

This paper appeared in the *Journal of Philosophy* CI: 10 (October 2004) 499-526.

## Moral Rationalism<sup>1</sup>

Basic moral principles are known to us a priori. I will be arguing for this claim, trying to say what it means, and discussing its ramifications.

The claim that basic moral principles are a priori was emphasized by Leibniz and, on some natural readings of the texts, endorsed by Kant.<sup>2</sup> Even a self-proclaimed empiricist like Locke sometimes veered towards endorsing this claim of a priori status.<sup>3</sup> Yet the character of this a priori status, and its significance for the epistemology and metaphysics of moral claims, have both been very largely lost in recent discussions of moral thought. I will be arguing that the nature of this a priori status is incompatible with subjectivist, judgement-dependent and mind-dependent treatments of moral thought. Part of the task in establishing this incompatibility is to articulate more precisely the kind of a priori status that is in question here. It is easy to underestimate the problem for mind-dependent theories of moral thought if one starts by understating the sense in which basic moral principles are a priori.

If basic moral principles are a priori in a way that is incompatible with mind-dependent treatments, various tasks become pressing. One task is to develop a conception of the metaphysics and epistemology of morals that respects this status. Another is to address some of the motivations that have made mind-dependent views of this territory so tempting. This is evidently not a task for just one paper. But after attempting to make out

the case against mind-dependent theories, I will try to outline some possible directions of development; and also to identify something I will call “the Subjectivist Fallacy” which can make mind-dependent views of morality seem more attractive than they really are.

One can pursue these questions about the a priori status of basic moral principles as issues of interest in their own right in the subject of morality and its epistemology. But the questions also have a wider significance. The case of moral thought is of interest as a test case for anyone sympathetic to a more general program of moderate rationalism.<sup>4</sup> Moderate rationalism seeks to explain all cases of a priori knowledge by appeal to the nature of the concepts that feature in contents that are known a priori. For the moderate rationalist, the explanations of a priori knowledge in various domains will not involve the postulation of causal interactions with non-physical or non-mental realms. That is what makes it a moderate rationalism. The explanations will also treat the a priori ways of coming to know as rational, as an exercise of reason. That is what makes the moderate position a form of rationalism. What I have to say in this area can be seen as some early steps towards carrying through the moderate rationalist’s program in the special case of moral thought. I hope that some of the considerations I offer will be of more general application, and will help in the development of a moderate rationalism in other areas.

### *1. The Claim of A Priori Status*

Here is a first formulation of the claim of a priori status:

Every moral principle that we know, or are entitled to accept, is either itself a priori, or it is derivable from known a priori moral principles in conjunction with non-moral propositions that we know.

For an illustration of this Initial Thesis, consider the moral proposition that national high school examinations which assume that candidates have first-hand knowledge of vocabulary needed in snowy climates are unfair to those who live in southern states. That is not itself an a priori principle. No amount of a priori reflection would succeed in excogitating it. The moral proposition does, however, follow from two other truths: from the a priori principle that fair examinations will not include questions requiring background knowledge likely to be absent in one geographical group, together with the empirical, non-moral fact that it rarely snows in the southern states.

The Initial Thesis implies that for any moral proposition we are entitled to accept, there is a similar division: into its a priori moral grounds on the one hand, and its a posteriori non-moral grounds on the other. What the Initial Thesis excludes is the irreducibly a posteriori moral ground. The Initial Thesis is in the spirit of, indeed is a formulation of, Kant's claim that "all moral philosophy is based entirely upon its pure part".<sup>5</sup>

Why should we believe the Initial Thesis? All sorts of heavy-duty theories - theories of the a priori and theories of morality - might be offered in its support. I shall be touching on, and endorsing, some of them later. But the primary reason for accepting the Initial Thesis is not theoretical at all. The primary reason rests on the consideration of examples. Consider your belief that prima facie it is good if the institutions in a society are just; or your belief that prima facie it is wrong to cause avoidable suffering; or that

prima facie, legal trials should be governed by fair procedures. These beliefs of yours do not, and do not need to, rely on the contents of your perceptual experiences, or the character of the conscious states you happen to enjoy, in order for you rationally to hold them. Understanding of what justice is, of what pain and suffering are, of what a trial and what fairness are, makes these several beliefs rational without justificational reliance on empirical experience. Experience, as Kant said, may be necessary for the acquisition of these concepts, but that does not mean there cannot be propositions involving them that are a priori. Nor is it clear how empirical experience could rationally undermine these beliefs. Empirical information about extraordinary circumstances might convince us that it would be better on this occasion that a trial not be fair. That would not undermine the proposition that prima facie trials ought to be fair; and it is not clear what could. Take any other moral principle that you are entitled to accept. I suggest that on examination, it will always involve an a priori component, in the sense employed in the Initial Thesis.

The epistemic situation in the case of moral principles seems to me broadly similar to that concerning the status of logic and arithmetic. All sorts of heavy-duty theories – philosophical theories about logic, arithmetic and the a priori - can be offered to support the view that logic and arithmetic are a priori. Those theories may or may not be convincing, but they could not be more convincing than the evidence they attempt to explain, such facts as that we are, apparently, justified in accepting that  $2+2=4$ , or that  $A \vee B$  follows from  $A$ , without justificational reliance on the content of our perceptual experiences, or other conscious states. In both the moral and the arithmetical and logical cases we must of course be prepared for the possibility that these appearances of a priori status are misleading. Anyone who defends the Initial Thesis must address all sorts of

challenges, not all of which I can consider here. All I am emphasizing at this point is that there is strong prima facie support for the Initial Thesis from consideration of examples, in advance of any detailed philosophical theory of how or why the Thesis holds.

The Initial Thesis is neutral on the question of whether every true moral principle could be known by us. People who disagree about that could both accept the Initial Thesis. The Initial Thesis concerns only the cases in which a principle is known, and says something about the existence of a priori ways of coming to know the principle.

This does not make the Initial Thesis a mere de facto claim about the moral principles we happen to know. The reasons for accepting the Initial Thesis go beyond what is provided by inspection of the particular moral principles we actually accept. I will be offering some general grounds for the Initial Thesis that are not dependent upon the particular moral principles we are currently entitled to accept. There is some plausibility in the further claim that the Initial Thesis, if true at all, is itself a priori. In any case, it has the status of a philosophical, not an empirical, claim.

For those who think that it is begging too many questions to formulate a thesis in a form that presupposes the possibility of moral knowledge, we could frame a version, which may be more comfortable for those doubters, that mentions only entitlement to accept. (I myself doubt that this really is weaker, but I mention it so that we can focus on the essential issues.) Any interesting version of the Initial Thesis must, however, make some use of some distinction between proper and improper acceptance of a moral principle. It could not be formulated in terms of mere acceptance.

The Initial Thesis is cagily formulated using “we”. It will not be true of each individual thinker that every moral principle he is entitled to accept is either a priori, or

derivable from a priori moral principles and non-moral propositions he knows. Moral knowledge, like any other kind of knowledge, can be acquired by testimony. An empirical moral principle may be so acquired, and when it is, the acquirer himself need not know the a priori grounds of the empirical moral principle he learns through conversation. Nevertheless, someone must know or once have known them if the moral belief he acquires by testimony is to have the status of knowledge. The Initial Thesis is a thesis about actual epistemic grounds, in the epistemic community as a whole over time. The Thesis goes far beyond claims about the mere possibility of grounds.

