DESIRING TO DESIRE:
RUSSELL, LEWIS AND G.E MOORE[1].

By

Charles R. Pigden

1. Moore's warning

'I am the person Moore warned you against,' joked David Lewis[2] before reading his now famous paper, 'Dispositional Theories of Value', to the Aristotelian Society in 1989[3]. The alleged warning occurs in §13 of *Principia Ethica*[4], the crucial passage in which Moore expounds the Open Question Argument. 'To take, for instance one of the more plausible, because one of the more complicated of such [naturalistic] definitions, it may easily be thought, at first sight, that to be good may mean to be that which we desire to desire.' (*PE*: §13.) But plausible as it may be, Moore goes on to contend that this definition is false, since it is possible to wonder whether what we desire to desire is good. If 'good' *meant* 'what we desire to desire', the question 'Is what we desire to desire good?' would be a silly question since the answer would be very obvious – 'Yes'. Since the question is open and the answer is *not* obvious to every competent speaker, the definition cannot be correct.

---

[1] This paper is dedicated to the memory of David Lewis

[2] So I was told by an ear-witness, though Lewis himself had no recollection of the remark. However he certainly subscribed to the sentiment. 'The position defended is similar to the one that G. E. Moore chose as the target for his "naturalistic fallacy" argument.' See Lewis, David, *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press, 2000), 2.

[3] Lewis, David, 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society Supplementary Volume,* 63 (1989), 113-137. Reprinted in Lewis, David, *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press, 2000), 68-94. Henceforward DTV II with page references both to the original and to the reprint.

[4] Moore, G. E. *Principia Ethica*, revised edn., ed. Thomas Baldwin, (Cambridge: Cambridge University Press, 1993). Henceforward *PE*. On the whole I give section rather than page references to accommodate those with earlier editions.

Now Moore's 'warning' certainly *applies* to Lewis since he develops an analysis of value as what we are ideally disposed to desire to desire (an analysis which would fall foul of the Open Question Argument if that argument were sound). But Moore can hardly have had Lewis in mind when he penned this notorious passage in the early 1900s. Since he gives no citations it is tempting to suppose that Moore plucked his opponent out of thin air. But in fact, the person Moore warned us against was neither David Lewis nor Mr. Nobody but Bertrand Russell. For the definition of 'good' selected for dissection is precisely the definition suggested by Russell in 'Is Ethics a Branch of Empirical Psychology' a paper read to the Apostles in 1897[5]. (*RoE*: 71-78/*Papers* 1: 100-104.) 'The criterion [of morality] must be supplied, therefore, by the contrast between ideal and actual desires, by the contrast between the desires we desire and desires we dislike.' Moore does not just criticize Russell's definition in 'Is Ethics a Branch of Empirical Psychology?' - he criticizes the chief thesis of Russell's paper, namely that ethics is indeed a branch of empirical psychology. This thesis follows fairly obviously from the definition, since if 'good' means what we desire to desire, to find out what is good we need to ascertain what we desire to desire, which is a matter of psychological fact. And since [normative] ethics largely consists in the enquiry into what is good, ethics becomes a 'matter for purely psychological investigation' (*RoE*: 78/*Papers* 1: 104). The conclusion follows whether goodness consists in what *we* (the community, reasonable people or whatever) desire to desire (which is roughly Lewis's line), or whether goodness *for each individual* consists in what *that person* desires to desire (which is what Russell seems to suggest). Moore is very severe with this sort of thing. Naturalism, he says, consists in fixing on some natural property and then supposing that 'to be "good" *means* to

_____

[5] Most the works of Russell referred to in this paper are collected in Pigden, Charles R. (ed.) *Russell on Ethics* (London: Routledge, 1999), henceforward *RoE*. References are to *RoE* and to the relevant volumes of *The Collected Papers of Bertrand Russell* (abbreviated, for example, as *Papers* 1).

possess the property in question ... thus replacing Ethics by some one of the natural sciences'. 'In general,' explains Moore, 'Psychology has been the science substituted, as by J.S. Mill' (*PE*: §26). Moore does not add 'and by Mill's secular godson, Bertrand Russell', presumably because the substituting was done in a confidential paper read to a secret Society (the Apostles). Moore was very scrupulous about keeping the Society secret, so much so that he was worried about discussing its doings by postcard. See Griffin (2002: 186). But the consequence was that until very recently nobody realized that at least *one* of Moore's targets in *Principia Ethica* was Bertrand Russell.

