

### Seminar 4: Moore's Ethics

In the preface to *Principia Ethica*, Moore distinguishes two kinds of ethical questions:

- A. What kinds of things ought to exist for their own sakes?  
are intrinsically good?
- B. What kinds of actions ought we to perform?  
are right?

The different versions of A are equivalent. The different versions of B are almost, but not quite, equivalent. Every act we ought to perform is right, but an act can be right, without being what we ought to perform when there is a different act that is also right and what we ought to do is to perform one or the other. But for this exception, Moore takes rightness and moral duty to coincide.

Corresponding to A and B questions are A and B ethical statements. Example A-statements are:

- The apprehension of beauty is (intrinsically) good.
- Knowledge is (intrinsically) good.
- Friendship is (intrinsically) good.

Examples of B-statements are:

- Keeping one's promises is right.
- Telling the truth is right.
- Helping others is right.

In the preface, Moore announces two theses about A-statements and B-statements.

- T1. If the conclusion of an argument is an A-statement, but none of its premises are, then the premises do not entail the conclusion, and their truth provides no evidence for it, or any compelling reason to believe it.
- T2. If the conclusion of an argument is a B-statement, then the premises entail the conclusion only if they include both an A-statement and a "causal statement" (or another B-statement).

Thesis 2 expresses Moore's *consequentialism* – the view that the rightness of an action is wholly determined by the goodness or badness of the states of affairs it brings about.

The classical *utilitarianism* of Bentham and Mill is a theory of this kind.

- 1a. An act is right iff it produces more good consequences than any alternative act open to the agent.
- b. Happiness and happiness alone is good.
- c. Therefore, an act is right iff it produces more happiness than any alternative act open to the agent.

The first premise is common to all consequentialist theories, in accord with T2. The second premise is a statement of type A. Different versions of consequentialism result from taking different A-statements to play the role of (1b). For Moore, principles of type A provide the foundation of all ethical judgments.

They are the subject of his main thesis, T1, which was viewed as a bold claim. Ordinarily, one would think that the claim that something is good can sometimes be supported by evidence and argument. In such cases, it, we say that x is good *because* x is so and so – where *x is so and so* is not itself an explicitly evaluative claim, requiring still further defense. If T1 is correct, this natural idea is mistaken.

The main premise Moore uses to support T1 is thesis T3.

T3. *Good* is indefinable.

Moore thinks he can demonstrate the truth of T3. He also thinks that the truth of T1 follows from the truth of T3 plus one of his further theses T4.

T4. It is impossible to know what constitutes evidence for the proposition that something is good unless one knows the definition of *good*.

His reasoning seems to be this: If *good* is indefinable, then no one can know the definition of *good*, because it has no definition. If, in addition, T4 is true, then no one can know what constitutes evidence that anything is good. This in turn suggests that there can be no evidence for the proposition that a particular thing is good, or any compelling reason to accept it. Hence, T1.

#### Points of Clarification

Concerning Definition: Immediately after asserting T4 in section 5 of chapter 1, he tells us that concepts or properties expressed are the primary targets of definitions, rather than the words that express them. For Moore, a concept or a property is what a word contributes to propositions expressed by sentences containing the word. What then is it for a concept to be definable, or indefinable?

“My point is that ‘good’ is a simple notion, just as ‘yellow’ is a simple notion...Definitions of the kind that I was asking for, definitions which describe the real nature of the object or notion [i.e. of the concept] denoted by the word...are only possible when the object or notion [concept] is something complex. You can give a definition of a horse [i.e. of the concept *horse*] because a horse has many different properties and qualities, all of which you can enumerate. But when you have enumerated them all, when you have reduced a horse to his simplest term [i.e. to the simplest concepts/properties that make up anything that is a horse], then you can no longer define those terms [concepts]... And so it is with all objects [concepts] ... which we are able to define: they are all complex; all composed of parts [simpler concepts], which may themselves... be capable of similar definition, but which must in the end be reducible to simplest parts [concepts], which can no longer be defined. But yellow and good, we say, are not complex: they are notions of that simple kind, out of which definitions are composed and with which the power of further defining ceases.” (chapter 1 section 7)