What do I mean by “a priori”? For an intuitive, overarching characterization of a standard notion, we can say this: a thinker’s judgement is a priori if it has an operative justification or an entitlement that is independent of the representational content or kind of the thinker’s perceptual experience, and of her other current conscious states. So the judgement “There’s a window over there”, when the thinker makes it because he sees a window to be over there, is not a priori, because it endorses the content of the thinker’s perceptual experience. The judgements “I’m in pain” and “I’m imagining standing on a beach” are not a priori when the thinker’s operative justification or entitlement lies in the character of his current conscious states, his pain or his imaginings. In all these cases – of seeing the window, of the pain, and of the imagination – there is a way in which the judgement comes to be made and whose status as justifying or entitling is dependent on one or another features of perceptual experience, or of other conscious states. The way itself is not a priori, we might say. By contrast, judgements to which the thinker is entitled because the thinker, or someone else, has a proof of their contents are a priori by this umbrella criterion.

The umbrella characterization covers two fundamentally different species of the a priori. As I implied in the introductory remarks, it is important to distinguish them, both from each other and from related notions in the territory, if we are to have a clear view of the significance of the senses in which basic moral principles are a priori.

The two species of the a priori can be introduced by first considering a much more general auxiliary notion. This more general notion in its most abstract form stretches far beyond the a priori. It is the notion of a judgement with a given intentional content being true in any circumstances in which it is reached in a given way. A judgement “I’m in pain” that the thinker makes rationally because she consciously experiences pain falls under this general notion. In any circumstances in which a thinker comes to make the self-ascription of pain by rationally responding to her conscious experience of pain, her self-ascription will be true. A judgement of a logical truth reached by accepting a proof of it equally falls under the same notion. I label this very general notion that of  $p$ 's being *judgementally valid* with respect to a given way.

It is important that the judgemental validity of a content with respect to a given way turns only on the truth of the content in circumstances in which it is in fact judged (and reached in the given way). In assessing judgemental validity with respect to a given way, we do not have to consider whether the content is true in circumstances in which it is not reached in that way. Nor do we have to consider whether the content has any kind of necessity.

Various famous concepts in philosophy are variants of this core notion of judgemental validity. Descartes was particularly interested in those contents with the following property: that there exists a way with respect to which they are judgementally

valid, and which is indubitably so. Descartes' description of something that is "necessarily true whenever it is put forward by me or conceived in my mind" is a variant, with additional restrictions, of the core idea of the judgementally valid.<sup>6</sup>

We can make use of this auxiliary notion of the judgementally valid in distinguishing the two species of the a priori that I want to distinguish. The first notion of the a priori to be distinguished is simply a restriction of the notion of judgemental validity. I say that

*p* is *judgementally a priori* with respect to a way *W* just in case it is judgementally valid with respect to *W*, and the way *W* is an a priori way.

The judgementally a priori includes some classical self-verifying cases. When the content "I am thinking" is judged, not as a report on the thinker's own recent conscious states, but because the thinker appreciates, on the basis of his grasp of the concepts it contains, that it will be true in any circumstances in which he judges it, the content is judgementally a priori with respect to this way. The same applies to "I hereby judge that water is H<sub>2</sub>O". The judgementally a priori will also include such traditionally acknowledged examples of the a priori as contents reached by mathematical proof.

A second notion of the a priori I call the "contentually a priori":

*p* is *contentually a priori* with respect to a way *W* if *W* is an a priori way of coming to know *p*, and *W* is also a way that ensures the following: the content *p* of the judgement it yields is true in the actual world, whichever world is labelled as the actual world, and is true regardless of whether that way *W* is used, and of whether the conditions of its use are met, in the world that is labelled as the actual world.



Here the phrase “whichever world is labelled as the actual world” does not mean “I don’t care what the actual world is like”. “ $p$  is true in the actual world, whichever is labelled as the actual world” means: for any possible world, if it were actual,  $p$  would be true when evaluated with respect to it.

To say that something comes to be known in a way that ensures that it is true in the actual world, whichever is the actual world, is not to say this: that someone who comes to know something in this way thereby comes to know *that* it is true in the actual world, whichever is the actual world. The situation is quite parallel to the more straightforward case of the intuitive notion of an a priori way of coming to know some content. A person’s entitlement can be a priori without her exercising, or even possessing, the concept of the a priori. The same point applies to the contentually a priori. A person can come to know something that is contentually a priori with respect to the way in which she comes to accept it, without herself exercising or even possessing the concept of the contentually a priori. The fact, however, that there is a way of coming to accept a given content that does ensure that it is true in the actual world, whichever is the actual world, is something striking, and in need of philosophical explanation.

Being contentually a priori is a relation, between a content and a way. It will often be convenient to use an existential quantification of the relation. We say that something is contentually a priori tout court if there is some way with respect to which it is contentually a priori.

Those who are not made queasy by the whole idea of the a priori would count amongst the contentually a priori propositions the following: the known logical truths; known arithmetical truths; and propositions such as “If I exist, and this place here exists,

then I am here”, “No shade is both a shade of red and a shade of green”, and “If  $p$ , then Actually  $p$ ”. As some of these examples illustrate, and as the writings of Kripke and Kaplan made clear, something can be contentually a priori without being metaphysically necessary.

In modal semantics, Martin Davies and Lloyd Humberstone very helpfully introduced an operator “Fixedly”.<sup>7</sup> Its semantical clause states that “Fixedly  $p$ ” holds at a given world in a given model just in case it holds in that world in any model differing only in which world is labelled as the actual world. All the contentually a priori propositions I just mentioned hold Fixedly Actually in the sense of Davies and Humberstone. That is, if we preface them with the pair of operators “Fixedly” and “Actually”, in that order, the result is true. Enthusiasts for philosophically significant formal semantics will also be struck by the affinity between the contentually a priori and David Kaplan’s notion of validity in the logic of demonstratives, that is, the notion of truth with respect to every context in every structure.<sup>8</sup>

There is a sharp difference in extension between the judgementally a priori and the contentually a priori. Not everything that is judgementally a priori is contentually a priori. Simply considering the matter in the abstract, one should expect this. For a content to be judgementally a priori it is required only that it be true in each world in which it comes, by a certain route, to be judged. By contrast, to be contentually a priori a content must be true in the actual world, whichever is the actual world, regardless of whether it is judged, or how it comes to be judged.

The examples bear out the expectation of a difference in extension. Some self-verifying judgements are judgementally a priori, but they do not have the property of

being true in the actual world, whichever is the actual world. Worlds in which I am not thinking now, or not judging that water is  $H_2O$ , could have been actual.

Consider a way  $W$  which, when used, leads to judgement of a content that is judgementally a priori but not contentually a priori – that is, it leads to something which is *merely* judgementally a priori, as I will say. The explanation of why such a way  $W$  leads to a true judgement has to mention that fact that certain contents are actually accepted, or stand in other psychological relations, when the judgement is reached in that way. This applies to the explanation of the truth of such self-verifying judgements as “I am thinking” and “I (hereby) judge that water is  $H_2O$ ”. The explanation of why their contents are true must mention the fact that the judgements are actually made.

All these cases contrast with acceptance of the first-order content “ $13 \times 5 = 65$ ” on the basis of an arithmetical computation. The computational method is guaranteed to yield a result that is true in the actual world, whichever is the actual world, without reference to anything involving acceptance of the intermediate stages, or indeed anything psychological at all. That is why the first-order judgement of  $13 \times 5 = 65$  meets the stronger condition of being contentually a priori.

With the distinction between the contentually a priori and the judgementally a priori in hand, we can return to the Initial Thesis. At first blush, moral principles that are a priori do not seem to be merely judgementally a priori. They do not seem to be true only in worlds in which they come to be judged in a certain way. I want to propose, consider and defend the Initial Thesis in a sharpened and strengthened form, in which it concerns the contentually a priori. The Sharpened Thesis states:

Every moral principle that we know, or that we are entitled to accept, is either contentually a priori, or follows from contentually a priori moral principles that are known in conjunction with non-moral propositions that we also know.

