold that the synonymy of two sets of predicates is a necessary condition for an acceptable definition of one set by the other? Once again, Hare's criticism of naturalistic definitions of 'good' presupposes an affirmative answer. But this surely is questionable, for as we shall now see, such a condition eliminates perfectly acceptable definitions.

Hare's principal argument against the definability of 'good' is roughly as follows: If 'good' meant the same as some set of naturalistic predicates 'C', then we would be prevented from saying something we sometimes want to say (or prevented from doing something we sometimes do). We sometimes want to, or actually do, commend an object which is C by calling it 'good'; for instance, we sometimes want to, or actually do, say that this is a good strawberry because it is sweet, juicy, firm, red, large, and ripe. But if 'good strawberry' meant 'sweet, juicy, firm, red, large, and ripe strawberry', then we couldn't commend a strawberry for being sweet, etc., by calling it 'good'. The reason here is that upon substitution of the proposed definiens for the definiendum the commending sentence turns into the non-commending "This is a sweet, juicy, firm, red, large, and ripe strawberry because it is sweet, juicy, firm, red, large, and ripe." Hence the proposed definition prevents us from saying something we want to say or from doing something we actually do.¹³

Hare's strategy in this argument is to show that when a proposed definiens is substituted for the definiendum the new sentence fails to perform some function performed prior to the substitution. Hence the proposed definiens is unacceptable. But the problem with this kind of argument is that if it establishes that 'good' does not have all the same functions as any set of naturalistic predicates, then it establishes some differences in function between sets of predicates which are interdefinable. For example, let us take as 'P' the pred-

¹³ *LM*, pp. 85-86.

icates 'rectangular' and 'as 'Q', the predicate 'rectilinear and having all its angles of a certain size, namely ninety degrees'. Hare believes that here 'Q' does define 'P'.¹⁴ Employing a Harean type argument, we may reason as follows: Surely, in geometry classes, an instructor sometimes wishes to tell his students why a certain figure is rectangular (or he may want to categorize in an *informative* way a certain geometric figure in virtue of its having certain geometrical properties). This is sometimes done by saying, "This object is rectangular because it is rectilinear and has all its angles of a certain size, namely ninety degrees." But if 'rectangular' meant 'rectilinear and having all its angles of a certain size, namely ninety degrees', then this sentence would mean "This object is rectilinear and has all its angles of a certain size, namely ninety degrees, because it is rectilinear and has all its angles of a certain size, namely ninety degrees". But this is not what the instructor wanted to say. The latter sentence does not have the same pedagogical function as the former one. So, if the proposed definition is accepted, the instructor is prevented from saying something he wanted to say.

A more general type of critical example is always available. For we can always find a sentence in which a definiendum cannot be replaced by a proposed definiens without a loss in function. Following Hare we can argue as follows: If the predicate 'rectangular' meant 'rectilinear and having all its angles of a certain size, namely ninety degrees' then the sentence "This A is rectangular" would mean "This A is rectilinear and has all its angles of a certain size, namely ninety degrees". But this latter sentence *is used to translate* the former sentence, whereas the former sentence obviously cannot be used (correctly) to translate itself. Hence, we have a context in which there is a function performed by the latter sentence but not by the former sentence. By Hare's criterion the two sentences thereby differ in meaning, and hence the two predicates differ in meaning. So, once again,

¹⁴ *LM*, p. 82.

following Hare's line of reasoning, we establish that if the two predicates had the same meaning we would be prevented from doing something (i.e. translating) which we actually do. Hence, they cannot have the same meaning.

These examples show that there are contexts in which the job done by the 'rectangular' sentence differs from the job done by the 'rectilinear' sentence. The two corresponding predicates, then, do not have the same meaning, and by Hare's criterion for definitions they cannot form an acceptable definition. Hare's criterion, then, is obviously too strong. Naturalistic thesis (a) is rejected at an unacceptable cost.

Something has gone wrong. It is important to note that Hare's argument does not establish that 'good' commends but that 'C' does not commend. At most, the argument establishes that there is one type of context in which 'good' but not 'C' commends. And even here, Hare establishes, at most, that in sentences such as "This strawberry which is C is C", or "This is a C strawberry because it is C", the *second* occurrence of 'C' does not commend.¹⁵ Fully compatible with this is that the first 'C' commends and that 'C' frequently commends in other sentences; for example, in sentences of the form "This A is C". Surely, we sometimes commend a strawberry by noting it is sweet, juicy, firm, red, large, and ripe. Similarly, we sometimes commend an action by noting that it has alleviated suffering, or that it was useful to society, or even that it was pleasing to God.

