13

What’s Right with the Open Question Argument*

Susana Nuccetelli & Gary Seay

Ethics . . . [is] partly analysis of what’s meant by ‘good’, ‘ought’, ‘right’, ‘wrong’, ‘valuable’, etc. And if certain analyses of these are right, then other ethical propositions, ones which aren’t analytic, wouldn’t be philosophical at all, but belong to psychology, sociology, and the theory of evolution.

G. E. Moore, Lectures 1933–34, 196

Naturalism: the recognition that it is within science itself, and not in some prior philosophy, that reality is to be identified and described.

W. V. Quine, *Theories and Things*, 21

I

The substantive issues of moral ontology raised by G. E. Moore’s Open Question Argument (OQA) are bound up with difficult problems of knowledge, language, and philosophical analysis. Offered as a refutation of ethical naturalism, the OQA has been thought to founder on some such problems, as can be seen from the long and varied history of attempted rebuttals it has inspired.¹ In spite of widespread

---

* We wish to thank Stephen Schiffer, Gary Ostertag, Ken Williford, and Fritz McDonald for helpful suggestions that have made this a better paper than it would otherwise have been.

¹ The version of the argument most prominent in chapter 1 of *Principia moralia* has come under fire both from ethical naturalists and non-naturalists, who have criticized it on several counts—including that it is
objections, however, the argument has remained a central point of contention in the naturalism/non-naturalism debate, rivaled in longevity among Moore’s most controversial arguments not even by his 1939 ‘Proof of an External World’. When an argument is as resilient as that, we may suspect that there is something to be learned from trying to determine the source of its enduring appeal, even though much might still be wrong with its original formulation. Here we propose a sympathetic reading of the OQA that neither endorses Moore’s moral ontology nor focuses on exegesis.² Our qualified defense distinguishes two arguments, OQA₁ and OQA₂, roughly according to the type of reductive ethical naturalism each attempts to undermine. OQA₁ targets any such program grounded in a semantic thesis, OQA₂ any such program grounded in a metaphysical thesis. This distinction is consistent with Moore’s own reasoning against what he thought of, in broad terms, as naturalism in ethics, which also points to two independent, though compatible, arguments. Although OQA₂ might at first appear more fundamental, here we devote most of our efforts to the other argument, OQA₁, whose scope is often underestimated (cf. Pigden, this volume). If we are right, OQA₁ amounts to a refutation of semantic reductive naturalism, a family of doctrines that will be shown to be of greater metaethical significance than sometimes thought. Given OQA₁, OQA₂ can be made out as an independent argument to the effect that there may in fact be no good grounds for metaphysical varieties of reductive naturalism.³

II

One of Moore’s goals with his OQA was to show that the value predicate ‘good’ is indefinable. Clearly, if ‘good’ were indefinable, then a fortiori it could not be defined in terms of any descriptive predicate standing for a purely natural or metaphysical property, and any version of naturalism that rested on such definitions would therefore be false. Yet the indefinability of ‘good’ would have no tendency to undermine all doctrines of ethical naturalists—though Moore himself seems to have thought that it did disprove them as well. As we shall see later, there is a certain casting of the invalid, that it leads to the paradox of analysis, that it presupposes an inflated moral ontology, and that it begs the question.

² Exegetical readings sympathetic to the open question argument can be found in Sylvester (1990) and Ball (1988). For more critical treatments, see Hutchinson (2001), Baldwin (1990), and White (1958).
³ The achievement of Moore’s OQA may be largely that it points to the normativity moral agents associate with evaluative concepts, as recently argued by Scanlon (1998); Stratton-Lake (2002); and Darwall (2006), among others. But we believe that substantiating any such claim requires that we first undertake a charitable construal of Moore’s OQA, whether this is reconstructed as deductive or inductive. Cf. Shafer-Landau (2003); Strandberg (2004); Ball (1991); and Snare (1975).
What’s Right with the Open Question Argument / 263

Moorean argument that does refute a semantic brand of ethical naturalism. This relatively modest conclusion, however, is not an obvious consequence of Moore’s original argument, which can be reconstructed in a first approximation as follows:

OQA

1. If a certain variety of ethical naturalism is true, then a value predicate such as ‘good’ is analyzable into a purely descriptive predicate such as ‘pleasure maximizing’.

2. If ‘good’ is analyzable into ‘pleasure maximizing’, then the question, ‘Granted a is pleasure maximizing, but is it good?’ is closed.

3. But the question, ‘Granted a is pleasure maximizing, but is it good?’ is open.

Therefore,

4. ‘Good’ is not analyzable into ‘pleasure maximizing’.

5. Steps (1) through (4) can be iterated for each attempted naturalistic analysis of value predicates into purely descriptive predicates.

Therefore,

6. No value predicates are analyzable into purely descriptive predicates.

Therefore,

7. Ethical naturalism is false.

A number of arguments may result from this OQA, depending on how some unclear terms are cashed out. Note, however, that no such argument could have force against non-reductive naturalism in ethics: that is, against a doctrine holding, for example, that evaluative and normative properties supervene on, but are not reducible to, purely descriptive properties. Of course, Moore himself allowed that value properties do (as we would now say) supervene on natural properties. Thus the logical space of theories vulnerable to OQA is limited to reductive theses of ethical naturalists such as

Semantic Reductive Naturalism (SRN)
Value predicates are analyzable without evaluative or normative remainder into purely descriptive predicates.

Metaphysical Reductive Naturalism (MRN)
Value properties are analyzable without evaluative or normative remainder into purely descriptive properties.

A common motivation for proponents of either of these theses seems to be

Epistemic Reductive Naturalism (ERN)
Ethical inquiry into value properties needs no other justification beyond that attainable through the standard methods of science.
Views such as ERN were rejected by Moore, who rightly saw them as challenging the status of ethics as a part of philosophy with a subject-matter and method entirely independent of the natural and social sciences. Such views, which do not entail the repudiation of ethics but rather its assimilation to the sciences, usually fuel theses along the lines of MRN or SRN. Of the two, only MRN entails that value properties are nothing over and above natural properties. SRN could not have that consequence unless there were nothing else to the content of the relevant predicates than the properties they denote. Since that would be, in effect, only a highly plausible ‘Fido’-Fido theory for those predicates, therefore when more is allowed into their content, SRN does not entail MRN. It follows that no OQA that defeats SRN could refute MRN.