“When we say, as Webster says, “The definition of horse is “a hoofed quadruped of the genus *Equus*,” we may mean three different things. (1) We may mean merely: ‘When I say “horse,” you are to understand that I am talking about a hoofed quadruped of the genus *Equus*.’... (2) We may mean, as Webster ought to mean: ‘When most English people say “horse,” they mean a hoofed quadruped of the genus *Equus*.’... But (3) we may, when we define horse, mean ... that a certain object [concept]... is composed in a certain manner: that it [really anything falling under it] has four legs a head, a heart, a liver, etc., etc., all of them arranged in definite relations to one another. It is in this sense that I deny good to be definable. I say that it [the concept *good*] is not composed of any parts [concepts], which we can substitute for it in our minds when we are thinking of it. We might think just as clearly and correctly about a horse [or about horses in general], if we thought of all its parts and their arrangement [or of all the simple concepts in terms of which the concept *horse* is analyzed, and how they are related to one another in the analysis] instead of thinking of the whole... but there is nothing whatsoever which we could so substitute for good: and that is what I mean, when I say that good is indefinable.” (chapter 1 section 8)

Moore wants to show there is no definition of the word ‘good’ that gives an analysis of the property *goodness* that we use it to express. He assumes that if P is a predicate, a definition of P is a definition (or analysis) of the property we use P to express – a definition expressed by a true sentence [the property *being P* is the property *being D*], where D is a word or phrase. E.g., a definition of the word ‘square’ tells us that the property *being square* is the property *being rectangular with 4 equal sides*. So the word ‘square’ is standardly used to express a complex property the constituents of which are *being rectangular* and *having 4 equal sides*. Since this property is also expressed by the phrase ‘rectangular with 4 equal sides’ the word ‘square’ means the same as this phrase, and one can be substituted for the other in any sentence without changing its meaning, or the proposition it expresses. In saying that *good* is indefinable, Moore is saying that the word ‘good’ can’t be defined in this way; the property we use it to express is unanalyzable because it has no constituent properties.

The second clarification requires us to distinguish *knowing the meaning of a word* from *knowing its definition*. If *being good* is an unanalyzable property, then the word ‘good’ has no definition, but it still

has a meaning, which is just the simple property it expresses. So, one can know what ‘good’ means, even though the word is indefinable.

The third point of clarification (needed to understand T4) involves the relationship between *knowing that something is good* and *having evidence that it is*. For Moore, the statement that one can’t know what is evidence for the claim that *x* is good does **not** entail that one can’t know that *x* is good. On the contrary, he believes that some things are known without evidence – i.e., without inferring their truth, or even their probable truth, from other, more basic, claims. E.g., he thinks we can know that something is *yellow*, not by inferring the proposition that it is from more basic claims that provide evidence for it, but simply by looking at it under proper conditions. He believes it is similarly possible to know that certain things are *good* simply by considering the question of their goodness, and properly distinguishing that question from others with which they might be confused.

Even with these clarifications, Moore moves very quickly from indefinability (T3) to the impossibility of evidential support (T1), without explaining the connection between the two. To understand this we must first look more closely at the connection between definitions and analytic truth for Moore.

### Definition and Analyticity in Moore

An analytic statement for Moore is one that is either a logical truth, or can be turned into a logical truth by putting synonyms for synonyms. Thus, in claiming *good* to be indefinable, Moore takes himself to be saying that no statement *For all x is good iff x is D* can be turned into a logical truth by replace the expression *D* by a descriptive word or phrase that is synonymous with *good*, by virtue of expressing the same property that *good* does. However, he didn’t take all statements of the form *For all x is good iff x is D* to be false. In fact, he held that something along the lines of (2) is true.

2. For all *x*, *x* is good iff *x* is the contemplation of a beautiful object, or *x* is the enjoyment of human companionship.

His point is that no such truth is analytic. He makes a similar claim about statements of the form *For all x, if x is D, then x is good*. According to Moore, no such statement is analytic. In section 6 of chapter 1 sums things up as follows.

“If I am asked, ‘What is good?’ my answer is that good is good, and that is the end of the matter. Or if I am asked ‘How is good to be defined?’ my answer is that it cannot be defined, and that is all I have to say about it. But disappointing as these answers may appear, they are of the very last importance. To readers who are familiar with philosophic terminology, I can express their importance by saying that they amount to this: That propositions about the good are all of them synthetic and never analytic; and that is plainly no trivial matter. And the same thing may be expressed more popularly, by saying that, if I am right, then nobody can foist upon us such an axiom as that ‘Pleasure is the only good’ or that ‘The good is the desired’ on the pretense that this is ‘the very meaning of the word.’”

