

## **Necessary Moral Principles**

Richard Swinburne

[Swinburne, Richard, 2015, "Necessary Moral Principles", *Journal of the American Philosophical Association*, 1 617-634.]

Moral realism, as I shall understand it, is the doctrine that some moral propositions, that is propositions which attribute a moral property to some action, are true<sup>2</sup>. Moral supervenience, the highly plausible doctrine which I shall take for granted, that the moral supervenes on the non-moral, then becomes<sup>3</sup> the more precise doctrine that moral properties strongly supervene on non-moral properties in Kim's (1993: 80) sense of 'strongly supervene'. In this sense a property of kind A supervenes on a property of kind B iff it is metaphysically necessary that for every object that has a property F of kind A, there is some property G of kind B, such that any object which has property G has property F, and has property F *because* it has property G. So the supervenience of the moral on the non-moral would consist in a metaphysical necessity that any particular action or kind of action is morally good or bad, right or wrong, because of some non-moral property (normally, a conjunction of such properties) which it has. Thus, plausibly what Hitler did in ordering the German invasion of Poland in September 1939 was morally wrong because it was an act of attacking a peaceful independent country; what Florence Nightingale did in organizing a hospital in the Crimea in 1856 was morally good because it was an act of caring for the sick, etc. No action can be just morally good or bad; it is good or bad because it has certain non-moral properties – those of the kinds which I have illustrated. And any other action which had just those non-moral properties would have the same moral properties. I shall call any set of the most general metaphysically necessary moral propositions stating which moral property supervenes on which non-moral properties, which entail all other such necessary propositions a set of the true fundamental moral 'principles'. The conjunction of non-moral properties which gives rise to the moral property might be a long one or a short one; and the examples which I provided of the non-moral properties which conferred on the actions of Hitler or Florence Nightingale the moral properties, may well be too short a list. It may be that all acts of attacking a peaceful independent country are bad, or it may be that all acts of invading a peaceful independent country, which is not harbouring groups of foreign fighters about to attack your country or [a list of different circumstances in which it would not be bad to attack such a country] are bad. But it must be that if there is a metaphysically possible world W in which an action *a* individuated by a conjunction of all its

non-moral properties is bad, there could not be another world  $W^*$  which is exactly the same as  $W$  in all non-moral respects, but in which  $a$  was not bad<sup>4</sup>. So, for every true contingent moral proposition (e.g. 'action  $a$  was morally good',  $a$  being individuated by some unique description) there is a metaphysically necessary moral proposition (e.g. 'all actions of non-moral kind  $Z$  are morally good') which together with a contingent non-moral proposition ('action  $a$  was an action of kind  $Z$ ') makes the contingent moral proposition true.

In its strongest form in which I shall consider the doctrine that are such necessary moral propositions, I assume that a proposition is metaphysically necessary iff either it is 'logically' (in a narrow sense) or 'conceptually' necessary, or as strongly necessary as either of these. I shall group together the 'logically' and 'conceptually' necessary, and call any proposition whose modal status as metaphysically necessary, impossible, or whatever can be determined a priori, *logically* necessary, impossible, or whatever. Yet with this understanding of metaphysical necessity it seems initially to many thinkers that it is implausible to suppose that there are metaphysically necessary moral propositions; and that reason is a major reason why moral realism has not seemed very plausible to a substantial number of thinkers.<sup>5</sup> The purpose of this paper is to argue that it is very plausible to suppose that there are metaphysically necessary moral propositions, and in particular (in my wide sense) logically necessary ones. For the purposes of this paper I shall understand 'moral' goodness, badness etc as overall goodness, badness etc; but the results of this paper can easily be applied to narrower conceptions of the moral, for example as the kind of goodness etc. involved in obligations to other sentient creatures .

Some recent writers, including Shafer-Landau (2003: 87-98) and Brink (1989: 179)<sup>6</sup>, have tried to make it plausible to suppose that there are a posteriori metaphysically necessary moral propositions. The examples by which Kripke (1980) and Putnam (1975) persuaded us that there are a posteriori necessary propositions such as 'Hesperus is Phosphorus' and 'water is  $H_2O$ ' seem to exhibit a common pattern. An object (substance, property or whatever) is picked out by a rigid designator, but our understanding of that particular rigid designator is that it designates an object which is the object it is in virtue of an underlying essence which it needs a posteriori work to discover. Having that essence, it could not not have that essence; and so any sentences which state that it has that essence or state something which is entailed by it having that essence are necessary with as hard a necessity as logical necessity. Thus, when 'water' is understood – as Putnam supposed that it was

in the late eighteenth century - as a rigid designator of the transparent drinkable liquid in our rivers and seas, what makes the stuff that stuff is its chemical essence - being  $H_2O$ . Having that essence, it could not not have that essence. But it needs a posteriori investigation to discover what that essence is. So, the suggestion is that moral principles might be like this.

But, while of course a posteriori investigation is needed in order to discover the truth of contingent moral truths, all argument about the underlying moral principles relies on a priori considerations. When examples of particular situations (e.g. the trolley problem) are adduced in order to persuade us that some general moral principle is or is not true, it is quite irrelevant whether the examples are examples of an actual event or of an imagined event. What matters is what it would be right to conclude about which actions in that situation would be good or bad - whether or not the situation actually occurred is irrelevant. So any metaphysically necessary moral principles must be logically necessary. Hence, either there are logically necessary moral principles, or moral realism is false.