I have two aims in this paper. In §§2-4 I contend that Moore has *two* arguments (not one) for the view that that 'good' denotes a non-natural property not to be identified with the naturalistic properties of science and common sense (or, for that matter, the more exotic properties posited by metaphysicians and theologians). The *first* argument, the Barren Tautology Argument (or the BTA), is derived, via Sidgwick, from a long tradition of anti-naturalist polemic. But the *second* argument, the Open Question Argument proper (or the OQA), seems to have been Moore's own invention and was probably devised to deal with naturalistic theories, such as Russell's, which are immune to the Barren Tautology Argument. The OQA is valid and not (as Frankena (1939) has alleged) question-begging. Moreover, if its premises were true, it would have disposed of the desire-to-desire theory. But as I explain in §5, from 1970 onwards, two key premises of the OQA were successively called into question, the one because philosophers came to believe in synthetic identities between properties and the other because it led to the Paradox of Analysis. By 1989 a philosopher like Lewis could put forward precisely the kind of theory that Moore professed to have refuted with a clean intellectual conscience. However, in §§6-8 I shall argue that all is not lost for the OQA. I first press an objection to the desire-to-desire theory derived from Kripke's

famous epistemic argument. On reflection this argument looks uncannily like the OQA. But the premise on which it relies is weaker than the one that betrayed Moore by leading to the Paradox of Analysis. This suggests three conclusions: 1) that the desire-to-desire theory is false; 2) that the OQA can be revived, albeit in a modified form; and 3) that the revived OQA poses a serious threat to what might be called *semantic naturalism*.

2. Moore's two arguments

Though Moore managed to convert Russell to non-naturalism (*RoE*: 73 & 75-104), there is reason to suspect that the desire-to-desire theory continued to be a worry. It is not always noticed that Moore has not one but *two* distinct arguments against naturalism, the Open Question Argument and the Barren Tautology Argument[6]. The first contends that 'good' cannot be synonymous with any naturalistic predicate 'X' since 'Are X things good?' is a significant or open question for every 'X'. The second contends that 'good' cannot be synonymous with any naturalistic 'X', if 'X things are good' is supposed to be a reason for action rather than a 'barren tautology'. The first is set forth at *PE*: §13, whilst the second crops up at *PE*: §11, though variants of it recur throughout the first four chapters (*PE*: §§14, 24 & 26). Russell (who was rather more succinct than Moore) summarizes it thus:

> Chapter II, on *Naturalistic Ethics*, discusses theories which hold that the only
>
> good things are certain natural objects, in so far as these theories are
>
> advocated as derivable from the very meaning of *good*. It is shown that such
>
> theories always confuse good, in its correct and indefinable sense, with the

---

[6] It seems to me that the discussion of Moore in Darwall, Gibbard and Railton's justly famous 'Towards *Fin de Siecle* Ethics: Some Trends', *The Philosophical Review*, 101 (1992), 115-189, is vitiated by a failure to distinguish clearly between the Open Question Argument and the Barren Tautology Argument.

sense which they assign to it by definition. For example, Evolutionist Ethics are apt to argue that *good* means *more evolved*, and on this to base practical recommendations. Yet, if their contention were correct, no practical consequences could follow. We ask: Why should I prefer this to that? And they reply: Because the more evolved is the better. But if they were right in the reason they give for thinking so, they have only said that the more evolved is the more evolved; and this barren tautology can be no basis for action. The meaning of two phrases cannot be the same, if it makes any difference whether we use one of them or the other; and, applying this test, it is easy to see that *more evolved* does not mean the same as *better*. (*RoE*: 100/*Papers* 4: 572.)

More formally, we can restate the argument as follows:

1#) For any naturalistic or metaphysical 'X', if 'good' meant 'X', then (i) 'X things are good', would be a barren tautology equivalent to (ii) 'X things are X', or (iii) 'Good things are good'.

2#) For any naturalistic or metaphysical 'X', if (i) 'X things are good', were a barren tautology, it would not provide a reason for action (i.e. a reason to pursue or promote X-ness).

3#) So for any naturalistic or metaphysical 'X', *either* (i) 'X things are good',

does *not* provide a reason for action (i.e. a reason to promote X-ness), *or*

'good' does not mean 'X'.

To put the point another way:

3#') For any naturalistic or metaphysical 'X', *if* (i) 'X things are good',

provides a reason for action (that is, a reason to promote X-ness), *then* 'good'

does not mean 'X'.

In other words, if you want the basic principles of your naturalistic ethic to be true by

definition, they can't at the same time be action-guiding. (Note: this argument does not entail

or presuppose that factual considerations cannot provide reasons for action.)

This is, I think, the *real* argument for the naturalistic fallacy, since it suggests that *most*

naturalists actually commit an intellectual mistake that can reasonably be described as a

fallacy - they propound as a reason for action some such principle as 'X things are good', or

even 'Only X things are good', and then try to defend it by claiming that it is some sort of

analyticity, 'the very meaning of the word', etc. (See Prior, A. N. (1949: chs. 1 & 9) and

PE: 11, 24.) But this is to subvert the action-guiding power of their original pronouncement.

It cannot *both* be that 'X things are good', is analytic (and thus secure from all shocks) *and*

that it provides a reason for promoting X-ness. To suppose that it can, or to propound such an

inconsistent view, is to make a mistake in reasoning that might reasonably be dubbed 'the

naturalistic fallacy'.