This denial that any statements involving ‘good’ are analytic is what leads Moore to think that no conclusion that *x* is good can be *derived* from premises not involving *good*, even though he grants that some synthetic statements about goodness are true – e.g., *that human companionship is good*.

He also maintains that premises that don’t mention goodness can never even provide *evidence* that something is good. This point can be illuminated by considering statements that predicate the simple property *being yellow* of things.

- 3a. Lemons are yellow

One can say this without saying that *to be a lemon is the same thing as to be yellow*. Moore would further hold that *being yellow* not an essential property of lemons, since there is nothing incoherent about a world in which lemons are orange. So, he would say, *the statement that lemons are yellow is synthetic*.

A similar point can be made about equivalences.

3b. For all  $x$ ,  $x$  is yellow iff  $x$  reflects light waves in such and such ways.

Suppose investigations into the physics of light established the truth of an equivalence of this kind. Though Moore would admit that (3b) is true but deny it is analytic. In support of this denial, he would point out that an ordinary person might know that something is yellow without having any idea about how light waves are reflected. Thus, Moore would say, it's not part of our concept *being yellow* that anything that is yellow must reflect light waves in certain precise ways. Rather, we use one set of criteria to determine whether something is yellow – namely just looking at it – and another set of criteria to determine how it reflects light waves. It is an empirical discovery, not a conceptual truth, that the two sets of criteria are satisfied by the same things. So, (3b) is synthetic. Moore says the same thing about equivalences involving 'good'. Although there are true statements of the form  *$x$  is good iff  $x$  is so and so*, none is analytic. The reason 'good' and 'yellow' are alike in this way is that *being good* and *being yellow* are simple, unanalyzable properties. They differ in that while we can tell that something is yellow by sense perception, we determine that something is good is by intellectual intuition. This is a sign that *being yellow* is a natural property, while *being good* is non-natural.

Simple, Indefinable Properties: *Being Good and Being Yellow*

Given all this, we can understand Moore's claim that the conclusion that something is good isn't entailed, or in any way supported, by premises that don't mention goodness. It is analogous to a claim that could be made about being yellow.

T1y. If the conclusion of an argument is a statement that something is yellow, but none of its premises are, then the premises do not entail the conclusion and their truth provides no evidence for, or any compelling reason to believe, it.

One typically establishes that something is yellow, not by argument, but by looking. Although there are cases in which an argument might be given, Moore wouldn't take them to falsify T1y. Consider the dialog: Q. *What's in the box? Is it something yellow?* A: *It's a lemon.* C: *Then it is probably something yellow.* Here it might seem that the premise *It's a lemon* provides evidence for the conclusion *It's yellow*, and hence a reason for believing it. Moore wouldn't regard this as a genuine counterexample to T1y. He would say the argument relies on a suppressed premise, *All (most) lemons are yellow*, which itself depends on observation rather than argument. Once this premise is added to the previous dialog, the argument's premises will contain a statement about what things are yellow, and so will cease to be a counter example to T1y. I suspect Moore would say the same thing about the following case.

*What color is that thing at the blast sight? It reflects light waves of frequency  $n$ . Then, it must be yellow*

Moore would say this argument is sound only because it relies on a suppressed premise that has already been established – *Anything that reflects light waves of frequency  $n$  is yellow*. Examples like this make it seem plausible that conclusions about what is yellow must ultimately rest on observations, rather than on demonstrative arguments the premises of which don't mention *yellow*. Moore made a similar claim about goodness. The main difference is that we don't come to see that something is good in the same way we see that a thing is yellow. We see that something is yellow with our eyes. We come to see that something is good with our intellect – simply by clarifying what is before our minds.

**Good is Indefinable: The Open Question Argument**

The argument is given in section 13 of chapter 1. Moore says that 'good' is indefinable because no matter what definition is offered it is always meaningful to ask of whatever satisfies the defining complex whether it is good. He illustrates this point by considering a sample definition.

G. For all  $x$ ,  $x$  is good iff  $x$  is what we desire to desire.

Moore reasons that if G were a genuine definition, it would also give us the meaning of 'good', in which case 'good' and 'what we desire to desire' would express the same property, and so mean the same thing. But, he thinks, we can easily show that 'good' does not mean this by considering Q1.

Q1. Granted that x is what we desire to desire, is x good?

No matter what you might think the answer to this question is, Moore says, it is clear that the question is just as intelligible, and makes just as much sense, on reflection, as Q2.

Q2. Is x good?