Those who claim that fundamental moral principles are (in my wide sense) logically necessary sometimes claim that their necessity is a different kind of necessity from logical necessity (in a narrow sense) or conceptual necessity. Some writers have argued that the necessity of moral propositions is like that of fundamental epistemic propositions, which in effect (since in effect they claim that these propositions are true in all logically possible worlds) they claim to be a special kind of logical necessity (in my wide sense). Parfit (2011: 524-5) compares other normative propositions to 'normative truths about credibilities and epistemic reasons'. Wedgwood (2007 :153-173) argues that the possession of 'intentional' concepts and in particular the concept of belief, entails a view about what are the rational procedures for acquiring and using them, and that rationality is a normative concept. Cuneo (2007) has written a whole book arguing that the supervenience of epistemic propositions on non-epistemic ones is just like the supervenience of moral propositions on non-moral ones. By 'epistemic' propositions these authors have in mind propositions claiming that some belief is 'rational' or 'justified' or 'probably true', and by 'epistemic' principles principles which determine when this is the case. But there seems to be a sharp distinction between the supervenience of moral propositions and the supervenience of epistemic propositions, in that the normative force of the supervenient epistemic proposition is hypothetical - e.g. 'this belief is justified' entails 'if you want a true belief, you should hold this one'; whereas the normative force of the supervenient moral proposition is categorical - e.g. 'this action is obligatory' entails 'whatever you want, you ought to do this.' For this reason it would be better not to argue to the logical necessity of

moral principles on the basis of their similarity to epistemic principles.

Fine (2002) has taken the heroic stance of claiming that the necessity of moral principles, which he calls 'normative necessity', is *sui generis*, quite distinct from metaphysical necessity or natural necessity (the necessity possessed by laws of nature), but that leaves it very obscure how we can have access to such necessity - something much less obscure in the other two cases. So I shall argue in this paper for the logical necessity of moral principles on the grounds of the similarity of ways in we establish them to ways in which we establish obvious logical (in the narrow sense) and conceptual necessities.<sup>7</sup> In order to do this, I need to devote the next section to putting forward a view of what it is for some proposition to be logically necessary, and of how we can show that probably it is. It is a view resulting from conceivability accounts of logical possibility (such as Yablo (1993) and Chalmers (2002) which - I suggest - are not very different from 'understanding-based' accounts.

## II

Following Chalmers (2002: 147), I shall understand conceivability as 'a property of statements', that is token declarative sentences uttered (or whatever) on a particular occasion. Sentences express propositions, and the propositions are the propositions they are in virtue of the rules of the language governing the meanings of sentences. Sentences contain words, words express concepts which are contained in propositions. The rules of a language governing the meanings of words (and so of the sentences which contain them) are of two kinds - syntactic and semantic. Syntactic rules relate a word to other words, describing what they entail and what is entailed by them, the kind of rules you find in a dictionary, such as 'a philatelist is someone who collects postage stamps or who knows about different kinds of postage stamps.' Semantic rules relate a word to objects or properties to which it paradigmatically applies, such as 'red is the colour of ripe strawberries, British post boxes, and London buses'. By 'paradigm examples' of the application of a word or sentence, I mean examples such that it would be inconceivable that that word did not apply to almost all of those examples (while allowing that it might not apply to a few of them). Some words have only syntactic rules for their use, as for example words introduced by arbitrary definition. But plausibly unless the words to which a given word is related are themselves logically related to other words and they to other words and so on, until we come to words which have semantic rules for their use, we wouldn't know what that word

meant. At any rate the word could not designate a (metaphysically) contingent property possessed by some object or property. While different syntactic rules for a word entail that it expresses a different concept, different semantic rules need not have that consequence. For clearly two words can mean the same even if they are introduced by different paradigm examples of their correct application – ‘red’ can be given a meaning by quite different sets of examples of red objects. But two words cannot mean the same if they don’t apply to the same (paradigm or non-paradigm) objects – although it might need much investigation to show whether or not that is the case.

I shall assume for the next few paragraphs that speakers of a language agree about what would make a sentence true or false in any narrow set of circumstances to which it has been applied in the past or would be applied in future, and so in that sense they have a common understanding of the meanings of words and sentences. I shall call a sentence which expresses a logically necessary (possible/impossible) proposition, a logically necessary (possible/impossible) sentence. I shall equate being logically (im)possible with being ‘ideally’ ‘(in)conceivable’, that is ‘conceivable on ideal rational reflection’ (Chalmers loc.cit.); and that a sentence being ‘apparently’ (in)conceivable (‘prima facie’(in)conceivable) is, in the absence of counter-reason, good reason to suppose that it is (in)conceivable; and that a sentence being ‘obviously’ (in)conceivable is very strong reason to suppose that it is (in)conceivable. I shall however understand ‘conceivable’ as ‘such that it makes sense to suppose that it is true’, a sense which may be a wider sense than that espoused by Chalmers,<sup>8</sup>.

Then we can come to see by a priori reflection on the public usage of the words and sentence forms involved that there are some declarative sentences which could not be true whatever the world was like, because it is obviously ‘inconceivable’ that there is any state of affairs which would constitute those sentences being true. The paradigm example of such a sentence is a self-contradictory sentence (e.g. one of the form ‘both  $p$  and not- $p$ ’.) But there are innumerable other examples of declarative sentences which are not explicitly self-contradictory, but are also such that it is obviously inconceivable that there is any state of affairs which would constitute them being true. These include what used to be called ‘category mistakes’ – ‘the Prime Minister is a prime number’ or ‘I had two helpings of democracy for breakfast’. We can see too that there are some declarative sentences such that it is obviously inconceivable that there is any state of affairs which would constitute them being false, for example the negations of self-contradictory sentences, and so must be true and so logically necessary. We also come to see to which other sentences a speaker who utters a given sentence is explicitly committed in virtue of the

rules of the language. If it is obviously inconceivable that ( $p$  and not- $q$ ), then in uttering  $p$ , a speaker is explicitly committed to  $q$ ; and in that case (see Swinburne (2013:18)) I shall say that  $p$  'mini-entails'  $q$ . We can see also that various sentences other than logically necessary ones are obviously not logically impossible, and so are logically possible.