The two revised versions of OQA that we shall present here are in part the products of distinguishing these two theses of reductive ethical naturalism in premise (1) of the above argument. OQA\(_1\) starts out with a premise to the effect that, given SRN, a predicate such as ‘good’ is analyzable (in a way to be spelled out presently) without evaluative or normative remainder into a predicate such as ‘pleasure maximizing’. By deploying a Moorean strategy against that thesis, the argument demonstrates that no such analysis is possible. This, together with a premise along the lines of (5) above, entails that no evaluative or normative predicate is analyzable into a purely descriptive predicate or set of predicates—which amounts to demonstrating the falsity of SRN. (As sometimes noted, a similar Moorean strategy could also be deployed to show that value predicates are not analyzable into metaphysical predicates such as ‘follows the will of God’ or ‘creates positive karma.’)

OQA is also an argument against MRN, to the effect that no value property is either identical or reducible to a purely descriptive property or set of properties. A major problem facing the argument for this conclusion is its alleged commitment to a ‘Fido’-Fido theory of general terms/concepts. Perhaps it is Moore’s own unclarity on the semantic features of general terms that fuels such an objection.⁴ But, as will be shown, there is a related argument, OQA\(_2\), that doesn’t rest on that problematic semantic theory for predicates. On our view, support for OQA\(_1\) stems from the reasons for OQA\(_1\)’s success—together with the difficulties facing metaphysical varieties of reductive naturalism in accommodating widely accepted intuitions about the modal status of identity statements entailed by their preferred analyses. But before we can examine these arguments, some further points of clarification are in order about Moore’s OQA above.

⁴ Although Moore seems to have held, at different times, more than one theory of meaning, there is no evidence that he was a ‘Fido’-Fido theorist about predicates. See, e.g., ‘Classes and Incomplete Symbols’, a selection from a course of lectures given in 1925–6 and later included in Moore (1966). According to Sylvester (1990), and White (1958), at the time of *Principia*, Moore endorsed a concept theory for predicates: namely, the view that a predicate stands for a concept. Cf. Baldwin (1990).
What’s Right with the Open Question Argument / 265

First, in OQA, questions of the form, ‘Granted a is X, but is it Y?’ are offered in order to tease out intuitions about certain equivalences. Let’s call any such question ‘Moorean’ and note that it would admit of two readings, depending on whether we take the intuitions it attempts to elicit to concern the equivalence of predicates or that of properties. For, clearly, although it might appear that a question of that sort concerns only the equivalence of the latter, it could also involve that of the former. To find out about a certain conceptual identity, surely one may unambiguously ask, for example, ‘Are concepts “X” and “Y” the same, semantically individuated?’ or ‘Do predicates “X” and “Y” have the same meaning?’ But an equally effective way to proceed would be by asking, ‘Is to be an X to be a Y?’ or ‘Is an X a Y?’ After all, imagine a sincere and competent speaker of English, who masters the predicate ‘attorney’ but is uncertain about its synonymity with ‘lawyer’. To find out whether these terms mean the same, that speaker may simply ask, ‘Is a lawyer an attorney?’ Yet any unequivocal OQA should make plain whether it is a conceptual question that is being raised.

However ambiguous the appeal to the Moorean question in the OQA above may be, it is clearly part of a reductio strategy against a certain doctrine. It constitutes, then, an exercise of philosophical analysis involving a thinker whose intuitions are sought for the issue at hand, similar in this respect to the Gettier counterexamples or Putnam’s Twin Earth. In all such cases, the intended thinker is oneself, provided one is a competent user of the terms involved and there is no reason to believe that one’s intuitions on the topic of concern are atypical and therefore irrelevant to the folk conception of them.⁵

A related unclarity undermining OQA concerns its appeal to analysis: which of the two philosophical methods standardly designated in that way, the conceptual or the factual, is at work in this argument? Needless to say, although methods of both sorts have been characteristic of reductive programs in philosophy, of the two, it is chiefly conceptual analysis that is associated with semantic theses such as analytical behaviorism in philosophy of mind. On the other hand, factual analysis is a strategy typical of metaphysical theses, such as the Strawsonian account of speech acts in terms of speaker-meaning in philosophy of language or Hume’s analysis of causation (Mackie 1973). Our reconstructions of Moore’s argument also accommodate these two different methods of analysis, since OQA₁ targets the conceptual analyses of semantic naturalists, and OQA₂, the factual analyses of metaphysical naturalists. Thus, where Moore talked equivocally of naturalistic definition in both cases, we talk of content equivalence—but only in connection with the naturalistic program targeted by OQA₁. On the other hand, the naturalistic program targeted by OQA₂ invokes identifications of value properties with purely descriptive properties.

⁵ For more on the purpose of analysis along the lines suggested here, see Mackie (1973); and Jackson (1998).
Regarding conceptual analysis, Moore sometimes denied emphatically that it concerned linguistic expressions, noting that a word may stand for different concepts. He often insisted that his interest was not merely in the words of a public language but in the objects of thought—"the notions themselves as they are present to the mind" (for example, 1966, 159). But a proper treatment of this subject would require discussion of how words are individuated and of the type/token distinction—both topics that Moore seems not to have developed at all. Even so, he later acknowledged (1942) that he could hardly have made out his arguments without some appeal to linguistic categories, as his use of the word 'predicate' suggests. Here we accept the point first made by Vendler, that when speakers are sincere and competent, what they say is syntactically and semantically parallel to what they think. That is, we accept that, when speakers are sincere and competent, many of their sentence-size utterances convey the contents of their psychological attitudes, as their singular and general terms convey the concepts they entertain. Throughout, then, instances of the Moorean question itself, together with its building blocks, are considered linguistic or mental entities of the standard sorts.