But on this reading, not all forms of naturalism are fallacious. In some cases the suggested 'X' is not supposed to denote a property that its proponent wants to see promoted. Rather the 'X' constitutes an *analysis* of 'good' which is designed to explain why thinking something good provides (or might provide) some sort of motive to promote it. The analysis is supposed to forge a conceptual connection between moral belief and action. Moore's Russell-derived example of 'what we desire to desire' provides a case in point. When Russell and (later) Lewis claim that goodness (or value) is what we desire to desire, this is not because they have a special yen for what we desire to desire and think that by calling it 'good' they can get people to maximize it. Rather they think that if we construe 'good' as what we desire to desire, we can see why people have a rational motive to promote what they believe to be the good. If we desire what we desire to desire (which we don't always do), then we will have a desire (and hence a rational motive) to promote what we believe to be good. The aim of the proposed analysis is not *action* but *understanding*, specifically an understanding of the 'conceptual connection between value and motivation' (DTV II: 113/69). Thus Russell and Lewis would be willing to concede that 'What we desire to desire is good', *is* a barren tautology, in the sense that it is unlikely to beget anything very spectacular in the way of action. But though it is an analytic truth and hence, if you like, a tautology, it is fruitful rather than barren when it comes to understanding the action-guiding power of 'good'.

Remember that the conclusion of the Barren Tautology Argument is a (quantified) conditional:

3#') For any naturalistic or metaphysical 'X', *if* (i) 'X things are good',

provides a reason for action (i.e. a reason to promote X-ness), *then*

'good' does not mean 'X'.

And it is quite consistent with 3#') that 'good' means 'X' for some naturalistic or metaphysical 'X' s*o long as* (i) 'X things are good', (or 'Good things are X') does *not* provide a reason for action (in the sense of a reason to promote X-ness).

But this means that the Barren Tautology Argument is a much less powerful engine against naturalism than is commonly supposed. It is not just the Russell/Lewis theory that escapes the net.   The Hutcheson/Hume theory (that value consists in a disposition to excite the approbation of a suitably qualified spectator)[7], the Michael Smith theory (that rightness is what we would desire ourselves to do if we were fully rational)[8], even the Hobbes/Locke/ Paley theory (that rightness consists in obedience to some Authority – God, the Sovereign or even the Beau Monde)[9] - *all* of them are immune to the Barren Tautology Argument. This is most surprising in the case of Hobbes/Locke/Paley, but all three could concede that 'Obeying the Authority is right', is a barren and unmotivating tautology, since each supplies *another* motive for obeying the Authority – the fear of punishment or, in the case of the Beau Monde, the fear of ridicule and ostracism.   At all events, we are a long way from a simple and unanalyzable property of goodness, which is what Moore wants to establish.

---

[7] See the extracts from Hutcheson and Hume in Raphael, D.D, (ed.) *The British Moralists 1650-1800*, 2 vols., (Oxford, Oxford University Press, 1967).

[8] See Smith, Michael, *The Moral Problem* (Oxford, Blackwell: 1991).

[9] Again see Raphael (1967) for relevant extracts.

3. Why the Open Question Argument?

Now, I am not sure how clearly Moore distinguished between his various arguments or if he was fully aware of how far they succeed. But the above analysis suggests an interesting speculation. We know that the BTA was developed *before* the OQA, since *PE*: §11 (which contains the Barren Tautology Argument) dates back to Moore's 1898 draft 'The Elements of Ethics' whereas *PE*: §13 (which contains the Open Question Argument) was written rather later. (See *PE*, revised edn.: 312-313.) There are indeed hints of the OQA in the *Elements of Ethics*, but so far as I can see they are only hints. In fact, the BTA is cribbed (with due acknowledgment) from Sidgwick's *The Methods of Ethics*. Now, it may be that Moore realized that the BTA does not do everything that he wanted. It points to a fallacy committed by many naturalists and it shows that you cannot found an action-guiding ethic upon a mere definition. But it does not exclude *all* forms of naturalism. In particular, it does not exclude the definition suggested by Russell, that 'good' means what we desire to desire. For this definition is not *intended* to provide a reason for action but to explain why goodness is a property which furnishes us with such reasons. *If* he realized this (and it is a pretty big 'if'), Moore *may* have been driven to invent the OQA in order to deal with naturalistic definitions such as this. For the OQA (if sound) would dispose of *all* brands of naturalism including the kind of theory propounded by Russell and Lewis. If this is correct, Russell's intervention may have forced Moore to move from the BTA to the OQA, which, despite one or two vague anticipations, seems to have been his own invention. (The final chapter of Prior (1949), 'The Naturalistic Fallacy – the History of Its Refutation', which deals at some length with anticipations of Moore, is exclusively concerned with the BTA.) Indeed, Russell's intervention might explain the long delay in the publication of *Principia Ethica*, which did not come out until 1903, even though Moore had a pretty good first draft by 1898. Perhaps

it took Moore a long time to come up with an adequate response. The flaw with this proposal is that Russell propounded his definition in 1897, *before* Moore wrote *The Elements of Ethics*. But it might have taken Moore a while to realize that the BTA could not deal with this particular threat.