A priori reflection can then lead us to discover the logical modality of other sentences which was not initially obvious. The direct way of discovering these things is by discovering entailments between sentences whose logical status is obvious.  $p$  entails  $q$  iff ( $p$  & not- $q$ ) is logically impossible. One way of discovering that  $p$  entails  $q$  is to discover that  $p$  and  $q$  can be joined by a chain of mini-entailments, that is  $p$  mini-entails  $s$ ,  $s$  mini-entails  $t$ , and so on until we reach a sentence which mini-entails  $q$ . We can discover that a sentence is logically impossible by discovering that it entails an obviously logically impossible sentence. We can discover that a sentence is logically necessary by discovering that its negation is logically impossible. And we can discover that a sentence is logically possible by discovering that it is entailed by some obviously logically possible sentence.

While these methods will enable all investigators who make the same judgments about what is apparently conceivable and apparently inconceivable, to determine the logical status of many sentences (and so of the propositions which they express), there will clearly be many sentences whose status they cannot determine straight off-ones such that they cannot immediately find an obviously conceivable sentence which entails it, or a route of proof by which they can deduce from it a self-contradiction. In that case there is an indirect method which may help investigators to discover the logical modality (as defined above) of some propositions- the method which (in the context of discovering moral truths) Rawls (1972:18) called 'reflective equilibrium'. Although Rawls may seem not to have understood his method in quite this way (but see the last paragraph of this paper), it is a method which enables us to discover some general principle governing the correct use of some type of sentence. The method assumes that the simplest account of the use of sentences of that type in various narrowly described sets of circumstances is most probably the account which describes how such sentences would most probably be used in all circumstances, and so most probably shows the logically necessary and/or sufficient conditions for the truth of a sentence of that type. Such an account is simple insofar as it uses few predicates designating properties easily recognisable in many different kinds of paradigm example - a condition which I will illustrate with examples.

The application of this method to discover the logically necessary and/or sufficient conditions for the truth of a sentence of

some type has been a prominent feature of 'analytic philosophy'. Consider the procedures whereby philosophers have tried to set out the logically necessary and sufficient conditions of 'S remembers having done X'. A first suggestion by Hume (1888: Book1,Part1, section 3) was (in effect) that this was equivalent to 'S has a "lively" mental image of having done X which "preserves the original form" of S's awareness of having done X'. But then, it was objected, innumerable examples of usage showed that we allow that someone might have 'remembered' having done some action, even if they did not have a mental image of having done it. So it was suggested that almost all examples of usage showed that in order to 'remember' having done X, someone needs merely a true belief that he had done X. But then philosophers pointed out that we would not count someone as having 'remembered' having done X merely because they had acquired the belief that they had done X from having read in a book that they had done X. So then it was suggested that for S to 'remember' having done X, his belief that he had done X must have been caused by him having done X. Then it was pointed out that even if S having done X caused him to write it in his diary that he had done X, and his reading this in his diary caused him to believe that he had done X, we still 'would not say' that he 'remembered' having done X. Only if the route of causation went directly through S's body (in effect, his brain), and only if the belief was a basic belief not inferred from anything else, would the resulting belief count as a memory<sup>9</sup>. And so an account of memory began to emerge along the lines of 'S remembers having done X iff S has a true basic belief that he did X, caused by a chain of causes in S's body, itself initiated by S having done X'. This account was supposed to be logically necessary, and could be used to show sentences about personal memory to have various logical statuses- e.g. that 'if S remembers having done X, then S has a true belief that he did X' is logically necessary. Similar processes led to accounts of the logically necessary and/or sufficient conditions of 'S knows that p', 'S perceived O', and such like.

In each case a simple extrapolation from a few examples was seen as giving a probably correct account of some logically necessary or sufficient condition for the truth of some type of sentence, until investigators came across cases where a sentence of that type was clearly not being used in a way which satisfied that condition, and so they sought a different account. In seeking such an account, no one suggested that the only exceptions to a previous account were the exact narrowly described examples which did not satisfy the account. No one suggested that the only exception to the principle that for S to remember having done X, his belief that he had done X must have been caused by having done X 'unless the causal route proceeded through S's diary'. Why no one suggested that was because it would have been far too ad hoc and so too

complicated a principle. Rather, in both of these cases philosophers tried to discover a new account using few predicates designating properties easily recognizable in many different kinds of paradigm example (such as 'belief', 'true' 'cause' etc.) which fitted both the old and the new examples.

However an account of the logically necessary or sufficient conditions for the correct use of a word or sentence form, which captures almost all our usage but would need to become very much more complicated in order to capture one particular usage of that word or sentence form would still be recognised as the correct account of the normal meaning of the word or sentence form, and the odd usage would then be considered one which uses the word or sentence form in a different sense from the normal sense. Thus someone may sometimes be said to 'remember' or 'know' something which didn't happen, just because they claim to remember or know it in the normal sense; and then they are said to be using these words in an 'inverted comma sense'.

The simplicity of a general principle underlying the paradigm examples of the application of a word may be a matter, not of the fewness of the predicates involved (as in the previous examples) but of the 'easily recognizable' character of a property expressed by a single predicate. Suppose that the only red objects language users have seen are ripe strawberries, ripe plums, London buses, rubies, holly berries, and pink roses; and they are taught that these are paradigm examples of 'red' objects. Then, given that they can easily recognize the similarities of colour to each other of these different paradigm examples, they would see (as the simplest account of its meaning) the logically necessary general principle governing the use of 'red' as 'an object is red iff it is similar in colour to actual ripe strawberries, ripe plums, London buses, rubies, holly berries, and pink roses. This in turn will lead them to conclude, when they come to observe what we would call different shades of red and name them 'scarlet' or 'crimson', that such sentences as 'if it scarlet, it is red' are logically necessary and so express a logically necessary proposition. This example illustrates the point that the ability of language users to recognize necessary truths depends on the kinds of similarities between phenomena to which they are sensitive. If they could not recognize any similarities of colour between the paradigm objects, they would conclude that there are six different unconnected senses of 'red'. From that it would follow that the sentence 'if it scarlet, it is red' is logically impossible, and so expresses a different proposition from the one expressed by the previous group of language users.