The reductionist programs targeted by Moore’s argument involve analyses of predicates/terms/concepts (and properties) of two types: what we may call ‘value’, which are either evaluative or normative; and ‘purely descriptive’, which correspond to what Moore himself referred to as ‘natural’ predicates. In his early work (for example, 1922, 253 ff.), the first group comprises predicates for intrinsic value (‘good’, ‘beautiful’, etc.) as well as terms that have ‘a fixed relation’ to kinds of value that are intrinsic (‘right’, ‘what ought to be done’, etc.), the second group includes terms for natural properties,⁶ some that on Moore’s view are intrinsic to whatever possesses them (‘yellowness’) and some that aren’t (‘being better fitted to survive’). This, of course, falls short of providing a criterion for the distinction between natural and non-natural properties, a problem that Moore continued to wrestle with long after Principia (1922, 1942). On our view, an epistemic criterion might do the best job of drawing the distinction—for example:

---

⁶ Criteria for defining a ‘natural property’ are at best vague in Moore’s work. In Principia, they consist in any such property’s existing in time all by itself and being a ‘part’ of anything that it characterizes. According to the most adequate metaphysical construal that can be culled from his labors, a property counts as natural if and only if, for any object that possesses it, no description of that object could be complete without a reference to that property, unless it is entailed by one or more other properties listed in a complete description of that object. Yet, as many have pointed out, this falls short of settling the distinction Moore wishes to make between natural and non-natural properties. On the other hand, the distinction of intrinsic properties/intrinsic values was unsatisfactory from the outset. In ‘The Conception of Intrinsic Value’, he writes: ‘...intrinsic properties seem to describe the intrinsic nature of what possesses them in a sense in which predicates of value never do. If you could enumerate all the intrinsic properties a given thing possessed, you would have given a complete description of it, and would not need to mention any predicates of value it possessed; whereas no description of a given thing could be complete which omitted any intrinsic property’ (1922, 273).
Natural Property

A property counts as *purely natural* or *descriptive* if and only if any claim about whether something has it is defeasible, at least in principle, by empirical investigation alone.

A criterion of this sort seems not only to beg no questions, but to accommodate what naturalists often find most implausible about non-naturalist positions: namely, that they countenance objects, properties, and relations well beyond what could be revealed by even the most complete empirical science. At the same time, our criterion is consistent with Moore’s reasons for defending the autonomy of ethics and his related rejection of naturalistic reductions of value properties to psychological, sociological, or biological properties.

III

Freed of some of its baggage, the original OQA may now be recast as either OQA\textsubscript{1} or OQA\textsubscript{2}, depending on which program of reductive ethical naturalism the argument attempts to undermine, and also on the casting of the Moorean question and the sort of philosophical analysis at work in the targeted program. OQA\textsubscript{1} targets a program advancing a semantic thesis such as SRN above, now recast as holding that value terms are *conceptually* analyzable (in a sense to be determined later) into some purely descriptive terms. OQA\textsubscript{2} targets a program advancing a metaphysical thesis such as MRN, now recast as holding that value properties are *factually* analyzable into some purely descriptive properties. Unlike Moore’s OQA, however, neither of these arguments is undermined by equivocation. Since it is OQA\textsubscript{1} that provides initial support for the other argument, let’s begin by having a closer look at it.

**OQA\textsubscript{1}**

1. If semantic reductive naturalism is true, then tokens of one’s value terms and purely descriptive terms such as ‘good’ and ‘pleasure maximizing’ instantiate the same semantic type.
2. If one’s tokens of ‘good’ and ‘pleasure maximizing’ instantiate the same semantic type, then the question, ‘Do my thoughts that \textit{a is pleasure maximizing} and \textit{a is good} have the same content?’ is closed.
3. But the question, ‘Do my thoughts that \textit{a is pleasure maximizing} and \textit{a is good} have the same content?’ is open.

Therefore,

4. One’s tokens of ‘good’ and ‘pleasure maximizing’ do not instantiate the same semantic type.
Steps (1) through (4) can be iterated for each attempted conceptual analysis of a value term into purely descriptive terms.

Therefore,

Semantic reductive naturalism is false.

If OQA1’s deductive core, (1) through (4), is compelling, then the rest of the argument, which concludes with a generalization against the reductive program it targets, will be in the clear. We shall now show that it is. Crucial to OQA1’s deductive core are the premises that deploy a reductio strategy against the targeted thesis: (2) contends that, given the proposed naturalistic analysis, a related Moorean question would be closed; while (3) asserts that the question is open.

We’ll take up the Moorean question first. As noted above, its intended addressee is oneself provided that (a) in the vocabularies involved, one is a competent user of the predicates, both the evaluative or normative and the purely descriptive; and (b) one has no reason to think that one is an atypical thinker, likely to give answers unrepresentative of ordinary intuitions about the alleged content equivalence of those predicates. Suppose one is now seriously considering a question of that kind. It seems that any suitable answer would require looking into the two propositional contents one is entertaining not merely dispositionally but occurrently, and not only subconsciously but consciously. As is apparent in OQA1, the Moorean question therefore turns on a cogito-like judgment of content: one that consists in a present tense, self-attributed, judgment about whether one’s own propositional attitudes (for example, that one is judging that a is pleasure maximizing and that a is good) are content equivalent. To better capture that is precisely the sort of judgment prompted by the Moorean question, we have recast the question accordingly. We take any qualifying questions to be of the form

CMQ (Cogito-like Moorean Question)

‘Are my current judgments that D and V content equivalent?’

where ‘D’ and ‘V’ stand for a general term/concept in each of the aforementioned vocabularies. Since the term/concept ‘a’ is held constant, therefore other things are equal in the content-clauses of any CMQ. It follows that the comparative judgment of content teased out by the question boils down to determining whether general term/concepts ‘D’ and ‘V’ are content equivalent in a sense we’ll now spell out.