This needs argument and defence against a variety of challenges. I will try to provide some of what is needed a few paragraphs hence. First I offer some observations intended to bring out the nature of this Sharpened Thesis.

The Sharpened Thesis corresponds closely to parallel theses in two other areas in which knowable truth seems to be truth that is, at a fundamental level, contentually a priori.

The first of these areas is that of metaphysical necessity, whose partial parallels with the moral case I will consider at several points. Each truth that contains a metaphysical modality, and that is also known to us, seems to be either itself contentually a priori, or it seems to follow from truths each of which is either a modal contentually a priori truth, or is an posteriori non-modal truth. It is necessary that Tully is Cicero. That modal truth is posteriori. But it is a consequence of an a priori modal truth - the necessity of identity - together with the a posteriori but also non-modal truth that Tully is Cicero. It has become a familiar claim about metaphysical necessity that every modal truth has its source in principles which are either necessary and a priori, or non-modal and a posteriori.<sup>9</sup> What is excluded is an irreducibly a posteriori modal truth. The a priori modal principles that are fundamental under this conception of metaphysical necessity are plausibly contentually a priori, and not merely judgementally a priori.

The second case paralleling the Sharpened Thesis is that of evidential and confirmation relations. Many instances of evidential and confirmation relations are a

posteriori. But it is arguable that each of them has an a priori component. A certain kind of rash confirms that an illness is meningitis. That is certainly a posteriori. But it rests on the a priori principle that a suitable range of instances gives nonconclusive support for a generalization, together with the truths about the presence of the rash in previous instances only in cases of meningitis, truths that are not themselves about the confirmation relation. What is excluded is an irreducibly a posteriori truths essentially about confirmation. Again, the notion of the a priori on which these are plausible claims is that of the contentually a priori.

Since evidential and confirmation relations are normative relations, this second case does more than merely provide a parallel example. It further suggests a general hypothesis: that there is a significant range of normative kinds, such that each truth of that kind has an a priori component. This thought will be resurfacing at several points later on.

The Sharpened Thesis has a more general epistemological feature. There has in discussions of justification and the a priori long been circulating an argument to the effect that in any domain in which justifications and reasons exist, some reason-giving relations must have an a priori status.<sup>10</sup> (Here we have the general hypothesis that all normative truths have an a priori component resurfacing already.) It is hard to see how justification and the making of judgements for good reasons could ever get started if all reason-giving relations were a posteriori. My own view is that this traditional argument is sound, when it is properly framed. There are all sorts of ways of mishandling the idea, some of which have to do with certainty. One such way of mishandling the idea is the view that if anything is probable, something must be certain.<sup>11</sup> But the idea that justification or

entitlement could not get started unless some principles or relations a priori can be developed without any commitment to the existence of such certainties.

If the reasoning of the traditional argument is sound, it applies as much in the domain of moral thought as it does in the area of empirical thought. Our Sharpened Thesis that all moral principles we are entitled to accept have a contentually a priori component dovetails with the traditional argument about justification. The Sharpened Thesis alludes to what must exist within the moral domain if the traditional argument is sound.

The Sharpened Thesis also has metaphysical ramifications, but I will first attempt to understand and explain its epistemic aspects.

## 2. *The Claim Defended*

A first objection to the claim of a priori status for basic moral principles may be that a thinker's impression, perhaps after some reflection, that a moral principle is correct is something that plays both a causal and a rational role in the thinker's acceptance of the moral principle. Why then is this impression not a conscious state whose role implies that basic moral principles are not a priori after all?

There must be something wrong with this objection, because such conscious states, playing a causal and a rational role, are present in clear cases of a priori status. A thinker may reflect rationally, and after her reflection, be left with the impression that a principle is a logical law. The thinker's impression will be both causally and rationally operative in her acceptance of the principle as a law. It is rational, in the absence of

reasons for doubt, to accept the outcome of such processes of reflective thinking. This can be an a priori way of coming to know the law.

What more specifically is wrong with the objection is that in the examples in question, the impression is not a justification. The impression of correctness is itself a rational response to conditions that give grounds for thinking that (say) gratuitous infliction of pain is *prima facie* wrong, or give reasons for thinking that the logical law is valid. In the former case, the fact that pain is subjectively awful provides such grounds; in the logical case, the justifying condition for a reflective thinker must include the fact that the law is true under all relevant assignments, or can be derived from such laws. The thinker has an impression of correctness only because he appreciates these justifications. Since the impressions of correctness in these examples are not themselves justifications, they cannot be used to support the claim that the thinker's operative justification in the moral or the logical cases is the character of one of his mental states.

This point is entirely consistent with the impression playing a causal role in the rational process leading up to the thinker's acceptance of the content. Of course the thinker would not have made the judgement in question if he had not had the impression that the content is correct. But that does not make the impression into a justification.

We can further emphasize the distance between impressions and justifications by considering their relations to correctness. For anything that is a justification for accepting a given content, there must be an account of why that justification entitles the thinker to judge that the content is true – an account of the relation between justification and truth, in short. An explanation of how a judgement comes to be made that includes reference to an impression of correctness is not by itself an explanation of why that method of

reaching the judgement is a correct method. For that, we need an account that mentions that to which the impression is a rational response, when it is a rational response.

This treatment still sharply separates the a priori cases from those of perceptual knowledge. Suppose you come to have the perceptual knowledge “That flower is yellow”. Your impression that this is a correct content is one to which you are entitled by the character of your perceptual experience; so the judgement is squarely a posteriori, indeed the paradigm case thereof. The explanation of why this is a correct way of reaching a judgement “That flower is yellow” would certainly have to mention the perceptual experience, as a source of non-inferential information about the world. Your impression that the content “That flower is yellow” is correct in these circumstances is parasitic on the justifying or entitling role of the mental state of perceptual experience, with its relation to correctness.

Some theories treat the impression of the correctness of a moral principle as something which is not the appreciation of a reason which is explicable independently of the thinker’s reactions on thinking about the principle, or its instances. There is a large subclass of such theories that treat moral properties as mind-dependent. Many different varieties of theory involve such mind-dependence. It is present in Christine Korsgaard’s idea that the source of normativity is an agent’s endorsement of “a certain way she looks at herself, a description under which she finds her life worth living and her actions worth undertaking”.<sup>12</sup> It is present in judgement-dependent theories, in various forms of subjectivism, and in a range of dispositional theories, where the dispositions in question concern mental properties.<sup>13</sup> Mind-dependence also seems to me to be present in Simon



Blackburn's treatment of moral thought, even though he himself explicitly denies that his view involves mind-dependence.<sup>14</sup>

I now want to raise the following question: can theories which treat the correctness of moral proposition as mind-dependent explain the apparent fact that basic moral principles are contentually a priori?

To separate the issues clearly, I first consider what the mind-dependent theorist can explain. Suppose we have some specific form of mind-dependent approach to moral norms. Suppose too that a thinker judges in ways acknowledged by that theory as suitably sensitive to the mind-dependent properties that he says are constitutive of moral norms. It will then hold according to that theory that the moral principles so reached will be true in any circumstances in which they are so reached. That is, under this mind-dependent theorist's conception, there is a way of reaching moral contents with respect to which they are judgementally valid.

It is hard to see how they could also be judgmentally a priori. Under a mind-dependent treatment, the entitlement to make the moral judgements is constitutively dependent upon the instantiation of the mind-dependent properties to which the moral judgements are sensitive, when the thinker is judging knowledgeably.