4. The Open Question Argument stated

It may be useful at this point to state the OQA a little more precisely. The OQA (*PE*: §13) rests on three premises.

1) 'Are X things good?' is a significant or open question for any naturalistic or metaphysical predicate 'X' (whether simple or complex).

'Every one does in fact understand the question'; it is 'intelligible', it can be 'asked with significance' and 'we can understand very well what is meant by doubting' the answer. (*PE*: §13) Such questions would *not* be 'significant' (in Moore's sense) if an understanding of the words involved were enough for an affirmative answer. This is the case with 'Are X things X?', 'Are good things good?' and 'Are bachelors unmarried?' where the questions posed are, in effect, interrogative tautologies But since 'Are X things good?' *is* 'significant' for any 'X' (indeed 'significant' for 'every one' by which Moore would appear to mean all competent speakers) it follows that an understanding of the words involved (which is shared by all competent speakers) does not suffice for an affirmative answer.

2) If two expressions (whether simple or complex) are synonymous this is evident on reflection to every competent speaker.

3) The meaning of a predicate or property word is the property for which it stands. Thus if two predicates or property words have distinct meanings they denote distinct properties.

From 1) and 2) it follows that

4) 'Good' is not synonymous with any naturalistic or metaphysical predicate 'X' (or 'goodness' with any corresponding noun or noun-phrase 'X-ness').

If 'good' *were* synonymous with any predicate 'X', then this would be evident on reflection to every competent speaker and the question 'Are X things good?' would not be open or significant for that particular 'X'. Thus, the fact that 'Are X things good?' *is* significant or open for every 'X' shows that 'good' is *not* synonymous with any such predicate. But Sub-conclusion 4) does not give Moore everything he wants. It states that the *word* 'good' is not synonymous any natural *predicate*, not that goodness itself is not identical with any natural or non-moral *property*.   It is tantamount to what I call the *semantic* autonomy of ethics, the thesis that moral words are *not* susceptible to a naturalistic definition. (See Pigden (1991).) It is therefore incompatible with semantic *naturalism*, which is precisely the thesis that the moral *can* be reduced to the non-moral by means of definitions, i.e. by establishing that 'good' (or whatever) *means the same as* some (presumably complex) naturalistic predicate. Moore however professes a lofty disdain for mere semantics. 'Verbal questions are properly left to the writers of dictionaries and other persons interested in literature; philosophy, as we shall see, has no concern with them' (*PE*: §2). He has bigger fish to fry. He wants to establish

what I call the *ontological* autonomy of ethics, the thesis that for moral judgments to be true there must be a realm of distinctively moral facts and properties, of which goodness is the chief. Nevertheless, Sub-conclusion 4) is not without importance, since if it is true, the Russell/Lewis theory is false. For Russell and his unwitting disciple David Lewis are both semantic naturalists. Lewis is quite explicit about this. His theory, he says is 'naturalistic [i.e. semantically naturalistic] since it advances an analytic definition of value' (DTV II: 113/68). But so too is Russell's. 'Unless, therefore, the good can be defined otherwise than in terms of desire, ethics, properly studied, must always remain … purely a branch of empirical psychology' (*RoE*: 75/*Papers* I: 102). In Russell's view the good *can't* be defined otherwise than in terms of desire, which means that ethics is indeed a branch of empirical psychology. But the point is that it is a *definition*, a purported analysis of the concept 'good', that is supposed to do the trick. But important as it is for Moore to refute the likes of Bertrand Russell, he wants to go one better. He wants to go beyond the word 'good' to the property for which it stands. How does he get from *semantic* autonomy (a predicate, 'good', that cannot be defined in terms of the non-moral) to *ontological* autonomy (a non-natural property of goodness that cannot be *identified* with anything non-moral)? By appealing to Premise 3), the thesis that the meaning of a predicate is the property that it denotes and thus that if two predicates have distinct meanings they stand for distinct properties. Moore certainly believed in properties at the time he wrote *Principia*, and then and thereafter, he seems to have subscribed to a 'one-level' theory of meaning according to which the meaning of a word is the thing it denotes. (See Baldwin (1990: 39-50 and 203) and Hylton (1990: 140-141)). Premise 3) provides the bridge between semantics and ontology.

From 3) and 4) it follows that

5) Goodness is not identical with any natural or metaphysical property of X-ness.

Since 'good' has a distinct meaning from every naturalistic or metaphysical predicate 'X', it denotes a distinct and non-natural property. And this is precisely what Moore set out to prove.

5. The OQA discredited

Premises 1), 2) & 3) suffice to prove Moore's point. But Premise 3) is highly questionable. Bob Durrant (1970) was perhaps the first to point out a) that Moore's argument requires some such premise if it is to succeed but b) that the assumption depends upon a purely referential theory of meaning according to which there is nothing more to the meaning of a predicate than the property for which it stands. Once we admit that, non-synonymous predicates can refer to the same property (just as non-synonymous names can refer to the same thing), Moore's argument for 5) collapses and he is reduced to Sub-conclusion 4). 'Good' may not be synonymous with any naturalistic predicate 'X' (whether simple or complex) but this does not prove that goodness is not identical with some naturalistic property of X-ness. We can no longer proceed from an unanalysable and non-natural *predicate* 'good' to an unanalysable and non-natural *property* of goodness. Cornell realists rejoice in this fact and happily propound synthetic identities between moral properties and others analogous to the celebrated identity between water and H_2O. We can have moral truth without either metaphysical spooks or implausible attempts to give a naturalistic definition of the word 'good'.