Suppose one is seriously entertaining a certain CMQ and meets conditions (a) and (b) above. A suitable answer would require that one make a comparative judgment of content equivalence by looking into that CMQ’s content-clauses: the that-clauses widely taken to count as primary vehicles for individuating the content of whatever propositional attitudes they are used to report. Contra Moore, it doesn’t really matter whether one entertains such that-clauses in thought only, or also in language (recall we are assuming Vendler’s point). The differences among the tokens of a
CMQ's that-clauses may concern the syntactic type and/or the semantic type they instantiate. Of course, any two such clauses that differ syntactically in virtue of being instances of two different syntactic types may none the less have the same content in virtue of instantiating the same semantic type. Naturally, any two that-clauses that instantiate the same semantic type are content equivalent. Thus in a CMQ's that-clauses, since other things are equal except for the tokens of certain general term/concepts ‘D’ and ‘V,’ determining content equivalence amounts to determining whether those tokens instantiate the same semantic type. Given OQA1’s premises (2) and (3), whenever one meets the conditions for being a suitable addressee of the relevant CMQ, one is a reliable judge of the putative content equivalence of the two that-clauses one is consciously and occurrently entertaining. That is, the premises require that one would reliably judge that the question is closed whenever a CMQ’s that-clauses are content equivalent, but that it is open whenever they are not content equivalent.

It is now clear that to undermine the analyses of semantical reductive naturalists, OQA1 presupposes that competent users of the concepts involved in such analyses can reliably assess claims of content equivalence by entertaining corresponding CMQs. But this boils down to assuming the transparency of mental content: after all, when presented with a CMQ, one is supposed to be able to make a selfascriptive comparative judgment of content right off, by looking into the contents of one’s own thoughts prompted by the question. Yet there is tension between transparency and allowing for unobvious analyses, and the latter seems required by the argument in order to avoid the so-called paradox of analysis: the problem that any conceptual analysis, if correct, must be trivial.⁷ On the one hand, comparative judgments of content triggered by CMQs presuppose the transparency of mental content. On the other, to avoid the Paradox of Analysis, proponents of the argument would have to countenance the possibility of unobvious analyses. And once they do that, the appeal to CMQs could have no force against semantic theses of reductive naturalism. It appears, then, that if there were such unobvious correct analyses, then premise (2) would be false. That a is good might after all be content equivalent with that a is pleasure maximizing, even when the question, ‘Are my current thoughts that a is pleasure maximizing and that a is good content equivalent?’ remained open for some competent users of the concepts involved. Since at least some such analyses could then be both correct and open for competent users, OQA1 would founder in light of the Paradox of Analysis.

Although it may be thought that proponents of OQA1 need not worry about a paradox that has already been resolved in more than one way (see, for example, Mackie 1973; Katz 2004), on our view, meeting this objection really amounts to being able to formulate a plausible thesis of the transparency of mental content, one that

⁷ We are indebted to Ken Williford for this suggestion.
could also be deployed to resolve the paradox of analysis. In fact, nothing stronger than the following is required by the argument:

Transparency of Mental Content Thesis (TMC)
Under normal circumstances, first-person comparative judgments of content teased out by CMQs amount to a priori knowledge—i.e., to beliefs that are (i) default true and (ii) warranted without empirical investigation.

Now clearly proponents of OQA1 are not committed to denying the possibility of unobvious analyses. To qualify for a priori knowledge, a true belief need not be obviously warranted, but rather warranted just by thinking.

At the same time, it is not difficult to see how TMC is weaker, and therefore more plausible than an unqualified Cartesian thesis to the effect that one is always in a position to know whether any two self-attributed thoughts have the same content. For one thing, TMC’s scope is limited to cogito-like Moorean questions of the sort featured in OQA1’s premises (2) and (3). Such questions seek to refute attempted naturalistic analyses of value terms/concepts into purely descriptive terms/concepts. Since other things are equal in the Moorean question, TMC boils down to claiming a plausible privilege for one’s assessment of the content equivalence of certain concepts one is competently entertaining, which belong to those two types of term/concept. This thesis leaves open the possibility that, for other types no such transparency holds. Thus, TMC falls short of entailing that one is always in a position to know whether any two concepts one is entertaining are content equivalent. Furthermore, TMC requires ‘normal circumstances’, which are those meeting the specified conditions (a) and (b) for being an eligible addressee of the question—namely, that one is a competent and typical user of the concepts whose content equivalence one is assessing. Thus, TMC commits to neither the infallibility nor the incorrigibility of one’s comparative judgment prompted by a CMQ. In addition, a case could be made that ‘default true’, a weak, truth-related immunity required by TMC, is a privilege conferred by the first-person grounds available for judgments teased out by CMQs. After all, they amount to cogito-like thoughts whose privileged-truth status is widely accepted, given first-person authority. As we construe their truth status, although they fall short of being immune to falsity or corrections by others, they are predominantly true or true in the absence of evidence to the contrary—a qualification needed to accommodate those rare occasions where observation and the testimony of others may be necessary, for example, in order to disqualify oneself as a suitable addressee of a certain CMQ. But if generally true and warranted non-empirically (i.e., just by thinking) such judgments constitute a priori knowledge—at least in the sense of amounting to true beliefs warranted without empirical investigation. In fact, there is now dialectical space to contend that OQA1’s premises are all a priori in exactly this sense. If so, then since the argument is valid and non-question begging, it follows that such an epistemic status
What’s Right with the Open Question Argument / 271

may transmit to its conclusions. But there is another condition, to which we now
turn, that must also be met.

IV

OQA1’s conclusion (4) could inherit an a priori status from the argument’s premises,
provided that if one can know a priori both that \( p \) and that \( p \) entails \( q \), then one can
also know a priori that \( q \). Yet a principle sanctioning this is weaker than other closure
principles, and therefore at least prima facie very plausible. Given this principle,
since OQA1 is a non-question-begging valid argument, the apriority of its conclusion
(4) depends entirely on that of supporting premises. Moreover, if (4) is knowable a
priori, then since it is clear that premise (5) could be known by philosophical reflection
alone, conclusion (6) would qualify as a priori knowledge, too. Thus, support for
OQA1 amounts to showing that premises (1) through (3) can be considered by default
true and warranted non-empirically.