Can the mind-dependent theorist provide for non-introspective ways of coming to know moral propositions? It seems to me that he can allow for that. If statements of a certain kind are regarded as having mind-dependent truth-conditions, it does not follow that coming to know the truth of such a statement must (even in basic cases) involve checking on the thinker's own current mental states, or on anyone else's mental states. Statements about belief are certainly mind-dependent. Yet consider the Edgley-Evans

procedure for self-ascribing beliefs: “I get myself in a position to answer the question whether I believe that  $p$  by putting into operation whatever procedure I have for answering the question whether  $p$ ”.<sup>15</sup> The legitimacy of self-ascribing by this means shows that a procedure for self-ascription may nevertheless involve looking outwards towards the world, not inwards to one’s own mental states. Take the mental states to which, according to the mind-dependent theorist, a thinker must be sensitive if he is to be making moral judgements knowledgeably. If those mental states are not themselves about other mental states, the mind-dependent theorist can, it seems, consistently embrace the existence of non-introspective methods of coming to judge, knowledgeably, that certain moral propositions hold. Moral emotions, for example, are directed outwards to events, states of affairs and other people, and are not at all well-described as in general involving introspection of one’s own mental states. While there are many good questions about whether the mind-dependent theorist can properly characterize the mental states in terms of which he wants to explain moral thought, I think we can still grant the conditional that if, within the terms of his own theory, he has access to those mental states we normally express in our moral thought, he can legitimately claim that the ways of coming to know which he endorses are non-introspective.

Still, this is not to say that they are a priori. In particular, it does nothing to show that basic moral judgements are contentually a priori. The challenge to the mind-dependent theorist is to answer these questions: must not his theory imply that were our morality-generating sentiments to be different, what is actually wrong would no longer be so? If it does have that implication, he cannot explain the fact that basic moral principles are true in the actual world, whichever is the actual world, since it seems that they would

not be true if the actual world were one in which we had different morality-generating attitudes. Can the mind-dependent theorist show that basic moral principles are true in the actual world whichever is the actual world?<sup>16</sup>

This crucial question has multiple readings. On a theory according to which psychological states are the source of norms, in order to articulate this question more precisely, we need to introduce some double-indexing. We need to use the notion: proposition P, when evaluated from the standpoint of psychological states in world w1, holds with respect to world w2. We can abbreviate this to P(w1, w2).

It cannot be begging any questions against mind-dependent treatments to employ this doubly-indexed notion. The first parameter makes explicit the dependence that the mind-dependent theorist himself needs to use in articulating his own theory. The second parameter is just assigned whatever world is the one with respect to which the proposition P is being evaluated. So in the case of a mind-dependent theory of morality in particular, “P(w1, w2)” means that proposition P, when assessed according to the moral standards said to result from thinker’s psychological states in world w1, holds with respect to w2.

Here it helps to draw up some matrices, analogous to those introduced by Robert Stalnaker.<sup>17</sup> The mind-dependent theorist of moral thought is committed to holding that in each world, there is some set of basic attitudes in terms of which moral truth, or entitlement to moral judgement, is elucidated philosophically and on which the correctness of moral claims depends. We can use the notation “Atts<sub>i</sub>” for such postulated basic attitudes as are held by thinkers in world *i*. Each matrix corresponds to a given moral statement S (as we can neutrally put it). In each column of the matrix, we hold constant a parameter of the form Atts<sub>i</sub> for some fixed world *i*. The various entries in the

column specify the truth-value of the statement  $S$  at a given world, with respect to the constant parameter  $Atts_i$ . So suppose that under the basic attitudes of world  $i$ , an action of type  $A$  is prima facie good (in some given respect). We can suppose that this is a basic evaluation, and not subject to empirical variation, under the given standards. So in the column for  $Atts_i$ , every entry is a “T” for true. But under the different attitudes of worlds  $j$  and  $k$ , such an action-type is not prima facie good; again we suppose that this is a basic evaluation. So the matrix for the statement “Actions of type  $A$  are prima facie good” might be as follows:

	$Atts_i$	$Atts_j$	$Atts_k$
$i$	T	F	F
$j$	T	F	F
$k$	T	F	F

Our question was whether the mind-dependent theorist could explain the contentually a priori character of basic moral principles, that is, could explain the fact that they are true in the actual world, whichever is the actual world. There are clearly at least two possible readings of the phrase “ $P$  is true in the actual world, whichever is the actual world” when we have double indexing. One is Reading (A):

For any world  $w$ ,  $P(w, w)$ .

This is equivalent to having the entry True at each cell on the diagonal of the matrix that runs from top left to lower right. We can call this “the diagonal reading”. It means this: take any world, and the alleged basic morality-generating attitudes of that world, the proposition P will hold in that same world.

Reading (B) of “P is true in the actual world, whichever is the actual world” is:

For any world  $w$ ,  $P(@, w)$ .

This we can call “the vertical reading”, since it fixes on  $Atts_@$ , and for this reading to hold all the entries in the column  $Atts_@$  must be “True”. It means this: take our alleged morality-generating attitudes, and hold them fixed: then P holds in every world, when evaluated with respect to those attitudes so held constant.

A proper appreciation of the two-dimensional framework suggests a much more powerful formulation of the objection that the truth of moral principles cannot depend on psychological states of moral thinkers. On the quasi-realist’s theory, the acceptability of basic moral principles depends on some psychological attitudes. However this dependence is formulated, it must be possible to consider which propositions are correct when we vary the standpoint of evaluation, that is, when we vary the first parameter, as in (A). Take a specific moral principle identified by its content, say “Prima facie the infliction of avoidable pain is wrong”. Now consider the claim

For any world  $w$ , Prima face the infliction of avoidable pain is wrong ( $w, w$ ).

It seems to me that the quasi-realist, like other mind-dependent theorists, must say this is false. It is false at those entries in the diagonal for worlds in which we have different attitudes to the infliction of avoidable pain. The mind-dependent theorist has not, by his own lights, excluded those worlds. Unless the quasi-realist, or more generally any other

mind-dependent theorist, has some way of showing that our basic evaluations could not have been different, I do not see how the mind-dependent theorist can avoid a commitment to denying this most recently displayed claim. In short: the objection to mind-dependent views concerns the diagonal reading, and the objection is that the mind-dependent theorist has not explained, by his lights, why there cannot be an entry “False” somewhere on the diagonal. It cannot be an adequate answer to this objection to that there are no “False” entries on the vertical that corresponds to the actual world. In short: objections based on the diagonal reading cannot be answered by appealing to properties of the vertical reading.

It may be helpful in clarifying the distinction between the diagonal and the vertical readings to fix on some very simple concepts where we would also want to invoke double-indexing. It seems to be widely agreed that things would not stop being red if humans lost their colour vision, and saw only in shades of grey. It is entirely consistent with this point to hold that which colours things have is in some way constitutively dependent upon how humans actually perceive them (in circumstances in which they have not lost their colour vision). If one does hold that further claim, the right way to formulate the dependence is not in terms of counterfactuals like “If we were not to see things as red, they would not be red”, or any more sophisticated variants thereof. Such counterfactuals are evaluated from the standpoint of how humans perceive things in some central normal cases (that is, evaluated holding fixed the first parameter), and so cannot capture the intended dependence. But it is possible to formulate the proposed dependence all the same, either at a meta-level, or using some analogue of Davies and Humberstone’s “Fixedly Actually” operator. The best way of doing this would again

depend on the resolution of various auxiliary issues, but one simple formulation of the suggested dependence is this:

There is no physically individuated property Q such that it is Fixedly Actually the case that objects with Q are red.<sup>18</sup>

For an arbitrary physical property Q, our imagined mind-dependent theorist will be committed to rejecting the claim that

(DC) For any world w, Q-objects are red (w, w).

This is precisely parallel to the mind-dependent theorist of morality's commitment to rejecting the claim

(DM) For any world w, Prima face the infliction of avoidable pain is wrong (w, w).