But if 'good' and 'what we desire to desire' expressed the very same property, and so meant the same thing, then we could always substitute one for the other in a sentence without changing the proposition, or question, it expresses. So, if G were a correct definition, Q1 and Q3 would express the same question.

Q3. Granted that x is what we desire to desire, is x what we desire to desire?

But they don't, since the questions they express are different, G doesn't define 'good'.

Moore's argument can be reconstructed as follows:

- P1. If (i) [For all x, x is good iff x is D] is a definition of 'good', then 'good' expresses the same property as D, and the two expressions mean the same thing.
- P2. If 'good' expresses the same property as D, and the two expressions mean the same thing, then the sentences (ii) [Granted that x is D, is x good?] and (iii) [Granted that x is D, is x D?] expresses the same self-answering question (i.e. (ii) is on a par with (iv) "Granted that x is a male sibling of y, is x a brother of y?" in that properly understanding these sentences, and reflecting on the propositions they express, should enable one to see that the answer to the questions is 'yes').
- P3. There is no complex property (not itself containing goodness as a constituent), or simple natural property, p, and expression D, such that D expresses p, and (ii) in P2 expresses the same self-answering question as (iii); nor could we introduce such an expression D.
- C1. Therefore, there is no definition of 'good' [For all x, x is good iff x is D] in which D expresses either a complex property, or a simple natural property.
- C2. So, 'good' is indefinable, and must express a simple non-natural property.

The premises of this argument seem plausible. P2 embodies the natural assumption that the meaning of a sentence is a function of the meanings of its parts, while P1 is a reasonable statement of what we want from at least one significant kind of definition. Although not beyond question, these assumptions are attractive, and, for our purposes, may be accepted. Given this, our assessment of the argument depends on our assessment of P3. Moore's acceptance of P3 was probably based on the assumption that meaning is transparent.

### The Transparency of Meaning

If expressions  $\alpha$  and  $\beta$  mean the same thing (e.g. if two predicates express the same property), and if an agent x (fully) understands  $\alpha$  and  $\beta$ , then x will be in a position to know (i) that they mean the same thing, and (ii) that any two sentences (of the sort under consideration) differing only in the substitution of one expression for the other will mean the same thing, and express the same proposition (in the case of declaratives) or question (in the case of interrogatives).

Moore is tacitly relying on this principle when he takes it for granted that if D gave the meaning of 'good', then anyone who (fully) understood both could see by introspection that the interrogatives (ii) and (iii) of P2 expressed the same question. Since it is plausible that we do (fully) understand 'good' and related expressions, and that we wouldn't judge the questions to be identical, he takes P3 to be true. As

well he should, *given that he accepts the transparency principle*. However, it can be doubted. Although it has intuitive appeal, and was accepted by most analytic philosophers in the early to mid-20<sup>th</sup> century, counterexamples to it have been found in recent decades – many involving proper names and natural kind predicates. These counterexamples are important in other contexts. But their relevance to Moore's argument is limited and controversial. Since there are more serious problems with his overall argument, I won't use such counterexamples to the transparency thesis to challenge his conclusion that *good* is indefinable.

### The role of the indefinability thesis in Moore's argument for T1

The indefinability thesis, T3, is the 1<sup>st</sup> step in Moore's (implicit) argument for T1. In reconstructing the argument we take D to be a word or phrase (excluding those themselves defined in terms of 'good') that stands for either a complex property or a simple natural property. S4 elaborates T1.

#### 3.1 The Argument

- S1. There is no D such that [For all x, x is good iff x is D] is a definition of 'good'.
- S2. There are no analytic equivalences, [For all x, x is good iff x is D], and no analytic generalities, [If x is D, then x is good], for any D.
- S3. There is no entailment of the statement (expressed by) [ $\alpha$  is good] by the corresponding statement (expressed by) [ $\alpha$  is D], for any D.
- S4. No statement (expressed by) [ $\alpha$  is D], for any D, provides any evidence for the conclusion (expressed by) [ $\alpha$  is good], or any compelling reason to believe it. The claim that a particular thing is good can sometimes be derived from a general principle which states that all members of a certain class are good. But the fundamental principles of ethics – which state that all, or all and only, members of a certain class are good, and which provide the basis for justifying all other ethical claims – are self-evident propositions for which no justification is needed or possible; such propositions must simply be seen to be true.