Nevertheless, Sub-conclusion 4) is not without importance, since it suggests the *semantic* autonomy of ethics, the thesis that morals words are not susceptible to naturalistic

definition. It is therefore incompatible with *semantic* naturalism, which claims that the moral can be reduced to the non-moral *by definition*, i.e. by establishing that 'good' *means the same* as some naturalistic predicate 'X'. Thus if Sub-conclusion 4) is correct, the Russell/Lewis theory is false.

But Sub-conclusion 4) depends upon Premise 2). And Premise 2) is false. For it leads straight to the Paradox of Analysis, a problem that Moore recognized but did not succeed in solving. The Paradox first appeared in a paper by Langford (1942) but was probably discovered by Moore himself (Baldwin (1990: 208)).   The Paradox is that conceptual analysis (which was Moore's stock in trade) is either useless or productive of falsehoods. For suppose the *analysandum* (the expression to be analyzed) means the same as the *analysans* (or analyzing phrase).   Then by 2) this will be evident to every competent speaker and the analysis will teach us nothing new.   Suppose on the other hand that that the analysis teaches us something new, i.e. that it is *not* evident on reflection to every competent speaker. Then, again by 2), the analysis is false. For if it is not evident to every competent speaker that the *analysans* and the *analysandum* share the same meaning, then they won't share the same meaning and the analysis will be false.   (Baldwin (1990: 210-211, Pigden (1990: 427) Darwall, Gibbard and Railton (1992: 115).)   If conceptual analysis is to be a worthwhile enterprise, one *capable* of turning up new and interesting truths, Premise 2) which generates the Paradox had better be false. And it *is* false, since it presupposes that our concepts are transparent to us. This is a point now widely recognized. Baldwin, for instance states that what I call Premise 2) relies on 'the Cartesian conception of the content of thought as transparently available to the subject', whilst Darwall, Gibbard and Railton talk of 'assumptions about the transparency of concepts and the obviousness of analytic truth'. Moore therefore is in the embarrassing position of relying on an assumption which, if true,

would have sabotaged his philosophical career. It is an assumption that anyone who believes in the possibility of conceptual analysis - from Moore and Russell through to David Lewis and Frank Jackson - must reject. What is analytic isn't always obvious. Hence the fact (if it be a fact) that 'Are X things good?' is an open question for every naturalistic or metaphysical 'X' does not prove that 'good' is not synonymous with some such 'X'.

Thus the Open Question Argument has collapsed. It relied on three premises, 1) 2) and 3). Premise 3) was shot down in the 1970s, first by Bob Durrant and subsequently by others such as Putnam (1981: 205-211.). This opened the way for brands of naturalism such as Cornell Realism which rely on synthetic identities. It was a liberating thought that you can have moral truths without resorting to non-natural properties or dubious conceptual analyses. But the deletion of Premise 3) still left the OQA able to limp along as a disproof of semantic naturalism. But somewhere around 1980 people began to realize that Premise 2) 'the publicity condition', leads to the Paradox of Analysis and therefore had to go. The first person to make this point was Casimir Lewy in a paper published as far back as 1964. But despite the most memorably bizarre set of lecturing mannerisms that I have ever encountered, Lewy was not a philosophical superstar, and his paper went largely unnoticed. I heard the point first from the lips of David Lewis in 1981 (though at the time I did not understand what he was getting at). Not surprising then, that Lewis went on to reinvent a version of the desire-to-desire theory that the OQA was devised to disprove. The wheel had come full circle.

6. G.E. Moore redux?

At the moment Moore seems to be in a pretty bad way while Russell and Lewis are laughing. The BTA does not work against the desire-to-desire theory whilst the OQA, which was probably invented to dispose of it, relies on three premises, two of which are false. Moore

made his name with an argument, which, if it were sound, would have made mincemeat of much of his subsequent philosophy. But it was one of Lewis's pet theses that knock down refutations are rare to non-existent in philosophy. In the latter part of this paper I shall be illustrating this thesis by arguing that there is something to be said for the OQA - though in a suitably amended form, of course.

7. Colors, values and the epistemic argument.

As the title suggests, Lewis's version of the desire-to-desire theory is a *dispositional* theory of value. It stands in long a tradition which represents value properties as akin to secondary properties, and construes secondary properties themselves as dispositions to cause certain effects in us[10]. I want to consider an argument of Kripke's – the epistemic argument - that can be deployed to show that neither colors nor values should be understood as dispositional properties. It is my contention that the argument fails with respect to colors but succeeds with respect to values. Thus this semi-successful argument tends to show that goodness *unlike* yellow is *not* a secondary property, and hence that Lewis's desire-to-desire theory is false. Furthermore, the successful argument turns out to be a variant of the OQA.