This discussion should also make clear the strict limits of the earlier concession to mind-dependent theorists that allowed them to regard moral principles as judgementally valid. That concession can be granted only on the understanding that the first parameter, the attitudes that according to them are the source of moral truth, is held fixed.

We might pick out a moral principle not by its content, but by some definite description that relates it to those who accept it. The mind-dependent theorist does have access to some principle such as the following: in any world, a basic principle that is morally endorsed in that world will be one that holds in that world. This is identifying a moral principle by description, rather than by its content. Now a given matrix, of the sort I have introduced, corresponds to a statement identified by its content, by a that-clause. So the principle to which I have just agreed the mind-dependent theorist does have access

is not a principle that ensures that in a *given* matrix, all the entries along the diagonal are “True”. Rather, what it ensures is something concerning a set of many different matrices. It ensures that, for a given world  $w$ , if  $P$  is a statement endorsed by the basic morality-generating attitudes in  $w$ , then the entry in the matrix in the column labelled “Att $_w$ ” for the row for  $w$  will be “True”. This does not ensure what is required by the status of a given proposition as contentually a priori, viz.: possession of the entry “True” along the diagonal of a single given matrix. Rather, it gives only something weaker. It gives a diagonal of “True” entries in three dimensions, if you will, across a series of different two-dimensional matrices.

My position, in contrast to all mind-dependent views of moral principles, is that there is no sense in which moral principles fail to be contentually a priori. I hold this to be an epistemic and metaphysical truth. It is not itself a moral truth. The trouble for mind-dependent theorists is caused by variation with respect to the first parameter in  $P(w_1, w_2)$ .<sup>19</sup> If any form of mind-dependent theory of moral judgement is correct, that parameter must be articulable, at least at the level of philosophical reflection. My own view is that a proper appreciation of the contentually a priori status of moral principles ought to lead us to believe that any such parameter or argument-place is otiose. The moderate rationalist about morality who is also tempted to some form of subjectivism about colour will say that while basic moral principles are contentually a priori, so that - if the relativization is insisted upon - (DM) is true, in the case of colour, the characteristic consequence of one form of subjectivism holds, in that (DC) is false.

It is not in fact my view that our basic moral prima facie principles could intelligibly have been utterly different, in ways which have no connection with rationales



for the principles we in fact accept. That possibility was being entertained in the preceding part of the argument only for ad hominem purposes. My claim is that the mind-dependent theorists do not have the resources to rule out such variation, and so cannot explain why basic moral principles are contentually a priori.

In this discussion, I have focussed on the formulation used by Blackburn; but in fact the points I have been making seem to apply to any subjectivist or mind-dependent theory that tries to avoid the problems by using an “Actually” operator. Subjectivist and mind-dependent theorists are naturally tempted to appeal to our actual subjective states, or judgements, and to say that modal propositions about the moral should be evaluated always with reference to those actual states or judgements.<sup>20</sup> Contrary to the views of many writers in this area, I myself think that a proper deployment of the formal modal apparatus all things considered tells against mind-dependent approaches to morality, in a way in which it does not tell against mind-dependent approaches to statements about colours.<sup>21</sup>

If we step back to reflect on the argument I have given so far, it is apparent that it does not depend on features that are unique to morality. The argument I have offered so far can be developed in corresponding form to reject any mind-dependent treatment of any domain in which there are principles that are contentually a priori. The argument could be applied against mind-dependent treatments of metaphysical necessity, for instance (if further arguments against such treatments were thought to be needed). All the arguments against mind-dependent treatments in the moral case would carry through *pari passu* for the modal case.

### 3. *Explaining the A Priori Status of Morality: A Schema*

The claim that basic moral principles are contentually a priori does not by itself imply the view that they can be derived from the law of noncontradiction. The laws of modal logic, and other basic principles of metaphysical necessity, are also a priori. But they are not literally derivable from the law of noncontradiction alone. Otherwise modal logic would be a part of first-order logic, which it is not. Kant himself of course believed in a connection between what you can will without contradiction and the correctness of a principle. But his *Groundwork* also contains another idea, a more general idea which does not in its basic formulation mention noncontradiction. This more general idea contains the seeds of an explanation of the a priori status of moral principles. Kant writes:

the ground of obligation here must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but a priori simply in concepts of pure reason.<sup>22</sup>

This claim of Kant's is a consequence of the highly plausible principle that ways of coming to know a given proposition that are a priori ways have their source in the nature of one or more concepts in the given proposition. This principle is part of the moderate rationalism I mentioned at the start of this paper. If moral principles are a priori, and a priori ways of coming to know a proposition trace back to the nature of the concepts it contains, it follows that some ways of coming to know a moral principle have to do with the nature of moral concepts. Our task is to say how this is so.

I have already mentioned the modal case twice, and it will continue to help us to consider the partial parallel between modal and moral concepts. As I said, modal truth

seems to be fundamentally contentually a priori, like basic moral principles. Elsewhere, I argued that our understanding of modal truth is best explained by our having an implicit conception whose content is given by a set of principles that collectively determine which world-descriptions represent genuine possibilities.<sup>23</sup> Those principles I called the “Principles of Possibility”. The Principles of Possibility, whose details do not matter for present purposes, include principles entailing that genuine possibilities respect what is constitutive of the identity of the concepts, object, properties and relations they concern. What matters in considering a partial parallel with the moral case is the model of understanding, epistemology and metaphysics instantiated by this principle-based approach. Under the principle-based approach, to understand modal operators is to evaluate modal claims as true or false in accordance with these principles. The principles are at most tacitly known to an ordinary thinker when she evaluates modal claims. It takes philosophical thought to work out what those principles are.

The principle-based approach to modality has two features that that we equally need to provide for in the moral case.

First, it gives an account of how a way of coming to know, even one employed by a non-philosophical thinker, can be a way that ensures that what is known is true in the actual world, whichever is the actual world. In evaluating modal claims, the thinker draws on the content of tacit knowledge of the Principles of Possibility. These Principles state what it is, constitutively, for a description to represent a genuine possibility. The Principles are themselves true in the actual world, whichever is the actual world – they hold Fixedly Actually. Standard logical inferences will preserve Fixedly-Actual truth. Truths about what is constitutive of particular concepts, objects and properties are equally

plausibly truths that hold Fixedly Actually. If our thinker draws only on information which holds Fixedly Actually, by rules which preserve that property, when she evaluates modal truths, then the modal truth she comes to know thereby will hold Fixedly Actually. This is a way of coming to know a modal truth that ensures that what is known will hold in the actual world, whichever is the actual world.

This general method of evaluating modal claims is not infallible. Such general methods never are. A thinker may make mistakes about what is constitutive of the identity of a concept, object, property or relation; she may also make inferential mistakes. But when there are no such mistakes, the way in which a modal belief is reached can be one ensuring that its content is true in the actual world, whichever is the actual world. Since the existence of such ways of coming to know contentually a priori modal propositions relies on an account of understanding modal notions, and does not involve causal interaction with a modal realm, the principle-based account is a species of moderate rationalism for the modal case.

The other feature of the principle-based approach to modality that we also need to provide for in the moral case is its provision of a straightforward means for integrating the modal epistemology and modal metaphysics that steers between the extremes of mind-dependence on the one hand, and an epistemology that requires causal contact with a modal realm on the other. If the Principles of Possibility state what it is for something to be a genuine possibility, and those Principles are properly applied in reaching modal beliefs, we already have an explanation of how modal knowledge is possible. Such a middle course, avoiding both mind-dependence and interactionism is just what we need in the case of morality too.