#### Interpretation 1

Given Moore's strict sense of *definition*, his argument for S1, is plausible. But there are serious questions about his move to S2 – S4. First consider the move to S2. Moore treats his argument that 'good' is indefinable as if it were sufficient to establish that there are no *analytic equivalences or generalities* connecting goodness with properties expressed by any relevant D. It is therefore striking that he devotes so little attention to *analyticity* and related notions central to his argument. The few places in *Principia Ethica* at which he talks about analyticity suggest that he takes analytic truths to be *necessary truths the falsity of which is "inconceivable," and the negations of which are "contradictory."* This fits with his discussion in "The Refutation of Idealism," which supports the conclusion that he took analytic truths to be those that can be turned into logical truths by replacing synonyms with synonyms. When this conception of analyticity is combined with his strict conception of definition, and of what counts as synonymy, the analytic truths turn out to be a small subset of those that express necessary, a priori propositions. On this interpretation, the gap between S1 and S2 is minimized.

But the narrow conception of analyticity used to validate the move from S1 to S2 makes problems for the move from S2 to S3 and S4. To get to S3 one needs to say something about entailment. Moore identifies this relation with *logical implication* – a proposition p *entails* a proposition q iff p *logically implies* q – i.e. iff q is a *logical consequence* of p. But he didn't mean by *logical implication* and *logical consequence*, what we now mean. For Moore, logical implication and logical consequence were relations between *propositions* or sets of propositions; in modern logic they are relations between *sentences* or sets of sentences. Propositions are information sentences encode, agents believe, and assert. Synonymous sentences encode the same proposition, and different propositions may be encoded by different uses of the same sentence, if the sentence contains an indexical expression like 'I' or 'now'.

In addition, Moore took S3 to be a momentous philosophical thesis, but its counterpart S3<sub>1</sub>, involving the modern notion of logical implication, is a triviality.

S3<sub>1</sub>. For any relevant D, and name n, the sentence [n is D] does not logically imply [n is good].

S3<sub>1</sub> is a triviality because the mere fact that the word ‘good’ doesn’t appear in D is enough to ensure that [n is D] doesn’t logically imply [n is good]. A simple example illustrates this point. In modern logic, the sentence ‘The *object* is neither *round* nor *square*’ logically implies the sentence ‘The *object* isn’t *round*’ because any interpretation assigned to the non-logical words made the first sentence true would make the second true too. Similarly uniformly replacing the non-logical vocabulary with other non-logical words, while leaving the logical vocabulary intact, will never yield a pair of sentences differing in truth value. By this criterion the sentence ‘A *square* is *inside* the *circle*’ does **not** logically imply ‘A *rectangle* is *inside* the *circle*’, because the definition of logical implication doesn’t constrain the words replacing ‘square’ and ‘rectangle’ to be related. Since Moore would insist that the proposition that a square is inside the circle *does* entail the proposition that a rectangle is inside the circle, the entailment relation in Moore’s S3 *isn’t* logical implication in the modern sense. What is it?

At this point our interpretation of Moore is pulled in two directions – one aimed at validating the move from S2 to S3, and one aimed at validating the move from S3 to S4. First the former. Recall our account of Moorean analyticity – S is analytic iff S can be turned into a logical truth by putting synonyms for synonyms (where a logical truth is one comes out true no matter how its non-logical vocabulary is interpreted). This definition can be extended to propositions by defining a *proposition* to be analytic iff it is expressed by an analytic sentence. *Entailment* of propositions can then be defined in terms of *analytic implication* of.

#### Analytic Implication: Sentences

A sentence, or set of sentences, S analytically implies a sentence R iff there is a sentence, or set of sentences, S’ and a sentence R’ that arise from S and R by replacing synonyms with synonyms, and S’ logically implies R’.

#### Entailment: Propositions

A proposition (or set of propositions) p entails a proposition q iff there is a sentence (or set of sentences) S that expresses p (or if p and S are sets, the sentences in S express the propositions in p) and there is a sentence R, such that R expresses q, and S analytically implies R.

On this account of entailment and analyticity, S3 follows unproblematically from S2. But these notions of entailment and analyticity are very restrictive. Analytic *propositions* are those expressed by logical truths, and the entailment relation holds only between propositions p and q expressed by sentences s<sub>p</sub> and s<sub>q</sub>, one of which logically implies the other (in the strict formal sense). Worse, *the move from S3 to S4 becomes hopeless*. In order to reach S4, Moore must rule out the possibility that for some relevant D, (4) expresses a necessary a priori truth that is validated by the kind of reasoning available in philosophy.