The third aspect of my definition of the simplicity of an account of the logically necessary and sufficient conditions for the correct use of a word or sentence form is that it uses predicates which

designate properties recognizable in *many* different kinds of paradigm example. If language users were given only two kinds of paradigm example of red objects, for example rubies and ripe plums, it would not be obvious whether only dark red objects count as 'red' or whether all of what (we call) red objects count as 'red'. Neither extrapolation from the paradigm examples would be especially simple. Likewise if we had only a few paradigm examples of 'S remembers having done X' or 'S knows that p', there would be many different but not especially simple ways of extrapolating from the examples to a general logically necessary principle about knowledge. The evident simplicity of one account becomes apparent only when the semantic rules provide many different kinds of paradigm example of correct application.

So far I have been assuming that all speakers of a language would regard the same examples of words and sentences as paradigm examples of their correct application, that is ones such that it would be inconceivable to suppose that they did not apply to almost all of those examples. But of course most speakers will have been taught the meanings of words and sentences from different paradigm examples, and sometimes from very different kinds of examples. That may or may not lead them to acquire common concepts, expressed by inter-translatable words, depending on whether the concept which some speakers derive from one set of paradigm examples is one which they recognize as applying to the examples from which other speakers have derived their concept; which is to say, if reflective equilibrium leads to agreement about the different examples.

Suppose that there are two communities isolated from each other. One community lives in an environment where every object is of a different shade or mixture of blue or green: light blue, dark blue, blue-green, light green, dark green etc. The community recognizes that there is something common to those different shades and they call it being 'colorato'. The other community lives in an environment where every object is of a different shade or mixture of red or green: light red, dark red, red-green, light green, dark green etc. They too recognize that there is something in common to those different shades, and they call it 'coloré'. Suppose now that the two communities become acquainted with each other and their environments. Then it may happen that each community would come to recognize that what their objects of different shades have in common is the same kind of thing as what the objects of different shades of the other community have in common, which kind of thing we call being 'coloured'. So the community would be recognizing that their understanding of their general term ('coloré' or 'colorato') is that it applies not merely to objects of shades in the environment of the other community already known to them, but also to ones not

previously known to them. If that is what happens, the community would be showing that there was implicit in their previous understanding of 'coloré' or 'colorato' a natural fit to wider examples. If the first community came to recognize this, it would come to recognise that the sentence 'if it's light red [or whatever they came to call that colour], it is colorato' is logically necessary and so expressed a logically necessary proposition; and if the second community came to recognize this, it would come to recognize that 'if it is dark blue, it's coloré' expressed a logically necessary proposition. Then it would have turned out that 'colorato' and 'coloré' mean the same. Yet the processes by which the two communities originally learnt the concepts of 'colorato' and 'coloré' would not have included teaching them these logically necessary truths.

On the other hand it might happen that one or both communities did not come to apply their words 'coloré' or 'colorato' to the new shades. This would show that implicit in their previous understanding of those words was a much narrower understanding of them than as 'coloured' (in our sense). In that case they would have to claim that sentences such as 'if it's light red, it is colorato' expressed, not logically necessary but logically impossible propositions. Which outcome would result would depend on the kinds of sensitivity to colour which the teaching process has given them. Clearly the greater the overlap between the paradigm examples of 'colorato' and 'coloré' objects, the more likely it is that each group would come to see these words as meaning the same.

In order to discover the meanings of colour words and so the logical relations of colour concepts to each other, and indeed the meanings of any words which have semantic rules for their application and so the concepts which they express, communities will need in practice to have observed actual examples of differently coloured objects (or whatever). But the conclusion which they reach by reflecting on the concepts which they acquire is a conclusion which is logically independent of whether there are any objects which have these colours (or whatever) , let alone whether they know this; and so, like the conclusion about the nature of remembering, it is an a priori conclusion<sup>10</sup>. The moral of this story is that consideration of new examples, actual or imaginary, can get us to recognize previously unrecognized modal truths about old words used in the same sense and so expressing the same concepts as previously.

### III

It is well known that most attempts to show the truth of moral principles by the direct methods of deriving logically necessary truths, described in the previous section, are seldom successful for the kinds of reason that some attempts to derive logically necessary truths of other kinds also fail - for example, because disputants do not agree about what mini-entails what or when some sentence is obviously inconceivable. However many philosophers recognize the method of 'reflective equilibrium' as a method of helping us to discover (probably) true moral principles<sup>11</sup>. I now illustrate this, and show that the reason why this method works for moral principles is the same kind of reason as the reason why it works to show the necessary truth of non-moral sentences. That makes it very plausible to suppose that the correct moral principles are themselves logically necessary truths.