Arguably, one could come to know a priori the implication in premise (1). For it
appears to be both a true consequence of a semantic thesis such as SRN, and warranted
by philosophical theorizing with no empirical investigation necessary. Yet premise
(1) is sure to face one objection: has any naturalist ever held a thesis along such lines?
Is semantic reductive naturalism a position of any weight in ethics? Skepticism on
this count has historically taken the form of doubt about whether anyone really did
commit the ‘naturalistic fallacy’. Moore, of course, held an inflationary view of the
universe of offenders on this score.⁸ Yet the list of reductive naturalists willing to
advance semantic theses along SRN’s lines is by no means short. Bentham famously
defined ‘right action’ as ‘an action that is conformable to the principle of utility,’
insisting that the words ‘right’, ‘wrong’, and ‘ought’ have meaning only when interpreted
in terms of that principle (1789, ch. 1, §10). Westermarck, with whom Moore takes
issue in ‘The Nature of Moral Philosophy’ (1922, 332), appears to have held that calling
an action ‘right’ or ‘wrong’ is equivalent to saying that it produces in us certain
feelings of approbation or disapprobation. For clear examples of semantic naturalists,
however, Moore needn’t have looked further than his own contemporaries: R. B.
Perry famously wrote that ‘\( x \) is valuable = interest is taken in \( x \)’ (1926); and F. C. Sharp
presented the following definitions: ‘‘good’ means ‘desired upon reflection’’ and

⁸ Sidgwick appears Moore’s immediate predecessor in pointing out what might constitute one of the
so-called informal fallacies. As often noted, however, there is no agreement on precisely what these are.
Moore himself claimed to have found the naturalistic fallacy in Aristotle, the Stoics, Spinoza, Rousseau,
Bentham, Kant, Mill, Spencer, and Green (Baldwin(1990), 69). For a more conservative list, see Prior
(1949, 104–7).
"right" means "desired when looked at from an impersonal point of view" (1928; cited by Frankena 1963, 81). Nor are semantic naturalists absent from the present-day philosophical stage, as can be seen from the work of Frank Jackson, according to whom, in some cases, once a situation has been well described in nonmoral terms, certain ethical sentences will follow a priori from that description—or, as he puts it, ethical sentences are 'a priori equivalent to and analyzable in terms of nonmoral ones' (2003, 558). Clearly, semantic reductive naturalism is a view in moral philosophy that has had prominent defenders and remains currently a live option.

OQA1's premise (1), then, does not render the argument vulnerable to a straw-man objection.

We can now proceed to the next step of showing the apriority of the conditional in its premise (2). If warranted a priori and true, that premise also would qualify as a priori knowledge. That warrant for premise (2) is a priori follows from its being grounded entirely on philosophical theorizing about the cognitive value of a Moorean question involving concepts that would be content equivalent provided the implication of semantic naturalism in premise (1) is true. No warrant of that sort requires empirical investigation. For premise (2) to be true, it should be true that, given its antecedent, the Moorean question in its consequent is not open. As argued above, suitable assessments of the cognitive value of any such question involve first-person comparative judgments of content, which are a priori warranted and default true. Given that contention, not only does premise (2) come out a priori, but (3) does as well, a premise entirely based on a first-person judgment of content teased out by the Moorean question.

Further support for the contention that answers to such questions are a priori can be found in Moore’s own ways of cashing out the cognitive values they aim at eliciting. We may spell out one such way as follows:

\[ \text{OQ} \quad \text{For any CMQ and for any competent user of the predicates involved in it, the question is closed or settled if and only if those predicates are content equivalent, and otherwise it is open or unsettled} \]

Under a certain reading of 'settled'/‘unsettled’, determining whether a CMQ is open appears to hinge on whether or not any suitable answer to that question would be controversial, which might require evidence of the sort provided by opinion polls and other procedures of empirical investigation. Needless to say, if judgments about sameness and difference in content prompted by CMQs were at least in part warranted only by evidence, then the above argument’s premises featuring a question of that sort would fall short of a priori knowledge. Yet ‘settled’/‘unsettled’ also admit of a reading

---

⁹ Bertrand Russell and David Lewis may also be added to the list of semantic reductive naturalists. For arguments to the effect that they both attempted naturalistic analyses of the sort rejected by Moore, see Pigden (this volume).
What’s Right with the Open Question Argument / 273

more friendly to the apriority of answers properly teased out by CMQs—as captured by another common casting of their two possible values, ‘significant’/’non-significant’. This construal is consistent with our contention that proper answers to CMQs consist entirely in first-person judgments about the content equivalence of two self-attributed propositional attitudes prompted by any such question. If so, then those answers are default true and non-empirically warranted. They therefore qualify as a priori knowledge.

In invoking these privileges for cogito-like thoughts, we are, of course, not alone. As the current literature makes clear, much can be learned about the plausibility of such privileges from philosophers whose theories appear incompatible with them. Rightly fearing that any failure to accommodate these privileges would count as a reductio of their own theories, their standard reaction has been to make provisions for accommodation. Cases in point are the content-externalist theories of Tyler Burge and Donald Davidson. Burge has long been at pains to demonstrate that his externalism is, in fact, compatible with what he calls ‘the epistemic specialness’ of self-knowledge, offering an account of its privileges more ambitious than the transparency of mental content defended here. He writes: ‘some present tense self-attributed thoughts about propositional attitude content and type are self-referential, non-contextually self-verifying, and directly and non-empirically warranted’ (1996, 92–4). Likewise, Davidson proposed, though perhaps less emphatically, similar privileges for self-knowledge judgments—for example, when he argues

... [although] an interpreter must, if he is to get things right, look to relations between the mind he is interpreting and its environment, this does not prejudice the self-knowledge of the knower... I say 'I believe the Koh-i-noor diamond is a crown jewel'... And suppose, as is the case, that I know what the words I have just uttered mean, and that I am making a sincere assertion. Finally, let us suppose that you and I agree on these points... From these suppositions it follows that I know what I believe, but it does not follow that you know what I believe. The reason is simple: you may not know what I mean. Your knowledge of what my words mean has to be based on evidence and inference... it does not make sense to suppose I am generally mistaken about what my words mean; the presumption that I am not generally mistaken about what I mean is essential to my having a language—to my being interpretable at all. (1991, 212)

Efforts of these sorts indirectly support our presumption that the cogito-like judgements elicited by Moorean questions are a priori. Clearly, the view proposed here is consistent with Davidson’s claim that the assumption that one is not generally mistaken in evaluating the content of one’s own propositional attitudes is necessary for our having a language. Furthermore, it is consistent with arguments by Burge (1998) and (more especially) Paul Boghossian (1994) to vindicate the transparency of mental content, on the ground that it is a necessary condition of our being critical thinkers and rational at all. It should also be noted that the truth-related immunity we ascribe to judgments elicited
by Moorean questions is weaker than Burge’s claim of non-contextual self-verifiability cited above—though, like him, we also take *cogito*-like thoughts, of which Moorean questions are instances, to be warranted non-empirically. But when it comes to the truth-status of such thoughts, our more modest claim is closer to Davidson’s: i.e., that whenever one meets the conditions for being able to evaluate a certain CMQ, then if one considers that the question is open (or closed), that judgment is default true, a status that could yet be overridden by countervailing evidence. In any case, if default true and non-empirically warranted, that judgment is then a priori. Knowledge of this sort plainly falls beyond what can ordinarily be challenged by others, whether or not one is able to articulate that that is its normative force.