4. If α is D, then α is good

If there is such a D, then the proposition P<sub>G</sub> expressed by [α is good] will be an a priori, necessary, and philosophically validated consequence of the proposition P<sub>D</sub> expressed by [α is D], in which case P<sub>D</sub> might well constitute a proof of P<sub>G</sub>, or at least a compelling reason for drawing that conclusion. Such a result would falsify Moore’s most important meta-ethical thesis about goodness; namely that there can be no proofs of, or compelling arguments for, claims to the effect that something is, or isn’t, good.

Moore’s vulnerability is illustrated by the necessary, a priori, and self-evidently obvious examples in (5).

- 5a. For all x, if x is *chartreuse*, then x is *colored*.
- 5b. For all x, if x is an *automobile*, then x is a *vehicle*.
- 5c. For all x, if x *intends to win*, then x *doesn’t know that x won’t win*.

- d. For all propositions  $p$ , if  $p$  predicates being red of an object  $o$ , without predicating anything else of  $o$ , and without predicating anything of anything else, then  $p$  is true iff  $o$  is red.

Consider (5a). Someone like Moore, who believed in the transparency of meaning, might admit that a competent speaker who knew the meanings of both ‘chartreuse’ and ‘colored’ would thereby realize that (5a) expresses a truth, and that anyone who entertained the proposition it expresses would be in a position to judge it to be true. This might be so despite the facts (i) that ‘chartreuse’ isn’t *defined* in terms of ‘colored’ – since an individual might possess the concept *being chartreuse*, without possessing the concept *being colored* (under which anything chartreuse, magenta, crimson, etc. falls) – and (ii) that ‘colored’ isn’t *defined* in terms of ‘chartreuse’ – since one can know what it is for something to be colored without knowing all the colors. So, it is plausible to suppose that we could establish that ‘colored’ can’t be defined, in Moore’s strict sense.

Combining this result with our definitions of analyticity and entailment, we could get to  $S3_c$ .

$S3_c$  For any relevant  $D$ , and name  $n$ , the statement expressed by  $[n$  is  $D]$  does not entail the statement expressed by  $[n$  is colored].

*But nothing interesting follows.* The claim that something is chartreuse provides both evidence for, and compelling reason to believe, the claim that it is colored. One could *prove* or *establish* that a thing is colored by showing that it is chartreuse. So, the version of  $S4$  involving the predicate ‘colored’ is false, even though  $S3_c$  is true. Since the move from the one to the other parallels the original move from  $S3$  to  $S4$ , involving ‘good’,  $S4$  does not follow from  $S3$ . So, on this interpretation, Moore fails to establish his most important methodological conclusion.

### Interpretation 2

Perhaps the problem lies in an unduly narrow conception of analyticity and entailment. *Examples like (5) illustrate that two expressions can be conceptually connected even though neither is defined in terms of the other.* Similarly, the proposition expressed by one sentence may be a necessary and a priori consequence of the proposition expressed by another, even though neither sentence is transformable into the other by putting synonyms for synonyms, and no chain of definitions relates the two. One might take this to indicate the need for notions of analyticity and entailment that recognize conceptual connections not grounded in definitions.

#### Analytic Obviousness: Sentences and Propositions

Let  $S$  be any sentence that is necessary, that expresses something knowable a priori, and that is so obvious that anyone who understands it is disposed to accept it, and anyone who entertains the proposition it expresses is inclined to judge it to be true. Call any such sentence, as well as the proposition it expresses, *analytically obvious*.

#### Analytically Obvious Consequence: Sentences and Propositions

A sentence  $R$  is an *analytically obvious consequence* of a (finite) set  $S$  of sentences iff the conditional sentence the consequent of which is  $R$  and the antecedent of which is the conjunction of the sentences in  $S$  is analytically obvious. A proposition  $q$  is an analytically obvious consequence of a (finite) set  $p$  of propositions iff there is some sentence  $R$  that expresses  $q$  and some set  $S_p$  of sentences that express the propositions in  $p$ , and  $R$  is an analytically obvious consequence of  $S_p$ .

Let us grant, for the sake of argument, that the examples in (5) are analytically obvious and that (6b) as an analytically obvious consequence of (6a).

- 6a.  $n$  is chartreuse  
6b.  $n$  is colored



Next we need the notion of a sentence or proposition that can be derived from other sentences or propositions by a series of analytically obvious steps.