Given these arguments, one wonders why, if sameness of function in all contexts is taken as a necessary condition for sameness of meaning, that sameness of meaning is taken as

¹⁵ The absence of a commending function in such sentences is not peculiar to naturalistic predicates. Even the second occurrence of the predicate 'good' in the sentence "This strawberry which is good is good" may not commend. For some inexplicable reason, Hare sometimes seems to believe that he has shown that no naturalistic predicate 'C' commends, and not simply that there is one context in which the second occurrence of any given 'C' does not commend. Thus he writes, "Value terms have a special function in language, that of commending; and so they plainly cannot be defined in terms of other words which themselves do not perform this function..." *LM*, p. 91.

a necessary condition for an acceptable definition? Why demand of ethical naturalists that they produce such synonyms for ethical terms when such synonyms often are not and cannot be provided elsewhere? And even if such synonyms could be provided, why must all definitions in ethics but not in other fields satisfy this criterion? Answers to these questions are rarely considered by critics of the definability thesis. But until satisfactory answers can be found, whether the definability thesis is correct or not remains an open question.

V

If sameness of function in all contexts is not a suitable criterion of definition, several alternative criteria still remain. One might still require that in an acceptable definition the definiens be synonymous with the definiendum—under some other criterion of synonymy. However, until this criterion is made more explicit, it remains somewhat suspect. This is so for two reasons. First, but not foremost, the notion of synonymy is rather unclear and for some not philosophically respectable. Secondly, and more importantly, some definitions which we do, in fact, find acceptable simply do not satisfy any such criterion—under any reasonable construal of that notion. Many acceptable definitions, particularly in science, are, in Carnap's sense, explications.¹⁶ An explication may be viewed as a type of definition where the proposed definiens (explicatum) need not be presystematically synonymous with the definiendum (explicandum). Some definitions are considered acceptable provided that they are satisfactory explications.

¹⁶ See R. Carnap's *Logical Foundations of Probability*, Second Edition (Chicago, 1962), pp. 1-8. Hereafter *LFP*. Also see Carnap's "P. F. Strawson on Linguistic Naturalism" in P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap* (LaSalle, Ill., 1963), pp. 933-940; N. Goodman, "The Significance of *Der logische Aufbau der Welt*", in *The Philosophy of Rudolf Carnap*, pp. 545-558; and C. G. Hempel, *Fundamentals of Concept Formation in Empirical Science* (Chicago, 1952), pp. 10-14.

Naturalism in ethics can be viewed as the thesis that all ethical predicates can be explicated ultimately by non-ethical naturalistic predicates alone. An ethical naturalist might view himself as providing explications for the somewhat vague and imprecise predicates such as 'good'. In this way, the naturalist would be attempting to do for ethical terms what some philosophers and scientists have done for terms such as 'probability', 'confirms', 'knowledge', 'true', 'warmer than', 'fish', 'intelligence', 'red', etc.

Thesis (c) characterizes this version of naturalism, and although it is seldom critically examined, any serious criticism of ethical naturalism surely must come to grips with it. Its benign neglect may be attributed to one of two factors: either to the critics' ignorance about the nature of explication, or, as is suggested by some remarks of Hare's, to a mistaken view of explication. Hare writes:

It is therefore no answer to the above argument to claim that a 'naturalist' might if he pleased define 'good' in terms of some characteristics of his choice. Such an arbitrary definition is quite out of place here; the logician is, it is true, at liberty to define his own technical terms as he pleases, provided that he makes it clear how he is going to use them. But 'good' in this context is not a technical term used for talking about what the logician is talking about; it itself *is* what he is talking about; it is the object of his study, not the instrument. He is studying the function of the word 'good' in language; and so long as he wishes to study this, he must continue to allow the word the function it has in language, that of commending. If by an arbitrary definition he gives the word a different function from that which it now has, then he is not studying the same thing any longer; he is studying a figment of his own doing.¹⁷

¹⁷ *LM*, p. 92.