Given the presumption that qualifying answers to CMQs are privileged in these ways, OQA₁’s premises (2) and (3) are now established as clearly a priori. Since premises (1) and (3) have already been shown to have that epistemic status too, and OQA₁ is a non-question-begging valid argument presupposing a plausible closure principle, it follows that its conclusions (4) and (6) can be known a priori.

V

Our defense of OQA₁ is incomplete, however, for we still must dispose of some possible objections. Prominent among them is one likely to be raised by direct-reference semanticists—and, more generally, by any content externalists wishing to reject the transparency of mental content, even in the very modest version offered by our TMC above. Objections on these counts would likely invoke Pierre-style scenarios (Kripke 1979) thought to raise a paradox, or at least a puzzle, for belief ascription and rationality. Although we need not rehearse here the details of that case, recall that Pierre, a bilingual speaker of French and English, each independently learned, has two occurrent, conscious beliefs that he would sincerely and competently express by uttering ‘Londres est jolie’ and ‘London is not pretty’. By hypothesis unaware that the propositional contents thus expressed are contradictory, Pierre appears altogether rational. From Pierre’s own perspective, the only possible epistemic warrant for a knowledgeable comparative judgment of those contents would seem empirical. Furthermore, once we grant that it is the extension of a term that in part determines its content, we appear in a position to run parallel scenarios for propositions containing some general terms/concepts such as ‘cat’, ‘water’, and the like.

Here is one: imagine Pierre, now resettled in Wisconsin, periodically undertaking hunting expeditions to nearby states. Suppose that in the presence of a certain animal—say in Minnesota—Pierre has thoughts he would competently and sincerely
What's Right with the Open Question Argument / 275

express by uttering 'That's a woodchuck'; while in Illinois he has thoughts he would also competently and sincerely express by uttering 'That's a groundhog'. Ex hypothesi, Pierre fails to notice that groundhogs are woodchucks. To test whether he is able to put two and two together, we meet Pierre in Minnesota, at a moment when he has just competently and sincerely reported the presence of such an animal. We then pose this quasi-Moorean question: 'Granted that is a woodchuck, but is it a groundhog?' Pierre must now make a comparative judgment of content about the two propositional attitudes he is occurrently and consciously entertaining. By hypothesis, however, he is in no position to produce a knowledgeable first-person judgment of that sort.¹⁰ Thus, the transparency of mental content would fail in this scenario. Needless to say, it would also fail in relevantly similar scenarios run for pairs of natural-kind terms/concepts and even of non-natural-kind ones such as 'hotdog'/frankfurter', 'ketchup'/catsup', 'computer'/ordinateur', 'pencil'/crayon', and so on. Since, by hypothesis, Pierre would meet the conditions for being a suitable addressee of quasi-Moorean questions in those cases, his failed comparative judgments of content would seem to lead us straight to the conclusion that no transparency thesis can succeed—and thus neither can OQA₁, our Moorean argument presupposing one such thesis.

Yet more than this will be needed to defeat OQA₁. For, clearly, that would require that Pierre-style scenarios be compelling for propositional attitude contents of the sort relevant to the naturalistic analyses targeted by that argument, which trade on alleged content equivalences between value terms and purely descriptive terms. But could any intuition about the failure of transparency fueled by Pierre-style scenarios also be plausible in the case of propositional-attitude contents featuring predicates relevant to naturalistic analyses? It may be shown that there is no compelling reason in the offing for thinking that it could. In fact, there is considerable agreement that such scenarios are not compelling across the board for general terms/concepts of any type whatsoever. As Kripke (1979, 264–5) notes:

At the moment, at least, it seems to me that Pierre, if he learns English and French separately, without learning a translation manual between them, must conclude, if he reflects enough, that 'doctor' and 'medicin' and 'heureux' and 'happy' are synonymous, or at any rate, coextensive; any potential paradox of the present kind [Pierre-style scenarios] for these word pairs is thus blocked.

Here Kripke commits himself to upholding the transparency of content for terms of the sort he cites: competent users of such pairs ought to be able to tell by reflection alone whether the predicates involved are content equivalent. Fregeans would, of course, agree that in these cases (and many others) the transparency of the entertained contents remains unchallenged. Even direct-reference theorists such as Putnam (1970)

¹⁰ This case also shows that in Pierre-style scenarios nothing hinges on translation.
would concede Kripke’s claim, offering among salient counter-examples to Pierre-style scenarios ‘bachelor’, ‘sister’, ‘triangle’, and other ‘one-criterion’ general terms, whose definability or analyticity has never been an issue for direct-reference semanticists and content externalists. It appears, then, that no matter which of the available semantic theories will turn out to be the end-of-the-day theory of content for general terms/concepts, there is current consensus that the transparency of mental content must be assumed for a great number of such terms/concepts.

We now seem to have a principled way of distinguishing propositional contents for which transparency could fail from those for which it couldn’t. Such a criterion is simply vulnerability to Pierre-style scenarios. This would not only be accepted by Fregeans but also beg no question against reasonable direct-reference semanticists and content externalists who are none the less willing to jettison an unqualified thesis of the transparency of mental content. Clearly, they could still maintain that transparency fails for externally determined propositional-attitude contents while it obtains for contents that are internally individuated. The former are of the sort vulnerable to Pierre-style scenarios; the latter are not.