#### Extended Analytic Consequence: Sentences and Propositions

A sentence  $R$  is an *extended analytic consequence* of a set  $S$  of sentences iff it is possible to construct a proof of  $R$  each line of which is either a member of  $S$  or an analytically obvious consequence of earlier lines in the proof. A proposition  $q$  is an extended analytic consequence of a set  $p$  of propositions iff some sentence  $R$  expresses  $q$ , and the members of some set  $S_p$  of sentences express the propositions in  $p$ , and  $R$  is an extended analytic consequence of  $S_p$ .

#### Extended Analyticity: Sentences and Propositions

A sentence is *extendedly analytic* iff either it is analytically obvious, or it is an extended analytic consequence of some set of analytically obvious sentences. Extendedly analytic propositions are expressed by extendedly analytic sentences.

*Extended entailment* among propositions is the converse of *extended analytic consequence* –  $p$  extendedly entails  $q$  just in case  $q$  is an extended analytic consequence of  $p$ .

(7) illustrates the difference between analytic obviousness and extended analyticity.

7. For all  $x$ ,  $x = 2^{11}$  iff  $x = 2048$

$2^{11}$  is not a Moorean *synonym* for *2048* because one can understand both expressions without knowing that  $2^{11}$  is 2048. Thus, the question (7a) is a non-self-answering, question distinct from (7b).

7a. Granted that  $n = 2048$ , does  $n = 2^{11}$ ?

7b. Granted that  $n = 2048$ , does  $n = 2048$ ?

So (7) isn't a definition in Moore's sense. It isn't analytically obvious, since one can understand what the sentence means without realizing that it is true, and one can entertain the proposition it expresses without being in a position to judge it to be true. But it is necessary, a priori, and, arguably, extendedly analytic, since it can be proved from obvious, self-evident premises by obvious, self-evident steps.

Now suppose that analyticity in S2 and entailment in S3 of Moore's argument are defined as *extended analyticity* and *extended entailment*. Then, S3 follows from S2 and provides a reasonable basis for something approximating S4. If S2, and hence S3, are established, it will follow that there can be no *proof* of the claim expressed by [ $\alpha$  is good] from the premise expressed by [ $\alpha$  is D] each step of which is either analytically obvious, or an analytically obvious consequence of earlier steps. In *this sense*, no proof of the claim that  $X$  is good would be possible from premises not mentioning goodness. This isn't quite Moore's S4, which says the claim expressed by [ $\alpha$  is D] never provides any *evidence* for the conclusion [ $\alpha$  is good], or any reason to believe it. But it would be a step in that direction.

But this presupposes that S2 can be validly inferred from S1, when analyticity is interpreted as extended analyticity. It can't be. Moore's open question argument establishes only that 'good' is indefinable in his strict sense. A similar argument could be given that 'colored' is indefinable, despite the fact that generalities like (5a) may be extendedly analytic (if any sentences are). So, on this interpretation, Moore's overall argument fails to get beyond S1.

#### **Repairing Moore's Argument: Why Definability is Not the Issue**

The key question is whether the claim that  $x$  is good can be derived from the claim that  $x$  is D by a series of steps each of which is obvious *in the way that steps in a mathematical proof are*. The fact that non-obvious mathematical truths can be derived by a series of obvious steps is what makes it possible for many mathematical truths to be both surprising and rationally certain. If one could show that interesting claims about goodness can *never* be derived in this way from premises not mentioning goodness, then one would have taken a step toward Moore's desired meta-ethical conclusion. Definability is not the key question. To accomplish his goal Moore needs something stronger than indefinability.

### Expanded Open Question Argument

- P1. If [For all  $x$ , if  $x$  is  $D$ , then  $x$  is good] is analytically obvious, then the question -- (i) [Granted that  $\alpha$  is  $D$ , is  $\alpha$  good?] -- is a self-answering question on a par with (ii), (iii) and (iv) -- (ii) [Granted that  $\alpha$  is a male sibling of  $\beta$ , is  $\alpha$  a sibling of  $\beta$ ?], (iii) [Granted that  $\alpha$  is chartreuse, is  $\alpha$  colored?], (iv) [Granted that the successor of  $n =$  the successor of  $m$ , is it the case that  $n = m$ ?]. In each case, the proposition corresponding to the question is an obvious necessary, and a priori truth; anyone who truly understands the interrogative sentence, and entertains the question it expresses, is in a position to realize that the answer to it is 'yes'. Inability to see this would be evidence that one doesn't fully understand the sentence, or grasp the question.
- P2. There is no complex or simple natural, property  $P$ , and expression  $D$ , such that  $D$  expresses  $P$ , and the interrogative sentence (i) in P1 is a self-answering question on a par with (ii), (iii), or (iv).
- C1. Thus, there is no analytically obvious generality [For all  $x$ , if  $x$  is  $D$  then  $x$  is good] in which  $D$  expresses either a complex  $p$  or a simple natural property.
- C2. So, there is no extendedly analytic sentence [For all  $x$ , if  $x$  is  $D$  then  $x$  is good] in which  $D$  expresses either a complex, or a simple natural, property.