The moral analogue of the principle-based treatment of modality is a treatment under which to possess moral concepts involves having an implicit conception whose content is operative when one assesses moral propositions. Full grasp of a given moral concept, if such a thing is ever possible, would involve possession of an implicit conception whose content formulates what it is, constitutively, for something to fall under that moral concept. The general idea of a principle-based treatment is in itself neutral on what the content of the implicit conceptions are. Many different first-order moral views could avail themselves of a principle-based treatment in attempting to address epistemological and metaphysical issues about the status of morality. So equally could many different philosophical views about what unifies the principles that form the content of the implicit conceptions. I will not be taking on the task of addressing particular first-order moral views here, nor the question of what unifies them. My aim is rather to consider what resources a principle-based treatment makes available to a variety of conceptions when they turn to address epistemological and metaphysical issues.

The implicit conceptions possessed by a moral thinker will be complex and structured. They will concern values, ideals, their relative importance, and something about their underlying sources. Even from a description as brief as that, there are two apparent differences from the modal case. One of the most important differences is the need for some kind of “prima facie” or “pro tanto” operator in the moral case, which, in my view, has no analogue in the modal case. It is plausible that one will need to employ, in any principle-based account of moral truth and moral epistemology, principles of the form “Prima facie, given that an action is F, it is good in such-and-such a respect”. The same applies to evaluations of states of affairs. The presence of a prima facie operator has

many repercussions, including some for the issue of determinacy. There is nothing in such structures to rule out the possibility that some type of action may be *prima facie* good in certain respects, *prima facie* bad in others, and there be nothing further in the principles to settle outright whether it is good or bad.

A second difference from the modal case concerns completeness. There is some plausibility that we can give a very general characterization of what is required for a description to represent a genuine possibility. It is arguable that if a description respects what is constitutive of concepts, objects and properties, it represents a genuine possibility. Though we are certainly ignorant, for many concepts and objects, of what it is that is constitutive of them, such ignorance concerns whether the conditions for certain possibilities are met, and is not about what it is for something to be possible. It is not apparent that anything analogous has to hold in the case of moral thought. Even our implicit conceptions may be incomplete, may need further articulation from reflection on examples and other principles.

A thinker may have an implicit conception with a correct content involving a given concept, but nevertheless make mistakes when asked to formulate general propositions involving that concept. This is a familiar phenomenon of implicit conceptions in other domains, evidenced by the frequent inability of thinkers to, say, define “chair” correctly, or to state explicitly the rules of grammar they are following. Indeed, even the simple example I have been using needs qualification. The infliction of avoidable pain is not *prima facie* wrong in the case in which the pain still exists, but is not experienced as hurting, as is the case for one who has taken morphine. The infliction of pain is *prima facie* wrong only when it is a form of suffering, and is wrong for the

same reason as it is wrong to cause, say, avoidable depression or severe anxiety in a person. Reflection on the ways in which we can correct our initial impressions of wrongness or rightness will make the principle-based theorist say that not all cases are like those which Prichard described as immediate apprehension, in which “insight into the nature of the subject directly leads us to recognize its possession of the predicate”.<sup>24</sup>

A thinker who judges that some type of action is wrong may be more or less articulate in his ability to say why it is so. At the least articulate level, the thinker may just make some clear intuitive judgement that it is wrong, without being at all confident in any particular explanation of why it is wrong. At one step up from this, the thinker may be able to give a ground: “because it would be a betrayal”, “because it hurts him and the hurting is avoidable”. At another step up, the thinker may be able to say why these are grounds. Higher levels of justification involve abductions from a priori examples and other apparently a priori principles. At this level of description, the methodology is the same as that found in other domains in which truth is fundamentally a priori. The possibilities of error are the same as in other a priori domains.

A principle-based approach can share each of the features that made the parallel to the modal case tempting. Even if a thinker’s implicit conception of some moral property is incomplete, the content of that conception can still be a correct partial statement of what it is, constitutively, for something to fall under that concept. They will, for instance, be a correct partial statement of what determines the semantic value of a concept like *is prima facie wrong*. When they are so, and when the information is properly drawn upon in the evaluation of contents containing that concept, the contents thus reached will be true. And as in the modal case, since the rule determining the

semantic value of a concept applies whichever world is the actual world, propositions thus reached will be reached in a way that guarantees that what is known in that way will also hold in the actual world, whichever is the actual world.

The fact that implicit conceptions are involved in the evaluation of moral propositions does not by itself suffice to account for the contentually a priori status of basic moral principles. There is no contradiction in the idea of an implicit conception having an a posteriori content. In fact, an implicit conception with the content that the word ‘chair’ in one’s own language applies to things having certain properties is an implicit conception with an empirical content. What matters for a priori status is rather that the given way of coming to know is guaranteed to be correct by the way in which the semantic values of the relevant concepts are fixed. Implicit conceptions whose contents either consist of principles that state what it is for something to be wrong, for instance, or consist of consequences thereof, meet this further condition. Without this further condition, we would not have an explanation of the contentually a priori status of basic moral principles. The same applies to the modal case.

This integration of the metaphysics of the moral – what it is to fall under certain normative concepts – with an epistemology also steers the same middle course as the principle-based account of modality. It involves neither a mind-dependent account of moral truth, nor a causal epistemology for the contentually a priori principles.<sup>25</sup>

The point that fundamental principles help to determine the semantic value of concepts like *is prima facie wrong* is important in separating any principle-based conception from mind-dependent treatments of moral thought. Mind-dependent theorists can fairly insist that on their views, a certain set of moral principles is correct, and can



equally insist that some principles are more fundamental than others. That does not imply that mind-dependent theorists can simply take over the apparatus of the principle-based view. The objection remains outstanding against the mind-dependent theories that they cannot explain the contentually a priori status of basic moral principles. To try to meet the objection by saying that the principles themselves determine the semantic value of moral concepts, regardless of what attitudes minds take to them, would be to abandon any claim of mind-dependence. A principle-based conception is a very different animal from any mind-dependent view.

#### 4. *The Subjectivist Fallacy*

The Subjectivist Fallacy is the fallacy of moving from a premise stating that certain mental states are sufficient, or are necessary, for a given content to be true, to the conclusion that the truth of the content consists, at least in part, in something subjective or mental. I say that this is a fallacy even in the case in which the premise stating that certain mental states are sufficient, or are necessary, holds true a priori. To say that it is a fallacy is not of course to say that the conclusion is not true: only that it cannot be supported just from these premises.

The Subjectivist Fallacy is a fallacy because it may be possible to explain why the mental states are necessary or sufficient for the truth of the target content by exhibiting this necessity or sufficiency as a consequence of a more fundamental account of what is involved in the truth of the target content, a more fundamental account that does not mention mental states at all. The fact that there is in a certain sense no gap between

certain mental conditions obtaining and the holding of the target content may have a non-subjectivist explanation.

Here is an example of the Subjectivist Fallacy, an example which would be recognized as such on all but the most extreme views of the nature of meaning and rule-following. The case involves a hypothetical position on the understanding of arithmetical relations. We can imagine a theorist who starts from this true premise:

Within the accessible numbers, it is sufficient for  $n+m$  to equal  $k$  that a thinker who reaches his judgement about what  $n+m$  equals in accordance with certain recursive procedures will judge that  $n+m=k$ .

From this truth, our imagined theorist now moves to the conclusion

Equations involving addition have partially mind-dependent truth-conditions concerning what a certain kind of thinker would judge.

Almost everyone will agree that this hypothetical theorist's mistake lies in not realizing the judgements of his hypothetical calculating subject are correct only because they respect the recursive equations for addition. The fact that there is (and is a priori) a necessary and sufficient condition, framed in terms of the judgements of a hypothetical thinker, for the holding of the addition relation on the accessible numbers is just a by-product of something more fundamental. This more fundamental condition is the non-psychological truth-condition for equations involving addition determined by the recursive characterization of the addition relation.