Most children, although brought up in very different cultures, quickly acquire a concept of a kind of action which it is important to do, and of a kind of action which it is important not to do; and such that some of the former are more important to do than others, and some of the latter are more important not to do than others. They are praised and sometimes rewarded for doing actions of the former kind, and blamed and sometimes punished for doing actions of the latter kind. They thus acquire a concept of 'moral goodness' which they recognise as applicable to all actions of the former kind; and a concept of 'moral badness' which they recognise as applicable to all actions of the latter kind, concepts which are such that to have a belief that some action is morally 'good' (or 'bad') entails having some motivation to do (or not to do) the action. But they don't understand these concepts of moral goodness or badness as concepts of properties connected merely contingently with the concepts of the particular narrowly described kinds of actions which exemplify them. Rather, they understand the former concepts as determinable concepts of which the determinates are the different kinds of morally good or bad actions, just as 'coloured' is a determinable of which red, green etc are determinates. So they would regard it as inconceivable to suppose that almost all of the kinds of actions to which they originally applied the concept of moral goodness (or badness) are really not morally good (or bad). Someone who does allow that possibility for his concept has not got the concept which most children acquire and which I am discussing. For there would be then no semantic criteria for the application to actions of 'morally good' (or whatever). And some of the syntactic criteria, illustrated above, that a 'morally good' action is one which is 'important' to do, and 'gets praised' can themselves only be understood if we understand 'important' as 'morally important',

'praised' as 'recognised as morally good', and so on. These moral terms can only be understood in terms of each other, unless they are linked to paradigm examples which illustrate the kind of importance which makes an action morally good, the kind of praise which recognises a morally virtuous action, and so on. Otherwise 'importance' could be prudential importance, 'praise' could be praise for doing what the praiser likes to see done, and so on. While understanding moral terms *merely* in terms of action-guiding or similar notions, is incompatible with moral realism, without the tie to action guiding having a moral belief would just be *merely* having a belief that there are similarities of certain recognizable kinds between examples of morally good (bad) actions- which yields an implausible moral naturalism.

Further, the concepts of moral goodness, badness etc. which most children acquire is surely not one which they regard as applying merely to actions of narrow kinds, similar in almost all respects to those from which they originally learnt to apply them. For, at least as they get older, children are open to reflection on what they have been taught, to argument with members of other cultures, and to experiences of new situations, leading them to extend or retract the application of these concepts. Clearly use of reflective equilibrium makes it possible for two groups who largely agree with each other about moral matters but have a few disagreements, to reach agreement about disputed matters. They may both recognise a simple principle underlying some set of examples which both groups recognize as paradigm cases of 'morally good' actions for which praise is appropriate, which has the (probably true) consequence that one kind of action which only one group recognizes as a paradigm case of a 'morally good' action is not like the others; rather it is a kind of action to which praise is not appropriate, and so is not 'morally good'.

Here is an example where the simplicity of a principle depends on the fewness of the predicates designating properties which it lists, as with analyses of 'remember' and 'know'. Suppose that both groups have learnt that it is wrong for a person A to kill a person B unless B is about to kill A or some member of A's group, or A is acting on behalf of a court of law which has sentenced B to death for murdering a member of the group, or B is an enemy combatant in war with another group trying to kill a member of A's group. But one group, unlike the other group, may also have been taught that it is not wrong for A to kill B in a duel with B if B has insulted A or a relative of A. Then the first group may point out to the second group that all the examples except the disputed one are examples of situations in which B has killed someone close to A or is about to do so. So the first group suggests that a simple principle underlying all the agreed examples is 'it is wrong to kill except in revenge for a

killing or to prevent further killing (of a member of one's group)'. This simple principle treats life as so sacred, that it must not be taken away except in compensation for life or to prevent further loss of life. The consequence of this principle is that it is wrong to kill in a duel merely to defend your family's honour. The second group may then recognise this principle, and so move on from what they have been taught when young, that killing in a duel is not wrong. As with 'know' and 'remember' investigators come to see a previous paradigm example as not an example of the same concept as the other paradigm examples. And of course reflective equilibrium may eventually lead both groups to develop an even simpler principle, 'it is wrong to kill except to prevent further killing.'

Moral disagreement may turn, not on whether a suggested principle uses few predicates, but - as with 'red' - on whether the predicates which it uses designate properties easily recognizable in many paradigm examples. I take as an example of this a case where there is moral disagreement between two groups because, although both groups agree that the moral goodness or badness of some action depends on certain properties which it has, and they also agree about which of these properties make for the overall goodness of the action and which make for its badness, they disagree about which group of properties outweighs the other group. Each disputant group agrees that the considerations adduced by his opponent have some force; that is, would for example show the action to be morally good overall but for the considerations which it adduces against its moral goodness. Thus a group opposing the euthanasia involved in helping a depressed person (who has no close friends or family) to commit suicide may argue that such an act is overall bad because of the sanctity of human life, the possibility of helping a depressed person to recover from his depression, the value of his overcoming that depression, and so on. A group advocating euthanasia may argue that helping a depressed person to commit suicide is helping him to do what he clearly and firmly wants to do and what hurts no one else in any way. Both disputant groups appeal to considerations which the other group may admit to have some weight, although each holds that the considerations which it adduces outweigh those which the other group adduces. The consequence of this kind of disagreement is that for one group to persuade the other to change its view will only require them, on the basis of already agreed examples, to give a bit more weight to a feature which they already admit to have some weight; they would not need to persuade the other group to admit the relevance of some new good-making or bad-making feature.

So the group opposing euthanasia might persuade the other group that the simplest principle underlying their many shared beliefs about it being good to help the sick, the unemployed, the

unlovable etc is that it is always good not to 'give up' on helping people satisfy their basic *needs*. Hence they might persuade the group supporting euthanasia that this principle is probably true and so has the consequence that one should not give up on helping the depressed to satisfy their basic need for happiness, and so on helping them to recover from their depression; and so that one should not help anyone to commit suicide. Alternatively the group supporting euthanasia might persuade the group opposing it that the simplest principle underlying their many shared beliefs about it being always good to help those who want to get an education, to have children, to travel abroad is that it is good to help people to do what they firmly *want* to do, when that hurts no one else; and so persuade them that this applies to helping people to commit suicide.