On the assumption that Hare intends this argument to be directed against naturalistic explications of 'good' or other value terms, there are several reasons why it is mistaken. First, presystematically, the predicate 'good' need not be a technical term in order to be given an explication by naturalistic predicates. In fact, prior to a systematic explication, it is far from clear just what constitutes a technical term. Are the logical constants, 'or', 'and', 'implies', and 'entails', technical terms prior to their explications by a logical system? Surely they have frequent uses in ordinary language. And surely, it is not a precondition for their explication that they antecedently be technical terms.¹⁸ Further, there are ordinary nontechnical uses for subsequently explicated terms such as 'probably', 'warmer than', 'fish', 'red', etc. And just such uses provide the data and objects for a systematic explication. Similarly, for 'good'. To believe that a predicate must somehow be a technical term prior to its explication is to misconstrue seriously the nature and purpose of explications.¹⁹

Secondly, such explications need not prevent the explicated predicates from performing most of the functions which they have in ordinary language. On the contrary, as Carnap has noted, "the explicatum is to be *similar to the explicandum* in such a way that in most cases in which the explicandum has so far been used, the explicatum can be used."²⁰ And as we have previously noted, naturalistic predicates as well

¹⁸ For an illuminating analysis of an ordinary language criticism of explications of such presystematic terms by formal logic, see Quine's "Mr. Strawson on Logical Theory", *Mind*, 62, 1953. Reprinted in Quine's *The Ways of Paradox and Other Essays* (New York, 1966).

¹⁹ The reader is referred to the articles mentioned in footnotes 16 and 18. Carnap writes:

An explication replaces the imprecise explicandum by a more precise explicatum... The explicatum may belong to the ordinary language, although perhaps to a more exact part of it. Or it may be that it did not belong to the ordinary language originally but was introduced as a scientific term. Even such a term will frequently be accepted later into the everyday language, such as "at 4:30 P.M.", "temperature", "speed" as a quantitative term. "P. F. Strawson on Linguistic Naturalism", pp. 935-936.

²⁰ *LFP*, p. 7.

as ethical predicates can be used to commend. A naturalistic explication of ethical terms need not, then, take away the commending function of the explicated term or of the proposed explicatum.

Thirdly, explications are not arbitrary or purely stipulative definitions. Certain requirements must be satisfied by an acceptable explication. Just which requirements need to be satisfied is a matter of some debate. Goodman requires that an accurate definition satisfy the criterion of extensional isomorphism while Carnap, in the *Aufbau*, requires extensional identity. And in *Logical Foundations of Probability*, Carnap proposed four admittedly vague requirements for a given explicatum to be adequate: namely, similarity to the explicandum, exactness, fruitfulness, and simplicity. Whether a proposed explication satisfies these criteria is surely not an arbitrary matter.²¹

Thesis (c), then, cannot be satisfied by arbitrary or stipulative definitions. Nor is it the case that a proposed explication need either violate the function which ethical terms ordinarily perform or alter the ordinarily conceived truth conditions for ethical statements. Naturalistic thesis (c), as well as (a) and (b), escapes Hare's attack.

VI

Whether or not naturalistic thesis (a) or (b) or (c) is, in fact, correct is still an open question. Cogent naturalistic theories still await development. Perhaps, it will turn out that no acceptable naturalistic theory can be developed, and that some ethical predicates cannot be defined by, or explicated by, or deduced from any set of naturalistic predicates. If so, ethical naturalism would be refuted. Whether such a result is forthcoming is still too early to say. The

²¹ For further discussion of the non-arbitrariness of explications see Hempel's *Fundamentals of Concept Formation in Empirical Science*, p. 11, and Carnap's, "Nelson Goodman on *Der logische Aufbau der Welt*" in *The Philosophy of Rudolf Carnap*, pp. 945-946.

answer will be determined only after a thorough investigation of (hopefully) forthcoming naturalistic theories. Until then, I see no force whatever to the many wholesale and general condemnations of all naturalistic theories. The serious work of constructing substantive naturalistic theories remains.

RESUMEN

Los críticos del naturalismo ético han prestado poca atención a las tesis fundamentales de dicha postura filosófica. Se han limitado a refutarla, ya sea atacando una de sus versiones o distinguiendo primero entre sus diferentes manifestaciones y criticando después sólo una de ellas. En este artículo se formulan tres versiones diferentes del naturalismo ético y se trata de ilustrar la equivocidad de la crítica de Hare estableciendo una demarcación entre estas versiones. Por último, se ponen al descubierto algunas suposiciones falsas sobre los criterios de implicación, definiciones y explicaciones, contenidas en los ataques de Hare.