How does this bear on constructivism in ethics? Constructivists need not be mind-dependent theorists. Constructivists too can agree that the displayed transition about addition moves from a true premise to a false conclusion, provided their constructivism is

of a non-psychological variety. Their constructivism will be of a non-psychological variety only if the “can” in the phrase “can be constructed” which features in a statement of constructivism is not explained in psychological terms. It is also a necessary condition of the constructivism being non-psychological that the particular rules or recursions it mentions are not mentioned there by virtue of their meeting some mind-dependent condition. Some versions of constructivism meet these conditions. Hence constructivists need not be mind-dependent theorists.

The crucial step in the subjectivist fallacy as I have described it is acceptance of an incorrect criterion for the mind-dependence of a given property. So it is possible to make what seems to me the same mistake as is made in the subjectivist fallacy without actually being a subjectivist. Even a theorist who rejects subjectivism about a given domain may still be using a questionable account of mind-dependence of a given property. The theorist may even be relying on that account in his rejection of subjectivism. The writings of Crispin Wright and Mark Johnston contain examples of criteria for mind-dependence that seem to me open to question in this way. In his well-known discussion of the Euthyphro Contrast, Wright introduces the notion of a “provisional equation”, which is something having the form of a conditional whose consequent is itself a biconditional, i.e. the form  $A \supset (B \equiv C)$ . A provisional equation is something of the form “If CS, then (it would be the case that  $p$  if and only if S would judge that  $p$ )”.<sup>26</sup> A substantial provisional equation, says Wright, has an antecedent CS in which “a concrete conception is conveyed of what it actually does take” for the subject to be operating under conditions in which her opinion is true.<sup>27</sup> Wright endorses this conditional: “if a discourse sustains substantially formulated true provisional equations

which can be known a priori to be true, then that makes the beginnings of a case for regarding the discourse as dealing in states of affairs whose details are conceptually dependent upon our best opinions”.<sup>28</sup> Similarly Mark Johnston, in addressing the question “How then are we to demarcate the response-dependent concepts?” offers the answer that if a concept *C* is one interdependent with, or dependent upon the responses of subjects, “then something of the following form will hold *a priori*

*x* is *C* iff In *K*, *Ss* are disposed to produce *x*-directed response *R* (or *x* is such as to produce *R* in *Ss* under conditions *K*).”<sup>29</sup>

This biconditional will be fulfilled in our first, arithmetical, example, when we take the concept *C* to be the property of (say) being the sum of 7 and 5, the condition *K* to be the condition of exercising properly-functioning memory and perceptual systems, and the response to be that of making a certain judgement expressing the outcome of the subject’s computation in accordance with certain rules. A corresponding point could be made about Wright’s criterion. In both Johnston’s and Wright’s proposals, the test proposed for mind-dependence is too easily met by propositions whose truth is not mind-dependent. Nothing can be validly concluded from the existence of such a priori conditionals or biconditionals in a given domain about the mind-dependence of that domain.

There is nothing inimical, in these illustrations and arguments, to the idea that some contents do have mind-dependent truth conditions. They do when their truth-conditions concern a property whose nature – what it is, constitutively, to have to property – is to be explained in terms of properties of the mind. The burden of the preceding remarks is that this constitutive condition cannot be reduced to something involving a priori equivalence with conditions concerning certain mental states.

The Subjectivist Fallacy is an instance of a more general fallacy concerning the nature of properties. The more general fallacy is that of moving from the a priori truth of a biconditional of the form

$$F(x) \text{ iff } A(x)$$

to the conclusion that being *A* is what *makes* something *F*. I call this “the Biconditional Fallacy”. Just as in the subjectivist case, it is a fallacy because the correct account of what makes something *F* may have a consequence that it is a priori that something is *F* iff it is *A*; but the constitutive account may not mention properties or notions of the sort mentioned in the condition  $A(x)$ . One of the tasks facing those who want to develop Discourse Ethics, for example, is to show that it can be done without committing this fallacy. Habermas formulates the central claim of discourse ethics as follows: “Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse”.<sup>30</sup> Let us suppose that the theorist of discourse ethics makes a good case that this condition is a priori true. Nothing would follow about what makes something a valid norm. The approval, in the appropriate practical discourse, of those affected might be a consequence of more fundamental principles about norms that have this approval as a consequence. Consistently with the principle Habermas formulated being a priori, practical discourse might not be mentioned in an account of what is fundamentally constitutive of the notion of a valid norm. It is that further claim about fundamental constitution that Discourse Ethics would have to establish if it is to speak to the nature of morality.

I conclude with two more general reflections on rationalist positions. The first concerns the range of theoretical options available to us. When one reads the literature on

judgement-dependent and other mind-dependent approaches to ethics and other subject-matters, the impression is often conveyed that, when we do have an a priori biconditional linking some property with thinkers' mental states, there are only two options. Either we read the psychological material of the right-hand side as providing what is constitutive of the left-hand sides' holding: or else we must accept some form of "detectivism", with the overtones of "detection" involving a causal epistemology for the states of affairs detected.<sup>31</sup> To think that these are the only two possibilities is to overlook broadly rationalist approaches that are neither mind-dependent nor committed to the possibility of causal interaction. It is as if the only two possibilities in the philosophy of arithmetic, or the philosophy of modality, were either subjectivism, or else a commitment to causal interaction. I suggest that a good rationalist treatment of mathematics, modality and morality involves neither of those two positions, but genuinely offers a third way.

The other general reflection concerns the relations between moral rationalism and the nature of moral inquiry. Moral rationalism does not imply that moral inquiry cannot be advanced by partly empirical studies. The point is an instance of a much more general principle about our ways of acquiring knowledge of a priori truths. The general point applies even in the mathematical and logical cases. It was the empirical, physical theory of Newton's *Principia* that motivated the development of the calculus; and the calculus has as good a claim to a priori status as any other part of number theory. In the logical case, one can by reflection on empirical truths come to appreciate that the counterfactual conditional does not support transitive inferences. One could have established this non-transitivity a priori from a good semantics for the counterfactual conditional. It does not follow that we would have discovered, or indeed been confident about, that non-

transitivity without reflection on empirical examples. In the moral case, reflections on historical states of affairs and current actual situations may lead us to formulate concepts, principles and distinctions we might otherwise never have thought of. The same point applies to the moral emotions. Our moral indignation at a state of affairs, or our sudden guilt in reflecting on one of our own actions, may lead us to moral reflections we might not otherwise have attained. My own view is that the wrongness of some state of affairs never consists in its tendency to cause such moral emotions. Rather, the wrongness of the state of affairs in question is always explicable in terms of some emotion-independent condition. But once again, it does not follow that we would have reached these moral judgements without the help of the moral emotions. Moral inquiry would be a poorer thing if we had no access to empirical examples, empirical theories, and to the moral emotions. Moral rationalism does not license armchair philosophy.

CHRISTOPHER PEACOCKE

Columbia University

---

<sup>1</sup> I thank Richard Boyd, Alex Byrne, Paul Boghossian, Kit Fine, Christine Korsgaard, Wolfgang Kühne, Derek Parfit, James Pryor, Peter Railton, Stephen Schiffer, Tim Scanlon, Pekka Väyrynen, Ralph Wedgwood, David Wiggins, Aaron Zimmerman, and several anonymous referees for valuable discussion and comments. This material was presented at Cornell, London, NYU, and Oxford; at the fourth meeting of the

---

Gesellschaft für Analytische Philosophie in Bielefeld; and at Harvard as the first of my Whitehead Lectures in 2001.