The principal purpose of Kripke's *Naming and Necessity* is to argue that proper names are – at least typically – rigid designators and that they lack sense. But one of his subsidiary purposes is to argue against a dispositional account of secondary properties and, more specifically, colors. (Kripke (1980: 140n).) In Kripke's view, color terms such as 'yellow' are also rigid designators, and, like proper names, they too lack sense. 'Yellow' denotes yellowness, but *not* by abbreviating some such description as 'that (manifest) property of

---

[10] There are hints of this in Shaftesbury, but Frances Hutcheson (1694-1746) was probably the first philosopher of modern times to produce a well-worked out account of moral properties as secondary qualities. He found a distinguished, if heterodox, disciple in Hume. See Raphael (1967) for relevant extracts.

objects that causes them, under normal circumstances to be seen as yellow (i.e., to be sensed by certain visual impressions)'. That description fixes the reference but not the sense of 'yellow' since the word 'yellow' has no sense to fix. In my view Kripke is wrong about this. But I am not going to discuss the matter in detail. I am just going to focus on *one* of his arguments, the epistemic argument. We will first apply the argument to names, then to color terms and finally to 'good'.

1*) If 'Shakespeare' meant 'the actual[11] author of *Hamlet, Othello* etc.', then it would be analytic that Shakespeare (if he existed) was the actual author of *Hamlet, Othello* etc.

2*) But it is not analytic that Shakespeare (if he existed) was the actual author *Hamlet, Othello* etc.

3*) So 'Shakespeare' does not mean 'the actual author of *Hamlet, Othello* etc.' (Salmon (1981: 27-29).

I have no comment to make about this argument, which seems completely convincing at least with respect to the vast bulk of proper names. Can it be adapted to show that words like 'yellow' are senseless and should not be subjected to a dispositional analysis?

1**) If 'yellow' meant 'that property, if it exists, that given our actual optical propensities, excites yellow sensations under normal circumstances', then it

---

[11] As is well known, Kripke's *modal* argument (to the effect that names and other terms lack descriptive content) can be neutralized if we rigidify the terms in question with the aid of some well-placed 'actual's and 'actually's. It seems to me that 'yellow' and 'good' are both used rigidly which means that an 'actually' must be read into them. For ease of exposition I have given 'Shakespeare' the same treatment. See Salmon, Nathan U., *Reference and Essence* (Oxford: Blackwell, 1982), 26ff and Lewis, DTV II: 132-133/88-89.

would be analytic that given our actual optical propensities, yellow, if it exists, excites yellow sensations under normal circumstances.

2**) But it is synthetic not analytic that given our actual optical propensities, yellow, if it exists, excites yellow sensations under normal circumstances.

3**) So 'yellow' does not mean 'that property, if it exists, that given our actual optical propensities, excites yellow sensations under normal circumstances'.

I deny the second premise. It is *not* a matter of synthetic fact that yellow (if it exists) actually excites yellow sensations, but something we learn with the language. What *is* synthetic is that things with a certain range of surface microstructures and reflexive propensities excite yellow sensations. But this I can happily admit. Indeed it paves the way for a synthetic identity between instances of yellow and the microstructural properties which excite the sensations. In *this* case the epistemic argument fails.

So much for 'yellow', what about 'good'?

1***) If 'good' meant what Lewis thinks it means, then it would be analytic that what we are actually, ideally disposed to desire to desire is good.

2***) But this is, if true, a matter of synthetic fact.

3***) Accordingly the analysis is false and goodness is not a secondary property.

In the case of yellow I denied the second premise - that it is synthetic that if there is such a thing as yellow, it is what actually arouses yellow sensations in us. But this will not wash with good. For it is quite conceivable that what ideal human beings actually value is mostly

bad.   We can imagine Luther's opinion of the Lewis's theory.   The idea that the desires of unregenerate human beings should be a guide to the good would strike him as ridiculous. The whole project of converting oneself into an ideal desirer smacks of the impious vanity of the damned.   For Luther our second order desires (whether idealized or not) would be like reason - the Devil's whore.   (See Luther (1957: 46).)   The King of Brobdingnag, after hearing a somewhat slanted catalogue of human achievements, told his pet human that despite the high regard he felt for him it was obvious that mankind (well actually Englishmen) were 'the most pernicious race of little odious vermin that nature has ever suffered to crawl upon the face of the earth'. (Swift (1967: 173).) Is it likely that the desires of such contemptible creatures, even the best of them, will indicate what goodness is? Won't they rather give an absurdly high ranking to the interests of their own noxious species? Now I do not *agree* with Luther and the King of Brobdingnag (except in my most jaundiced moments). But their views embody no manifest contradiction. If it *is* analytic that 'What we humans ideally desire to desire is good.', it is certainly not an obvious analyticity.