To complete our defense of OQA\textsubscript{1}, we next need to determine how the predicates at issue in naturalistic analyses compare with the predicates so far considered. Are they similar to Pierre-style-scenario-vulnerable general terms/concepts such as ‘woodchuck’ and ‘water’—or to Pierre-style-scenario-resistant ones such as ‘happy’ and ‘bachelor’? Clearly, what works for ‘happy’ also works for ‘pleasant’, and many other predicates featured on the right-hand side of naturalistic analyses. But once we agree that a term such as ‘happy’ or ‘pleasant’ resists Pierre-style scenarios, we are committed to holding that any value term that is content equivalent with it would likewise resist such scenarios.

That is, given the consensus that those scenarios are blocked in cases involving predicates such as ‘happy’ and ‘bachelor’, the alleged problem facing OQA\textsubscript{1} is resolved. For to say that ‘happy’/‘heureux’, ‘bachelor’/‘unmarried man’, and the like are Pierre-style-scenario resistant is to say that the conceptual equivalence, or at least co-extensiveness, among the predicates in each pair is available upon reflection to any competent user of the pair. By analogy, Pierre-style scenarios are blocked in cases involving predicates in the two vocabularies at issue in the analyses of semantic naturalists. Although the purely descriptive predicate proposed on the right-hand side on any such analysis may, in fact, be a simple one such as ‘happiness’ or ‘pleasure maximizing’, the analysis could also feature a very complex predicate—perhaps of a disjunctive sort.\footnote{An analytical descriptivist of Jackson’s persuasion may object to our casting of naturalistic analyses on the ground that they appear too simplistic. After all, they could feature on their right-hand side very complex disjunctions of purely descriptive terms. On this view, adequate naturalistic analyses} Either way, any correct claim about the content equivalence of...
the terms featured in a naturalistic analysis ought to be available a priori to any rational thinker who is a competent user of those terms. Imagine Peter, a monolingual speaker of English who is both rational and a competent user of ‘good’ and ‘pleasure-maximizing.’ Peter is now considering a certain semantic naturalist claim to the effect that tokens of those terms instantiate the same semantic type. By analogy with the cases above (‘happy’/‘heureux’, ‘bachelor’/‘unmarried man’, etc.), we maintain that Peter ought to be able to settle this question just by thinking. If so, OQA1 is in the clear, since predicates such as ‘good’ and ‘pleasure-maximizing’ are, after all, Pierre-style-scenario-resistant. TMC, the transparency thesis fueling that argument, remains unchallenged.

Yet it might still be thought that this line of reasoning is vulnerable to Frankena’s (1939) objection against Moore’s own argument: i.e., that it begs the question against the reductionist program of semantic naturalists. But surely, just who is doing the question-begging is determined by where the burden of proof lies. Consider Peter again, who now is an avid fan of television’s ‘Jerry Springer Show’, even though he does not really like the fact that he enjoys it. Although making fun of vulnerable people has considerable entertainment value, Peter thinks, it is achieved at a morally questionable price. Peter now sincerely asserts ‘‘The Jerry Springer Show’ maximizes pleasure, but is not good.’ Semantic naturalists willing to take Peter’s tokens of ‘good’ and ‘pleasure-maximizing’ to be content equivalent would face a paradox. For they would be committed to denying at least one of the following: (1) that Peter is a competent speaker of English; (2) that he is sincere; or (3) that he is rational. Given that in common cases of this sort all three assumptions seem independently plausible, and that other equally common cases could be offered for equivalence claims involving different pairs of general terms in the two vocabularies of naturalistic analyses, there is room to argue that it is the naturalist who has the burden of proof. A mere rejection of the assumptions commonly made in light of such cases would simply beg the question. Furthermore, it counts against semantic reductive naturalism that, if this are of the form, ‘A is right if and only if it has whatever [descriptive] property it is that plays the rightness role in mature folk morality’(1998, 151). This proposal is not, however, beyond dispute. For one thing, it attempts to ground the relevant analyses on both folk morality and meaning analysis. But there is room to ask, do such analyses depend on the posteriori common-sense platitudes of a mature folk morality? Or are they rather a priori semantic equivalences between non-rigid descriptions? These seem to pull in opposite directions. Of course, the appeal to folk morality requires a complex analysans in order to preserve the view from the objection that the platitudes of common sense are often false. At the same time, since this theorist’s analyses purport to capture meaning equivalences in the two relevant vocabularies, those analyses ought to square well with the intuitions of competent users of the terms/sentences involved. But currently, they don’t. Maybe the theorist is right in claiming that there will be a convergence in a mature folk morality. Yet why should we accept this futuristic claim, given a plausible version of the open question argument? The burden of proof seems to be on the analytical descriptivist.
thesis were correct, then Peter’s assertion should be as paradoxical as those of Kripke’s Pierre, since belief ascription in Peter’s case would also have to be in tension with ascribing rationality to him. But belief ascription here seems tension-free: Peter simply believes that ‘The Jerry Springer Show’ maximizes pleasure but that it is not good.

In the absence of a convincing response by semantic naturalists, we may conclude that if tokens of ‘good’ and ‘pleasure-maximizing’ were content equivalent, then a competent thinker of both predicates who reflected long enough ought to be able to tell whether they are content equivalent—or, at least, coextensive. The Pierre-scenario objection to OQA₁ (our Moorean argument against the semantic theses of some reductive naturalists) is thus blocked.

VI

Although OQA₁ meets all conditions needed to defeat semantic theses of reductive naturalists—namely, being a non-question-begging argument that can transmit by entailment the apriority of premises to the conclusion that no such thesis is true—the argument falls short of refuting metaphysical theses of reductive naturalists.¹² Yet OQA₁ provides non-conclusive reasons for a view on the semantic properties of the relevant predicates that support OQA₂, our argument against metaphysical theses of reductive naturalists. OQA₂ starts out by maintaining that, given metaphysical reductive naturalism (MRN), predicates in the two relevant vocabularies—say, ‘good’ and ‘pleasure-maximizing’—must be coextensive. The reasons brought about by our defense of OQA₁ suggest that such predicates qualify for being Pierre-style scenario resistant. This entails that any alleged co-extensiveness between them must be transparent to competent users that consider those predicates. But this does not seem to be the case.