<sup>2</sup> Leibniz, *New Essays on Human Understanding*, esp. Book I, Chapter ii, pp.91-4 in the edition of P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981); Kant, *Groundwork of the Metaphysics of Morals*, at 4:408 (pp.62-3) in *Practical Philosophy* Cambridge Edition of the Works of Immanuel Kant, tr. and ed. M. Gregor (Cambridge: Cambridge University Press, 1996). Rawls argues that only the procedure of the Categorical-Imperative is a priori for Kant, and that moral principles are reached using it only in the presence of empirical information. See J. Rawls, *Lectures on the History of Moral Philosophy*, ed. B. Herman (Cambridge, Mass.: Harvard University Press, 2000), pp. 247-52. If Rawls's reading is correct, it remains that one of the sources of true moral principles is fundamentally a priori.

<sup>3</sup> *An Essay Concerning Human Understanding*, Book IV, Chapter 4, section 7.

<sup>4</sup> See my "Explaining the A Priori", in *New Essays on the A Priori*, ed. P. Boghossian and C. Peacocke (Oxford: Oxford University Press, 2000).

<sup>5</sup> *Groundwork of the Metaphysics of Morals* at 4:389 (p.45 in the Cambridge edition, op.cit.).

<sup>6</sup> Second Meditation, in *Philosophical Writings of Descartes: Volume II*, tr. J. Cottingham, R. Stoothof and D. Murdoch (Cambridge: Cambridge University Press, 1984), p.17.

<sup>7</sup> "Two Notions of Necessity", *Philosophical Studies* XXXVIII, 1 (July 1980) 1-30.

<sup>8</sup> "Demonstratives" in *Themes from Kaplan* (New York: Oxford University Press, 1989) ed. J. Almog, J. Perry and H. Wettstein, second definition on p.547.



---

<sup>9</sup> See my *Being Known* (Oxford: Oxford University Press, 1999), Chapter 4, “Necessity”.

<sup>10</sup> For an overview and one kind of defence, see L. BonJour, *In Defense of Pure Reason* (Cambridge: Cambridge University Press, 1988), p.5ff. For a somewhat different argument for the view that all instances of the entitlement relation are fundamentally a priori, see the ‘Third Principle of Rationalism’ discussed in my paper “Three Principles of Rationalism”, *European Journal of Philosophy* X, 3 (December 2002), 375-397. If it is true that all entitlement is fundamentally a priori, that is not in itself alone an argument for moral rationalism. There can be entitlements in domains for which forms of subjectivism or mind-dependence hold. In those cases, the entitlements are dependent upon certain mental conditions holding. My position is that there is no such dependence in our entitlement to hold moral propositions.

<sup>11</sup> A theme in C.I. Lewis’s writings. See his *Mind and the World Order: Outline of a Theory of Knowledge* (New York: Dover, 1956), for instance pp.311-12.

<sup>12</sup> *The Sources of Normativity*, ed. O. O’Neill (Cambridge: Cambridge University Press, 1996), p. 249.

<sup>13</sup> For discussion of such approaches, see C. Wright, *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press, 1992), Chs. 3, Appendix to Ch. 3, and Ch. 5; D. Wiggins, “A Sensible Subjectivism?”, in his *Needs, Values, Truth* (Oxford: Blackwell, 1987); and the papers by D. Lewis, M. Johnston and M. Smith in the Symposium “Dispositional Theories of Value”, in *Proceedings of the Aristotelian Society, Supplementary Volume LXIII* (1989) 89-174.

<sup>14</sup> For an overview of Blackburn’s more recent position, see his *Ruling Passions: A Theory of Practical Reason* (Oxford: Oxford University Press, 1998). For a statement of

---

his view that his quasi-realism does not commit him to mind-dependence, see for instance his answer to question 2, p.311-12 of *Ruling Passions*. Pages references appended to quotes from Blackburn are to this book. It is hard to see how it can be denied that, under his approach, the conditions under which someone is correct in asserting a moral proposition have something to do with expressed mental states.

<sup>15</sup> R. Edgley, *Reason in Theory and Practice* (London: Hutcheson, 1969); G. Evans, *The Varieties of Reference* (Oxford: Oxford University Press, 1982), Section 7.4, p.225.

<sup>16</sup> Since I already argued that being contentually a priori implies being judgementally a priori, and suggested that under the mind-dependent view, basic moral principles are not judgementally a priori, I already have an argument that the mind-dependent theorist cannot explain why basic moral principles are contentually a priori. But the arguments of particular theorists, such as those of Blackburn considered below, to the effect that there is no problem here, mean that we have to consider the case of the contentually a priori separately.

<sup>17</sup> *Context and Content* (Oxford: Oxford University Press, 1999), Introduction and Chapter 4, “Assertion”.

<sup>18</sup> Though of course this theorist will accept that

For any world  $w$ , Q-objects are red ( $@$ ,  $w$ )

in the case in which Q is in the actual world the underlying physical property of objects which are red

<sup>19</sup> The problem is specific to this feature. There is of course no general incompatibility between a domain of truths – such as the truths about which material objects have which colours - being mind-dependent and the existence of contentually a priori truths about the

---

properties attributed in that domain. Principles of colour incompatibility are contentually a priori. This is not merely consistent with truths about colour being significantly mind-dependent. In my judgement, the explanation of their status as contentually a priori has to draw upon the special relation of colour concepts to colour experience.

<sup>20</sup> For such use of an “Actually” operator in defending a subjectivist theory, see D. Wiggins, “A Sensible Subjectivism”, in *Needs, Values and Truth: Essays in the Philosophy of Value* (Blackwell: Oxford, 1987), p.206. Wright in *Truth and Objectivity* describes the use of an “Actually” operator as “an attractive strategy” (114), but, rightly in my judgement, goes on to caution that “no proposition whose necessity is owing entirely to actualisations can be known *a priori*” (116).

<sup>21</sup> Wiggins cites Davies and Humberstone pp.22-5 in support of his use of an “Actually” operator to meet the objections to subjectivism. David Lewis gives a very clear acknowledgment of the problem for a subjectivist theory: “The trick of rigidifying seems more to hinder the expression of our worry than to make it go away. It can still be expressed...”: “Dispositional Theories of Value”, op. cit., at p.132ff. My points in these paragraphs can also be seen as an endorsement and elaboration of some of G.E.Moore’s points in his paper “The Conception of Intrinsic Value”, in his *Philosophical Papers* (London: Routledge and Kegan Paul, 1922), 253-75.

<sup>22</sup> *Groundwork* 4:389, p.45 in the Cambridge edition, op. cit..

<sup>23</sup> *Being Known*, Chapter 4.

<sup>24</sup> H.A. Prichard, in his essay “Does Moral Philosophy Rest on a Mistake?”, repr. in *Moral obligation and Duty and interest: Essays and Lectures* ed. J.O. Urmson (Oxford: Oxford University Press, 1968), at p.8.

---

<sup>25</sup> It does not follow that moral properties may not be involved in other causal explanations, not having to do with knowledge of a priori principles.

<sup>26</sup> *Truth and Objectivity*, p.119. I have altered only Wright's notation for propositions.

<sup>27</sup> *Truth and Objectivity* p.112.

<sup>28</sup> *Truth and Objectivity* pp.119-20.

<sup>29</sup> "Dispositional Theories of Value", *Proceedings of the Aristotelian Society Supplementary Volume LXIII* (1989) 139-74, p.145.

<sup>30</sup> *Moral Consciousness and Communicative Action*, tr. C. Lenhardt and S. Weber Nicholsen (Cambridge, Mass.: MIT Press, 1990), p.93.

<sup>31</sup> See *Truth and Objectivity*, on the Euthyphro contrast, Appendix to Chapter 3.