Isn't the epistemic argument just the Open Question Argument all over again? After all, the OQA was that 'good' cannot mean 'X' because we can conceive of something being X without it's being good. (That is why the question 'Are X things, good?' makes sense or can be sensibly asked). We cannot in the same way conceive of someone as a bachelor without conceiving of him as an unmarried man - not if we know the meanings of the relevant words that is. Now the epistemic argument is very like this. It claims that 'good' cannot be analyzed as what we would be ideally disposed to desire to desire (i.e.: it is not analytic that 'What we are ideally disposed to desire to desire is good'). Why not? Because we can conceive that we are ideally disposed to desire to desire something which is not really good. Indeed we can imagine that it might be bad.

The candid answer to my question has to be yes - the epistemic argument *is* pretty much a rehash of the OQA. Then shouldn't it be discarded? Once we admit that there are unobvious analyticities, the fact that we can conceive of X-s which are not Y-s does not demonstrate that 'X' is a mistaken analysis of 'Y'. It might be that we have not thought the matter through. Hypothetico-deductive methods and even empirical research are required to establish non-trivial analyticities. Linguistic intuitions are not enough. Nor are linguistic intuitions enough to *dis*establish an alleged analyticity. Our intuitions may not penetrate to the buried rules and presuppositions that govern our use of language.

Nevertheless, our intuitions about what can and what cannot be conceived are not devoid of probative force. They reflect, albeit imperfectly, our understanding of the concepts we employ. Hence they can provide evidence for and against analytical hypotheses, though this evidence ceases to be decisive. Now it does seem to me clear that Luther and the King of Brobdingnag could be right, and that what we ideally desire to desire could be wrong. In which case we have evidence, though not conclusive evidence, against the desire-to-desire theory.

Now Lewis (DTV II: 132/88) wants to argue that in this case the intuitive evidence is misleading. My intuitions reflect my superficial thinking not the deep structure of the relevant concepts. If we try to flesh out Luther's story or the story of the King, we see that the hypothesis collapses. We cannot really imagine what it would be like for what we ideally disposed to value to be wrong. To talk largely of human depravity is not enough. What we need is corroborative detail: a plausible disvalue ideal humans are inclined to value or a value they are inclined to disvalue. Although Swift was about as misanthropic as they come, and although *Gulliver* is from first to last a satire on human nature, he does not manage to provide this. Yahoos, of course, value all sorts of nasty things, but they are far from ideal.

Englishmen likewise value things of doubtful worth, such as money, titles and military renown. But Englishmen too can be improved upon; for a start, they could be more disinterested. But what about truly ideal people? Could *ideal* humans be wrongly inclined to value domination - something Swift's repulsive puppets pursue? But although sub-ideal human beings in their fallen state may value domination, this is not something they need desire to desire, nor is it something they *would* desire to desire if the defects which make them less than ideal were removed.

I think Lewis' challenge can be met and corroborative detail supplied. Indeed it has been supplied in numerous works of Science Fiction. It is the oldest trick in Science Fiction's book to present human beings through the eyes of the aliens, and to present them as ridiculous and repulsive. They are greedy, savage, selfish and stupid, destroying other species and the ecosystems on which they depend. Given space-travel they will spread through the galaxy like a noxious slime, blighting every planet they touch. They must be stopped! If I pile it on thick enough, I can get you to sympathize momentarily with the extermination of the human race. It is at least conceivable that the destruction of the human race would be a Good Thing from the point of view of the Universe as a whole. But could ideal human desirers be brought to desire this, not just in a moment of ecological frenzy, but as a settled policy? Surely not. Human Chauvinism, in the weak form of a tendency to desire our continued survival is surely part of our make-up and it is not a desire that we desire not to have – not even under ideal circumstances. So the extinction of the human race is a conceivable (if not a plausible) value that ideal human beings would be inclined to disvalue.

Now if we can conceive of what is good and what the best of us are disposed to desire to desire coming apart, this is evidence (though not conclusive evidence) that the one does

not constitute an analysis of the other. The epistemic argument may be weaker than Kripke supposes, but in this context, it works after a fashion.

8. The OQA amended

But if the epistemic argument works after a fashion, then there is something to be said for the OQA too since in this context they come to much the same thing. The OQA failed because Premise 2), the publicity condition had to be rejected. However, even if we reject 2), and with it the assumption that our concepts are transparent to us, it does not follow that Moore's argument is entirely worthless. For there are weaker and more plausible variants of 2) which would take us to probabilistic variants of Sub-conclusion 4). Consider for example:

> 2') If it is evident to *some* competent speakers that two expressions 'X' and 'Y' are *not* synonymous (since it is not analytic that X is Y), then this is evidence (though not conclusive evidence) that they are not, in fact, synonymous.

This suggests that following reformulation of the OQA, though as reformulated it presupposes not an Open Question but a strong intuition on the part of some speakers that the two expressions do *not* mean the same. Nevertheless we shall call it the OQA*:

> 1') It is evident to some competent speakers that (so far as our understanding of the words is concerned) a thing could be X without being good.