The issue may then turn on how easily recognizable are the properties designated in rival principles. It is often very obvious what a person 'wants' (that is, desires), but the notion of what a person 'needs' may often be very unclear; the latter depends on a view about what constitutes human flourishing, about the application of which even its defenders may be unclear. The issue will also depend on how many are the different kinds of paradigm example which rival principles fit. If there are a vast number of different paradigm examples where, disputants admit, it is good to help people to do what they want, and relatively few paradigm examples where it is good to help people satisfy their basic needs; and especially if there is a paradigm example of the former where 'wants' trump 'needs', that shows that 'wants always trump needs' is the simpler moral principle. This parallels the 'colorato' case, where whether a red-green object is deemed to be 'colorato' will depend on how easy it is to recognize similarities between red-green etc objects and the previous paradigm examples of 'colorato', and how large and varied is that stock of previous examples. The more varied are the previous paradigm examples, and the easier it is to recognize the similarities between them and the new examples, the more evident it will be that the latter are 'colorato'. In the moral case, as in the colour case, it requires a common kind of sensitivity to paradigm examples, whether colour sensitivity or moral sensitivity, to recognize the relevant similarities and so to reach agreement – even if in the end the agreement is that the two principles are equally simple, and so when they clash it is equally good to give priority to wants as to give priority to needs.

Clearly, just as the more two groups share common paradigm examples of 'colorato' and 'coloré', the more probable it is that use of the method of reflective equilibrium will lead to agreement on the logical truths about colour, so the more shared examples of 'moral goodness' two groups have, the more probable it is that use of the method of reflective equilibrium will lead to agreement on moral

principles. As with the colour example, it is often in practice necessary for someone to have had personal experience of some situation in order fully to understand what is meant by someone doing some kind of action or having some kind of experience. One might need to experience acute depression oneself in order to understand that it is a necessary moral truth of the form that it is good to cure acute depression by means of drugs which will not cause pain, delusion, or mental impairment. But again -if this is a moral truth - it is an a priori one, because it is irrelevant to its status as a truth whether anyone ever does suffer from acute depression.

But surely, an objector may say, there is a big difference in the way people react to attempts to get them to change their views under the pressure of arguments appealing to reflective equilibrium, when the arguments concern moral principles from the way they react when the arguments concern ordinary (that is, non-moral) conceptual truths - the difference being that many people do not change their views about moral principles readily, whereas they do change their views about ordinary conceptual matters readily. And that suggests that there must be an important difference between the logical nature of conceptual truths and moral principles. There is however a ready explanation of the former difference which has nothing to do with the logical character of the propositions, but everything to do with the nature of humans, which makes it difficult for us to change our moral views in the light of good arguments. Our moral views are closely connected to our emotions, some of which hinder us from recognizing moral views which otherwise we would naturally adopt. The natural emotion of anger at those believed to have molested children may lead us not to recognize their right to a fair trial, and so not to adopt the principle that everyone is entitled to a fair trial. Our moral views are also closely connected with rival desires, some of which are desires to do actions which, if we accepted certain arguments we would have to recognize as morally wrong and so would find ourselves under internal moral pressure not to do. If I conclude that there is a true moral principle that everyone ought to donate much of their income above a certain amount to providing food and shelter to those in distant countries, I may have to conclude that I ought to give a lot, when I desire not to do so. So again I refuse to acknowledge a strong argument in favour of there being such a moral principle. Humans are only partly rational creatures. It is for this reason that questions about whether some suggested moral principle is a true moral principle remain 'open questions' for a long time. So there is no need to attribute the 'openness' of such questions to a difference between the logical nature of conceptual truths and moral principles.

But even if groups who have fairly similar moral views can resolve their differences by reflective equilibrium, what reason do

we have for supposing that groups whose views are much further apart from each other can do so? My response is (tentatively) to put forward a contingent hypothesis that most twenty-first century groups have enough moral views in common to make agreement between them possible - given enough time and willingness to understand each other. Almost all groups agree that it is good (though not always obligatory) to feed and otherwise care for their own (non-disabled) children and (mentally competent) parents and also some other members of their own group who need food or care, to keep their promises (not made under duress) except when very difficult to do so, not to steal what by community agreement belongs to another member of the group, and so on. Not merely are there some paradigm examples of moral goodness or badness on which almost all agree, but - I suggest - there is a chain between any groups of persons diametrically opposed on many moral issues, of groups who agree with one extreme group on most issues, different groups who agree with the semi-extreme group on most issues, and so on until we come to the group at the other extreme. When reflective equilibrium has moved someone away from an extreme position to a slightly less extreme position, that person will regard many of the new examples of morally good and bad actions which he comes to recognize as examples of the relevant concept just as satisfactory as the ones recognized previously and so as paradigm examples of such actions. And then that person will have much more in common by way of overlapping sets of paradigm examples with some others with whom previously he had less in common, which will provide a basis for reaching agreement which did not exist previously. And so it is possible for someone to move by a rational process consisting of many such steps from one general moral outlook to a quite different one, say from a narrow tribal understanding of which actions are overall good to a far wider one - although they will continue to hold some of the views about the goodness of helping members of their own tribe, which provided their original paradigm examples of good actions. My ground for putting forward my contingent hypothesis is that over many centuries the human race has made much progress in coming to share moral views. No longer do most humans regard slavery or duelling, let alone human sacrifice or suttee, as morally permissible. And that suggests that most of us do share a common sensitivity to what paradigm examples of 'morally good' (or whatever) actions have in common, and so a common concept of morality, the logically necessary truths of which can be discovered by a priori reflection, although it may take a long time for all of us to agree on what these true fundamental moral principles are.<sup>12</sup>