C1 follows from P1 and P2; C2 follows from C1. (Do you see why?) Also, S2 and S3 of Moore's argument, follow from C2 (when it is formulated using extended analyticity and extended entailment, and the transparency of meaning is assumed). A weakened version of S4 that limits itself to the claim that theses about goodness cannot be *proven* from claims not mentioning goodness might plausibly be taken to be established on the basis of S3. Thus, the weight of Moore's argument rests on P1 and P2.

Does Moore accept the premises? The case that he does comes from a certain interpretation of the Open Question argument, where he insists that the question expressed by [Granted that  $\alpha$  is  $D$ , is  $\alpha$  good?] is not the same as that expressed by [Granted that  $\alpha$  is  $D$ , is  $\alpha$   $D$ ?] or [Granted that  $\alpha$  is good, is  $\alpha$  good?]. Presumably he thinks that at least one of these questions--[Granted that  $\alpha$  is  $D$ , is  $\alpha$  good?] and [Granted that  $\alpha$  is good, is  $\alpha$   $D$ ?] must be open. But what does he mean by "open"? He says that the mere fact that "*we understand very well what is meant by doubting whether everything we desire to desire is good, shows that we have two different notions before our minds.*" This suggests that whereas it is *unimaginable* that anyone might doubt whether everything that is good is good, or whether everything that we desire to desire is what we desire to desire, it is quite definitely *imaginable* that someone might doubt whether everything we desire to desire is good. More generally, he seems to suggest that it is always imaginable that someone might doubt *whether everything that is  $D$  is good*.

Were 'good' truly definable--in the way that 'brother' is definable as 'male sibling' or 'square' is definable as 'rectangle with equal sides'--Moore would take this not be so. No one, he seems to think, could doubt what is expressed by 'Everything that is a square is a rectangle with equal sides' because that is just the proposition that everything that is a square is a square. Someone might, of course, be unsure whether the *sentence* 'Everything that is a square is a rectangle with equal sides was true'; presumably, however, Moore would maintain that this might happen only if the person was not a fully competent speaker of English, and failed to truly understand the sentence.

Thus, Moore may well have thought that *Granted that  $\alpha$  is  $D$ , is  $\alpha$  good?* always expresses an open question, in the sense that it is possible to understand the sentence, and entertain the question it expresses, without realizing that the answer to it is 'yes' (if indeed that is the answer). If Moore was correct in thinking that, it is enough to establish P2, and, indirectly, C1 and C2.

The only remaining interpretive issue is whether he would have accepted the characterization of the interrogative sentences (ii), (iii) and (iv) of P1 as *not* expressing "open questions" in the sense in which (i) supposedly does. I think he would have. He often speaks as if questions about goodness are

substantial and open-ended in ways that trivial questions like those expressed by (ii), (iii), and (iv) are not. *If there were no genuine contrast, then his supposedly far-reaching conclusions S3 and S4 would vanish, or be drained of significance.* Since he views them as of the highest importance, he would, I think, accept P1 and P2. Quite a few other philosophers would, I think, agree.

This, I believe, is the strongest reasonable reconstruction we can give of Moore's argument for his main meta-ethical conclusion about goodness. This argument is a combination of the expanded version of the open question argument, the weakened version of S4 that limits itself to ruling out *proofs* about goodness, plus the interpretation of steps S2 and S3 as involving extended analyticity and extended entailment (plus the transparency of meaning). However, as plausible as the resulting position may have seemed to some, Moore didn't *establish* that there is a clear and definite contrast between questions like those expressed by (ii), (iii), and (iv) of P1, on the one hand, and questions about goodness expressed by (i) on the other. Moreover, the claim that there is such a contrast for *all relevant D* is far from obvious. At best, we might regard it as an interesting, not altogether implausible, and historically influential conjecture.