In order to defend the thesis that terms such as ‘good’ and ‘pleasure-maximizing’ are coextensive even when their coextensiveness fails to be transparent, the ethical naturalist holding MRN must argue that terms of these two sorts are similar in the their semantic properties to pairs of coextensive natural-kind terms such as ‘woodchuck’ and ‘groundhog’, or ‘water’ and ‘H₂O’. Such a claim would, however, commit that ethical naturalist to an implausible view of identity statements entailed by his attempted factual analyses—since, arguably, those statements should then have the same modal status as identifications in science involving natural-kind terms.

¹² The objection that the open question argument fails as a valid argument against metaphysical varieties of ethical naturalism was made independently by Harman (1977); and Putnam (1981). See Brink (1989, 2001) for a view on the semantic properties of the relevant predicates that is vulnerable to the objections raised in this section.
in both their metaphysical and epistemic features. That is, if true, they should be necessarily so. At the same time, they should be epistemically contingent: i.e., knowable only \textit{a posteriori}. But it is implausible that the identity statements of this ethical naturalist do have such a modal status. Here is the argument:

1. Given MRN, identity statements involving a value term and a purely descriptive term have the modal profile of theoretical identifications involving two natural-kind terms.

2. But such identity statements lack the modal profile of theoretical identifications involving two natural-kind terms.

Therefore,

3. MRN is false.

The theoretical identifications in science relevant to this argument are those containing two coextensional natural-kind terms. Whether or not terms of this sort turn out to be rigid designators is a controversial matter that falls beyond our concern here. For the purpose at hand, it suffices to note that they do seem \textit{essentialist} in the sense that they appear to pick out structural properties of the substances and species falling within their extensions—often counted as hidden essences since they lie beyond the phenomenal qualities of those substances and species (Soames 2002). That a certain general term qualifies for being essentialist is shown by some well-known counterfactual scenarios such as Twin Earth cases. Given those cases, in any possible world where water exists, it is H$_2$O. Thus, the proposition expressed by instances of ‘Water is H$_2$O’ is not only true but necessarily so, in spite of any appearance of metaphysical contingency. Consistent with this intuition is the presumption that ‘Water might not have been H$_2$O’, ‘Woodchucks might not have been groundhogs’, and the like are self-contradictory. On the other hand, no such presumption arises when at least one of the featured general terms in the identity statement is not essentialist, as illustrated by ‘Water might not have being the stuff flowing in our lakes and rivers’. Furthermore, the scenarios invoked to show the necessity of theoretical identifications in science also explain away any appearance of contingency by suggesting that it would stem from the epistemic possibility of conceiving structurally different substances and species which none the less have exactly the same phenomenal qualities.

Now compare ‘Right is what maximizes pleasure’, ‘Good is what we desire to desire’, or any other more complex identity statement of the sort that would have to be true if certain factual analyses of reductive naturalists were correct. None of these would seem to have the modal profile of the theoretical identifications just discussed—for, if they had it, then there would be compelling counter-factual scenarios supporting this. But such scenarios are not forthcoming. It is not clearly self-contradictory to say that an action is right even though it fails to maximize pleasure, or good even
though it doesn’t coincide with what we desire to desire. Plainly, such judgments are by no means incoherent—as can be seen, for example, from the unresolved conflicts in normative ethics between the fundamental moral intuitions of consequentialism, on the one hand, and of deontology on the other. But this strongly suggests that the predicates involved in those judgments are not of the essentialist sort needed to make the ethical naturalist’s statements of identity relevantly analogous to theoretical identifications in science. If we are right, then there are good grounds to reject the reductive analyses of some metaphysical naturalists. The onus is now on them to show why we should accept a view that seems prima facie very implausible.

This objection generalizes, since at least in principle the same reasoning supporting that predicates such as ‘good’ and ‘pleasure maximizing’ are not coextensive could be offered for any other attempted analysis of value properties in terms of natural properties. It seems then that metaphysical reductive naturalism is faced with a reductio that runs as follows:

\[ OQA_2 \]

1. Given MRN, a certain value predicate such as ‘good’ and a descriptive predicate such as ‘pleasure maximizing’ are coextensive.

2. If ‘good’ and ‘pleasure maximizing’ are coextensive, then competent users of ‘good’ and ‘pleasure maximizing’ ought to be able to determine upon reflection that these predicates are coextensive, unless they are semantically analogous to natural-kind predicates.

3. But neither are competent users of ‘good’ and ‘pleasure maximizing’ able to determine upon reflection that those predicates are coextensive nor are such predicates semantically analogous to natural-kind predicates.

Therefore,

4. ‘Good’ and ‘pleasure maximizing’ are not coextensive.

5. Steps (1) through (3) can be iterated for each attempted factual analysis of a value predicate into purely descriptive predicates.

Therefore,

6. MRN is false.

VII

In the end, it is tempting to speculate about what Moore himself would have thought of our casting the dispute in these terms and trying to offer a qualified defense of his argument by reconstructing it in these two different ways. Although he was
optimistic in his early writings about the open question argument against naturalism, he later expressed disappointment. In ‘Is Goodness a Quality’, a 1932 paper read to the Aristotelian Society, he had this to say about his attempts to show the indefinability of ‘good’:

In Principia I asserted and proposed to prove that ‘good’ (and I think I sometimes, though perhaps not always, was using this word to mean the same as ‘worth having for its own sake’) was indefinable. But all the supposed proofs were certainly fallacious; they entirely failed to prove that ‘worth having for its own sake’ is indefinable. (Moore 1959, 98)

While it is true that his own OQA appears not to succeed as a proof of the indefinibility of ‘good’, Moore was plainly up to more than that in the first chapter of Principia. Relevant to our concern here have been two other claims that Moore-inspired arguments have been shown to support: the rejection of reductive naturalism; and the defense of the autonomy of ethics. Despite Moore’s own later judgment and those of many critics, we believe that properly construed open question arguments can succeed on these counts.

Works Cited

282 / Susana Nuccetelli & Gary Seay


Scanlon, T. M., What We Owe to Each Other (Cambridge, Mass.: Harvard University Press, 1998).


Queries in Chapter 13

Q1. please clarify author correction here