Of course it may sometimes be the case that those to whom new examples are presented and apparently simpler principles of

morality are suggested, show no tendency at all (even after prolonged exposure to them) to move towards agreement with most of us. That would suggest that these people do not have the same concept of 'morality', the same understanding of what it is for an action to be 'morally' good, bad, or whatever, not merely that they do not have the same views as most of us about which actions are morally good - just as we would reach a similar conclusion about those exposed to new colours who did not recognize them as 'colorato'. The process of acquiring concepts of 'morally' good, bad etc, would have produced in those uninfluenced in the ways illustrated by moral arguments a different concept of 'morality' from what it has produced in most of us. But for the reasons given above, I do not think that that is often the case. And how little tendency to change 'moral' views after how much exposure to the 'moral' views of others suffices to show that someone has a different concept of morality? 'Having the same concept of morality' is, like most linguistic expressions, vague, and so there are border-line cases for its application. But there are plenty of definite cases of people who used to disagree about 'moral' matters, yet are shown by their subsequent agreement resulting from reflective equilibrium to have had the same concept of 'morality', a concept possessed, I have suggested, by most of the human race. I have sought to show that - given moral realism and so given that there are true moral principles - the similarity of procedures by which we establish what are the true moral principles to those by which we establish what are the true non-moral conceptual principles, makes it very plausible to suppose that both kinds of principle are conceptually and so (in my sense) logically necessary - a result which alone, given the supervenience of the moral on the non-moral, is needed to make moral realism plausible.

The unwillingness of many moral realists to recognise that there can be logically necessary moral truths seems to me to arise from failure to recognize the importance of semantic rules as well as syntactic rules for the use of words, in determining the logical status of all sentences, and so of the propositions which they express; and a failure to recognize that the application of semantic rules requires a certain kind of sensitivity to similarities between the paradigm examples. There is no difference between non-moral conceptual necessities and moral principles in these respects. 'Anything red is coloured' is surely logically necessary, but only because the meanings of both 'red' and 'coloured' are fixed largely by paradigm examples, and language users recognize the same similarity between the paradigm examples of red things to which they are introduced, and the same similarity between the paradigm examples of coloured things to which they are introduced. and they recognize the former kind of similarity as a species of the latter. And since in general

speakers of languages other than English recognize the same colour distinctions as do English speakers, there are natural translations of 'red' and 'coloured' into their language which have the consequence that a sentence of their language means the same as, and so expresses the same logically necessary proposition as, 'anything red is coloured'. So too, I have been suggesting, logically necessary moral sentences, and so the propositions which they express, are logically necessary because the meanings of 'morally good, 'morally obligatory' etc. are fixed partly by semantic rules, that is by paradigm examples, to the similarities between which most of us have a similar sensitivity. Rawls may have recognised this when he wrote (1972: 44-5): 'It is obviously impossible to develop a substantive theory of justice founded solely on truths of logic and definition. The analysis of moral concepts and the a priori, however traditionally understood, is too slender a basis'. In other words, syntactic rules are not enough. However, he continued, 'If we can find an accurate account of our moral conceptions' which I take to mean that if we take account of the moral judgements we actually make, 'these questions of meaning and justification may prove easier to answer' As I put it, semantic rules fill out what is meant by moral concepts, and thereby (by the method of reflective equilibrium) enable us to determine their applicability to actions of different kinds.

Oriel College, Oxford

## REFERENCES

1. Blackburn, S (1993) 'Supervenience Revisited'. In his *Essays in Quasi-Realism*, (Oxford: Oxford University Press) pp.130-148.
2. Bonjour, L (1999) Article on 'A priori'. In R. Audi (ed.) *The Cambridge Dictionary of Philosophy*, second edition (Cambridge: Cambridge University Press) p.35.
3. Boyd, Richard (1988) 'How to be a Moral Realist'. In G. Sayre-McCord (ed.), *Essays on Moral Realism* (Ithaca, N.Y. Cornell University Press) pp.181-228.
4. Brink, David (1989) *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
5. Chalmers, David (2002) 'Does Conceivability entail Possibility?' In Gendler, T.S. and Hawthorne, J. (eds.) *Conceivability and Possibility* (Oxford: Oxford University Press) pp.145-200.
6. Cuneo, Terence (2007) *The Normative Web*. Oxford: Oxford University Press.
7. Dancy, Jonathan (2004) *Ethics Without Principles*. Oxford: Oxford: Oxford University Press.
8. Fine, Kit (2002) 'The varieties of Necessity'. In Gendler, T.S. and Hawthorne J. (eds.), *Conceivability and Possibility* (Oxford: Oxford: Oxford University Press) pp.253-281.
9. Horgan, Terence and Timmons, Mark (1992) 'Troubles on Moral Twin Earth: Moral Queerness Revised', *Synthese* 92 (2), 221-260.
10. Hume, David (1888) *A Treatise of Human Nature*, (ed.) Selby-Bigge, L.A. Oxford: Oxford University Press.
11. Kim, Jaegwon (1993), "'Strong" and "Global" Supervenience Revisited' In his *Supervenience and Mind* (Cambridge: Cambridge University Press), pp.79- 91.
12. Kripke, Saul [1980] *Naming and Necessity*, Oxford: Blackwell.
13. Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin.
14. Martin, C.B. and Deutscher, Max (1966) 'Remembering'. *Philosophical Review*, 75, 161-196.
15. Parfit, Derek (2011) *On What Matters*. Oxford: Oxford University Press.
16. Putnam, Hilary (1975) 'The meaning of "Meaning"' In his *Mind, Language and*
17. *Reality, Philosophical Papers*, vol. 2, (Cambridge: Cambridge University Press) pp.215-271.

18. Rawls, John (1999) *A Theory of Justice*, rev. ed. Oxford: Oxford University Press.
19. Scanlon, Thomas M. (2002) 'Rawls on Justification'. In S. Freeman (ed.) *The Cambridge Companion to Rawls*, (Cambridge: Cambridge University Press) pp. 139-167.
20. Shafer-Landau, Russ (2003) *Moral Realism: A Defence*. Oxford: Oxford University Press.
21. Smith, Michael (1994) *The Moral Problem*. Oxford: Blackwell.
22. Swinburne, Richard (2013) *Mind, Brain, and Free Will*, Oxford: Oxford University Press.
23. Wedgwood, Ralph (2007) *The Nature of Normality*, Oxford: Oxford University Press.
24. Yablo, Stephen (1993) 'Is Conceivability a Guide to Possibility', *Philosophy and Phenomenological Research*, 53, 1-14.