I. Hay por lo menos tres tesis importantes del naturalismo ético:

a) *La tesis de la definibilidad*, que sostiene que todos los predicados éticos son definibles en última instancia únicamente en términos de predicados naturalistas no-éticos. Tesis criticada por Moore y Hare.

b) *La tesis de la deducibilidad*, que sostiene que algunos enunciados éticos (predicados) pueden ser deducidos de un conjunto de enunciados (predicados) ninguno de los cuales es ético, y que todos los otros enunciados éticos pueden ser deducidos a su vez de aquellos que previamente se enlazan con los enunciados no-éticos. Tesis atacada por Hume.

c) *La tesis de la explicación*, que establece que todos los predicados éticos pueden ser explicados únicamente a través de predicados naturalistas no-éticos.

Hare, por su parte, reconoce estas distinciones, pero construye mal la relación entre (a) y (b), pues considera que la falsedad de la primera implica la falsedad de la segunda. Además, al rechazar las tesis (a) y (c), Hare acepta concepciones equivocadas de las explicaciones y definiciones. Veamos esto más de cerca.

II. Los argumentos iniciales de Hare no van dirigidos claramente contra una versión determinada del naturalismo ético. Pretenden mostrar que, puesto que el predicado "bueno" no significa lo mismo que un conjunto de predicados naturales, tampoco puede ser definido en términos de ellos. Por consiguiente, "bueno" no es implicado por estos predicados.

Suponiendo que el predicado "bueno" no pudiera ser definido por un conjunto de predicados naturales, ¿se sigue de esto que "bueno"

no podría estar implicado en tales predicados? Hare responde afirmativamente y lo ilustra con un ejemplo que consiste en comparar el comportamiento lógico de los predicados “bueno” y “rectangular”. En el caso del predicado “rectangular” podemos establecer una relación de definibilidad y por consiguiente, de implicación —afirmación básica de Hare— con los predicados “rectilíneo con ángulos de noventa grados”. Dados dos conjuntos de predicados ‘P’ y ‘Q’, si ‘P’ define ‘Q’, entonces ‘P’ implica ‘Q’.

En el caso de “bueno”, Hare pretende concluir del hecho de que “bueno” no sea definible por predicados naturales, la imposibilidad de que pueda estar implicado en ellos. Aquí, Hare se equivoca pues de la ausencia de definición no se puede desprender la ausencia de implicación. Ahora bien, si en lugar de asumir que la definibilidad implica la implicación —como lo hace Hare—, sostenemos la premisa conversa: que la implicación implica la definibilidad, entonces el argumento de Hare sobre el predicado “bueno” se convierte en válido, con el inconveniente de que el argumento anterior sobre el predicado “rectangular” se le convierte en inválido. Parece ser que la implicación no se sigue de la definibilidad sino viceversa.

Un segundo argumento de Hare contra las teorías naturalistas (dirigido contra la tesis de la deducibilidad únicamente) consiste en afirmar que la función valorativa de los juicios éticos se pierde cuando los hacemos deducibles de proposiciones de hecho. Esta afirmación implica la siguiente premisa: “Si ‘P’ implica ‘Q’, entonces no es el caso de que ‘Q’ tenga una función que no tenga ‘P’”. De acuerdo con esto la deducibilidad de “bueno” se realizaría a costa de su función prescriptiva o recomendatoria.

Sin embargo, esta premisa también es falsa, porque la relación de implicación entre dos predicados es absolutamente independiente de las funciones que cumplen. Por ejemplo, ‘P’ implica ‘PvQ’, pero este último, por ser una disyunción, puede funcionar en ciertos contextos expresando una indecisión, mientras que el primero, por ser sólo una afirmación, no la expresaría necesariamente. El hecho de que los predicados valorativos puedan estar implicados por predicados descriptivos, no implica a su vez, que los segundos deberán cumplir únicamente las funciones de los primeros. “Bueno” puede ser deducido de predicados naturales y conservar al mismo tiempo, su función valorativa.

III. Por otra parte, la noción de implicación mantenida por Hare es lo suficientemente amplia como para admitir la deducción de predicados valorativos a partir de predicados descriptivos. Esta noción dice así: “Una oración P implica una oración Q, si y sólo si,

el hecho de que una persona asienta P pero desienta de Q es un criterio suficiente para decir que ha entendido mal una oración u otra". Con esta noción es posible deducir el predicado valorativo "fresas buenas" de los predicados descriptivos "fresas dulces", "jugosas", "rojas", "grandes" y "maduras", y por lo tanto, utilizando el criterio de Hare, las oraciones descriptivas implicarían la oración valorativa. Si se desea evitar esto, la noción de implicación deberá ser afinada, cosa que Hare no hace. De ese modo la tesis naturalista se conserva intacta ante los argumentos de Hare.