2') If it is evident to *some* competent speakers that two expressions 'X'

and 'Y' are *not* synonymous (since it is not analytic that X is Y), then this

is evidence (though not conclusive evidence) that they are not, in fact,

synonymous.

4') 'Good' is *probably* not synonymous with the predicate 'X' (or

'goodness' with the corresponding noun 'X-ness').

This argument, the OQA* can be deployed against most naturalistic definitions of the moral

predicates, though whether it can be deployed against *all* such definitions is, to coin a phrase,

an open question. Furthermore the argument has to be deployed piecemeal. We no longer

have a blanket premise covering *all* naturalistic predicates. Instead we have a series of

specific premises for specific naturalistic predicates 'X', saying that according to some

competent speakers a thing could be X without being good. It is no longer enough *not* to

believe that the two terms *are* synonymous (which is all that was required for the original

OQA). Rather *someone* has to believe that they are *not* synonymous, because they can

conceive of a thing's being X without being good. And we can't be sure that 1') will be

satisfied for *every* naturalistic predicate 'X'. Thus the revised argument does not constitute a

refutation of semantic naturalism but a series of *potential* refutations of *specific versions* of

semantic naturalism   And even when Premise 1') is true for some 'X', the refutation is far

from conclusive. Nevertheless OQA* constitutes an argument schema which can be used to

refute – or perhaps since that's a weaker word, to *discredit* – a wide variety of semantic

naturalisms, among them the theory that to be good is to be what we are ideally disposed to

desire to desire. This is much less than Moore purported to prove, but it does indicate that he is not quite the undischarged intellectual bankrupt that he once appeared to be.

REFERENCES

Baldwin, Thomas, *G. E. Moore* (London: Routledge, 1990).a

Darwall, S., Gibbard, A. and Railton. P., 'Towards *Fin de Siecle* Ethics: Some Trends', *The Philosophical Review*, 101 (1992), 115-189.

Durrant, R. G., 'Identity of Properties and the Definition of Good', *Australasian Journal of Philosophy*, Vol. 48 (1970), 360-361.

Frankena, W. K., 'The Naturalistic Fallacy', *Mind*, 48, (1939) 464-477.

Griffin, Nicholas, (ed.) *The Selected Letters of Bertrand Russell, vol. 1: The Private Years 1884-1914* (London: Routledge, 2002).

Hylton, Peter, *Russell, Idealism and the Emergence of Analytic Philosophy* (Oxford: Oxford University Press, 1990).

Kripke, Saul, *Naming and Necessity* (Oxford: Blackwell, 1980).

Langford, C.H., 'Moore's Notion of Analysis', in Schilpp, P.A. (ed.), *The Philosophy of G. E. Moore* (Lassalle: Open Court. 1942), 319-342.

Lewis, David, 'Dispositional Theories of Value II' *Proceedings of the Aristotelian Society Supplementary Volume,* 63 (1989), 113-137.

Lewis, David, 'Dispositional Theories of Value', in Lewis, David, *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press, 2000), 68-94.

Lewy, Casimir, 'G.E. Moore on the Naturalistic Fallacy', *Proceedings of the British Academy*, 50 (1964), 251-262.

Luther, Martin, *On the Bondage of the Will*, trans. J.I. Packer and O.R. Johnston, (Cambridge: James Clarke, 1957).

Moore, G.E. *The Elements of Ethics*, ed. Tom Regan (Philadelphia: Temple University Press, 1991).

Moore, G.E. *Principia Ethica*, revised edn,, ed. Thomas Baldwin (Cambridge: Cambridge University Press, 1993).

Pigden, Charles R., 'Naturalism' in Singer, Peter (ed.), *A Companion to Ethics*, (Oxford; Blackwell, 1991).

Pigden, Charles R. (ed.), *Russell on Ethics*, (London: Routledge. 1999).

Prior, A. N., *Logic and the Basis of Ethics*, (Oxford: Oxford University Press, 1949).

Putnam, Hilary, *Reason, Truth and History*, (Cambridge: Cambridge University Press, 1981).

Raphael, D.D, (ed.) *The British Moralists*, 2 vols., (Oxford, Oxford University Press, 1967).

Russell, Bertrand, *The Collected Papers of Bertrand Russell. Vol. 1, Cambridge Essays, 1888-1899*, eds. Blackwell, K., Brink, A., Griffin, N., Rempel, R. and Slater, J. (London: Allen & Unwin. 1983)

Russell, Bertrand, *The Collected Papers of Bertrand Russell. Vol. 4, Foundations of Logic 1903-05*, eds. Urquhart, A. and Lewis, A. C. (London: Routledge, 1994).

Salmon, Nathan U., *Reference and Essence* (Oxford: Blackwell, 1982).

Sidgwick, Henry, *The Methods of Ethics*, 7th edn. (London: Macmillan, 1907).

Smith, Michael, *The Moral Problem* (Oxford, Blackwell: 1991).

Swift, Jonathan, *Gulliver's Travels* (Harmondsworth: Penguin, 1967).