## NOTES

1. Thanks to anonymous reviewers for helpful comments on a previous version of this paper, and to Roger Crisp for comments on a very early version.
2. I am concerned in this paper solely with propositions which ascribe 'thin' moral properties to possible or actual actions, that is properties normally expressed by a predicate consisting of the adverb 'morally' and a general evaluative predicate 'good', 'bad', 'obligatory' or 'wrong' or ones definable thereby; and with propositions which ascribe them on grounds independent of any agent's ability to do them. For example, I am concerned with actions which are good, even if there are no agents able to do them, and obligations which agents still have, even if they are unable to perform them and so are not culpable for not doing so. So I do not presuppose that 'ought implies can'.
3. The doctrine of supervenience can be stated in a more general form, so as not to entail moral realism, as by Blackburn (1993).
4. I take no stance on whether the only true propositions of the form 'it is metaphysically necessary that any action of kind A is morally good [or whatever]' are ones where 'A' specifies all the properties possessed by any action of that kind (apart from those concerned with some agent's ability to do it) - a conjunction so large that it could never be put into words; or whether -to my mind, far more plausibly- there are many true propositions of this kind where 'A' specifies only a small subset of such properties. Both kinds of proposition seem to deserve the name of 'principle'. My main claim concerns 'supervenience', not 'resultance'. The moral properties of an action of some kind may 'result from' it having certain positive non-moral properties which 'give it' the moral property, because in some sense they alone are positively relevant to its having that property. These are only a sub-set of the non-moral properties on which the moral property supervenes; among other properties are negative properties (the non-occurrence of properties whose occurrence would cancel the effect of the positive non-moral properties.) On this distinction between supervenience and resultance, see Dancy (2004: 38-52). Dancy includes among the 'other' properties 'enablers' (properties concerned with whether a person is able to do some action; but these are only relevant to kinds of propositions with which (see note 1), this paper is not primarily concerned.

My main thesis does nevertheless, I believe, have the consequence that, given a satisfactory account of what it is for non-moral properties to 'give' an action some moral property, propositions asserting which moral properties 'result from' which non-moral properties are also logically necessary.

5. Classically J.L. Mackie (1977: 41) wrote: 'What is the connection between the natural fact that an action is a case of deliberate cruelty – say, causing pain just for fun – and the moral fact that it is wrong? It cannot be an entailment, a logical or semantic necessity. Yet it is not merely that the two features occur together. The wrongness must somehow be "consequential" or "supervenient"; it is wrong because it is a piece of deliberate cruelty. But just what in the world is signified by this "because"?' Horgan and Timmons (1992) endorse this as a crucial objection to moral realism.

6. Russ Shafer-Landau (2003: 87-98) compares the relation of the moral to the natural (or physical) to the relation of 'chemical' to 'atomic' facts. David Brink (1989: 179) also claims that 'moral facts and properties are just one kind of higher-order fact or property'. He compares 'moral facts are constituted by natural facts' to 'tables are constituted by certain arrangements of microphysical particles'. Such examples are just like the 'water is H<sub>2</sub>O' example. Both writers seek to bolster their case by comparing moral supervenience to the a posteriori metaphysical supervenience alleged by physicalism of the mental on the physical.

But even if physicalism were true, that would show only that there is a large class of a posteriori necessary truths, not that moral principles belong to that class. That needs to be shown independently. Michael Smith (1994: 190-192) claims that just as we can know it is a necessary truth that the property A of a surface which actually (in our world) causes objects to look red does so, so there are necessary truths about the natural properties on which moral properties supervene. Strangely, he recognizes that the supervenience of redness on the reflecting properties of surfaces is a posteriori, whereas the supervenience of the moral on the non-moral is 'knowable a priori'; but he seems to think that this 'point of disanalogy' is relatively unimportant.

7. Parfit (2011: 524-5) also compares normative truths to 'the necessary truths of logic and mathematics.'

8. Thus Chalmers (2002: 153): 'A situation is [modally] coherently imagined where it is possible to fill in arbitrary details in the imagined situation such that no contradiction reveals itself...I will say that S is positively conceivable when it is coherently modally imaginable.' Although sentences containing category mistakes and other sentences which we can determine to be false solely a priori, do, I believe, 'reveal', that is 'entail' a contradiction, my wider definition of conceivability does not assume this.

9. This is a rough summary of the conclusion of Martin and Deutscher (1966).'

10. 'It is no objection to a claim of a priori justificatory status for a particular belief that experience is required for the acquisition of some of the constituent concepts' (Bonjour [1999] 35).

11. Thus Scanlon (2002: 149) regards reflective equilibrium as 'the only defensible method of making up one's mind about moral matters and about many other subjects. And (148) he sees it as a method of getting us to recognize the correct principles and correct particular judgments: 'the fact that a given judgment does not fit with principles that account for most of our other judgments can lead us to change our mind about that judgment itself, and we may also be led to change our mind when we see that the only principles that *do* account for a given judgment are ones that are seen in other ways to be clearly mistaken'.

12. One writer who emphasizes that although cultures start from different points, new experiences and the operation of reflective equilibrium lead to moral progress is Richard Boyd (1988). But Boyd seems to hold that in this process our concept of moral goodness itself changes, in the way that scientific concepts (such as the concept of a biological kind) change. The view which I have been advocating is that most of us have a common concept of moral goodness, and that progress normally consists in seeing that its extension is different from what we had believed hitherto.