IV. Recordemos que hasta ahora, hemos estado trabajando bajo la suposición de que la tesis de la definibilidad es falsa, y aún así, hemos demostrado la validez de la tesis de la deducibilidad. A partir de aquí, nos ocuparemos de defender la tesis (a) contra los argumentos de Hare. Este último sostiene que el predicado "bueno" funciona en forma diferente a los predicados descriptivos, que el funcionamiento diferente produce diferentes significados, y que esta diferencia de significados imposibilita la definición de "bueno" en términos descriptivos. De nuevo, el argumento de Hare se basa en dos supuestos discutibles: (1) que la misma función en *todos* los contextos es condición necesaria para la sinonimia (mismo significado), y (2) que la sinonimia es condición necesaria para la definibilidad.

Teniendo como verdadero el primero, se demuestra que el segundo es cuestionable pues elimina definiciones perfectamente aceptables.

De acuerdo con las afirmaciones de Hare, "fresas buenas" y "fresas dulces, maduras y jugosas" no son expresiones interdefinibles, y por tanto sinónimas, ya que no son sustituibles en ciertos contextos, pues "buenas" cumple una función de elogio o recomendación que no cumplen "dulces, maduras y jugosas". Esto es, "las fresas son buenas porque son dulces, maduras y jugosas" y "las fresas son dulces, maduras y jugosas porque son dulces, maduras y jugosas", no son enunciados equivalentes pues el segundo carece de la función de elogio que tiene el primero.

Sin embargo, esta forma de argumentar permite rechazar la definición de "rectángulo" en términos de "rectilíneo con ángulos de noventa grados" porque si sustituimos el definiendum por el definiens en la proposición "el instructor quiere enseñar a sus alumnos que el rectángulo es una figura rectilínea con ángulos de noventa grados", que cumple una función pedagógica informativa, obtenemos la proposición trivial "el instructor quiere enseñar a sus alumnos que una figura rectilínea con ángulos de noventa grados es una

figura rectilínea con ángulos de noventa grados” que carece de la función pedagógica anterior.

El criterio de Hare resulta demasiado fuerte, permite rechazar el naturalismo pero a un precio demasiado costoso.

Por otra parte, Hare no ha invalidado el hecho de que los predicados descriptivos pudieran tener una función de elogio en otras oraciones, como cuando elogiamos una fresa por ser “dulce, madura y jugosa”, o cuando elogiamos una conducta por ser “útil a la sociedad” o “aprobada por Dios”. Con esto mostramos el carácter discutible de los supuestos (1) y (2) del argumento de Hare contra la tesis de la definibilidad, por lo que la validez de esta tesis sigue siendo una cuestión abierta.

V. Por último se trata de esclarecer los errores cometidos por Hare en torno a la tesis de la explicación. Esta tesis surge del carácter especial de ciertas definiciones que, siendo aceptables, no satisfacen un criterio como el de Hare. Estas definiciones son en el sentido de Carnap, explicaciones. Una explicación es un tipo de definición en donde el definiens (explicatum) no tiene que ser sistemáticamente sinónimo del definiendum (explicandum). La tesis (c) del naturalismo sostiene, por lo tanto, que todos los predicados éticos pueden ser explicados por predicados no-éticos, descriptivos. “Bueno” se explica en la misma forma que “probabilidad”, “verdad”, etc.

Hare aduce en contra de lo anterior, que sólo los términos técnicos pueden ser explicados arbitrariamente a través de predicados naturales; que “bueno” no puede ser definido arbitrariamente como los términos de la lógica; que no es un instrumento técnico sino un objeto de estudio, y que si no respetamos sus funciones valorativas en el lenguaje, al acudir a otros predicados, éstas se perderían; lo que producirá un cambio en el objeto de investigación.

Parece ser que Hare se equivoca en los siguientes puntos:

a) Es falso que sea necesario que “bueno” sea un término técnico para que pueda ser explicado a través de predicados descriptivos.

b) Es falso que al acudir a otros predicados, se perdería la función valorativa de “bueno”; como ya vimos anteriormente.

c) Es falso que las explicaciones sean definiciones arbitrarias o estipulativas, pues éstas deben satisfacer ciertos requisitos: exactitud, simplicidad, utilidad, etc.

VI. De todo lo anterior se concluye que la crítica de Hare al naturalismo ético ha sido infructuosa pues no pudo derribar ninguna de sus versiones, y que la validez de dicha postura sigue siendo una cuestión en debate.