Abstract for

Reflections on Moral Disagreement, Relativism, and Skepticism About Rules

By Denis Robinson, University of Auckland

Part I of this paper discusses some uses of arguments from radical moral disagreement — in particular, as directed against absolutist cognitivism — and surveys some semantic issues thus made salient. It may be argued that parties to such a disagreement cannot be using the relevant moral claims with exactly the same absolutist cognitive content. That challenges the absolutist element of absolutist cognitivism, which, combined with the intractable nature of radical moral disagreement, in turn challenges the viability of a purely cognitivist account of moral judgments. Such a conclusion could be staved off if it could be held that a sufficient condition for commonality of cognitive content in moral judgments could consist, despite the presence of radical moral disagreement, in the parties' acceptance of a common set of fundamental moral principles. Part 1 begins, and Part 2 further develops, a destructive critique of that idea, leading thereby to a skeptical appraisal of the important role sometimes assigned, in meta-ethical theorizing, to moral rules. *Inter alia* the paper is intended to suggest the possibility of overlap between relativist and particularist agendas.

Reflections on Moral Disagreement, Relativism, and Skepticism About Rules

This paper consists of two main parts. Part I discusses the use of arguments from radical moral disagreement — in particular as directed against absolutist cognitivism[1] — and surveys a number of issues made salient by such arguments. Part 2 discuss the role of moral rules in underwriting moral judgment, and in particular, in determining the meanings of terms of moral evaluation, attacking a doctrine I call "RFMM" – "Rules Fix Moral Meanings". I hope each part has some independent interest, but the important link between them is that RFMM is sometimes suggested as a rejoinder to the kind of critique of absolutist cognitivism discussed in Part I.

Arguments from radical moral disagreement can be seen to challenge the objectivity of moral claims, or the epistemic accessibility of moral truths. But

[1] I comment briefly on my use of some key meta-ethical terms in a *Terminological Appendix* at the end of this article.

they are at their most powerful when given, broadly speaking, a semantic twist. So taken, they challenge what we might call the presumption of stable cognitivist univocity: the presumption that there is a single cognitive content *P* such that one party to the disagreement unequivocally asserts *P,* and the other unequivocally denies *P.* Such a presumption relies on the view that terms of moral evaluation used by disputing parties speaking a common natural language use terms of moral evaluation reliably to express the same cognitive content. Since this is hard to reconcile with the existence of radical moral disagreements, given simple cognitivist assumptions such disagreements demand some kind of relativist, rather than an absolutist, interpretation of those terms.[2]

It is commonly suggested that such threats to absolutist cognitivism can be rebutted by appeal to alleged pervasive agreement about Fundamental Moral Principles. This would only work if agreement about moral principles could somehow provide a guarantee, contrary to appearances, that the relevant terms are being used by the disputing parties with an exactly equivalent, absolutist sense. This is the doctrine I refer to above as "RFMM".[3]

The longest portion of Part 2 comprises an argument which surveys a range of ways in which moral rules might figure in determining agents' particular moral judgments. As these possibilities become more and more lifelike, I suggest, it becomes increasingly evident that some key components of moral judgment and understanding cannot plausibly be represented in terms of allegiance or conformity *merely* to a set of moral rules. I conclude by considering some possible implications of my discussion, and some illustrative analogies which may help to convey the attitude to moral principles which I advocate. This attitude has obvious affinities here with

[2] For the purposes of this paper I consider only disputes between speakers of what is *prima facie* a common natural language, mostly avoiding issues about translation of moral vocabulary between distinct natural languages. This somewhat artificial restriction, if anything, favours my opponents. Naturally, I think similar points can be made about moral judgments expressed in a range of languages, and issues about translating them. Tersman(2006) makes claims about constraints on translation central to his argument; see especially Chapter 6.

[3] My interest in RFMM was sparked by some remarks, reminiscent of it, appearing (in the context of a more complex and sophisticated position) in the work of Frank Jackson. See Jackson(1998); also Robinson(2009).

particularism. Without discussing the point at length, I certainly intend my discussion to suggest that particularism and relativism might have more in common than is often acknowledged.

I need to say clearly here that I do not principally talk about moral *choice,* but about moral *evaluation,* even though I talk as if a principal target of moral evaluation is *actions* — past or future, our own or others'. My supposition in effect is that the *sine qua non* of moral choice is the *general* ability to morally evaluate actions. We may favour various mental procedures in choosing an action which will stand up to our own retrospective moral evaluation. But for all that I say, these may be merely techniques for hitting a certain target, where merit attaches to hitting that target, not to the techniques. This may be anathema to anyone wishing to base moral evaluation of actions primarily on the processes by which actions are chosen or generated. My discussion may not absolutely exclude such a view, but certainly does not presuppose it.

Part 1: Radical Moral Disagreement, and Moral Rules: Setting Context, Surveying Landscape

1.1 On stabilizing meanings

It is reasonable to ask questions like the following about terms such as "right" and "wrong", used in English to express moral evaluations. Are there usefully expressible criteria which constitutively govern correct application of those predicates to particular cases, or which are constitutive of the concepts those predicates express? How are those predicates given their meaning, or those concepts given their content, and how are those meanings, or concepts, taught, transmitted, stabilized, and kept constant, so as to enable successful communication and mutual comprehension within a community which uses such terms, or employs such concepts — crucially, in applying them to particular events and actions?

We might ask, quite generally, what if anything *anchors, fixes or sustains constancy* of meanings across populations and times. When millions of people, over centuries, speak a common language, what ensures that words of that language are used with a common meaning, across that diverse and scattered population, and from one generation to the next? Notwithstanding the legacy of lexicography, the most general answer is surely "nothing in particular, beyond the mere practice of using and transmitting the language". Languages surely spread and develop, to a large extent, through unconscious and, broadly speaking, evolutionary processes. RFMM, if correct, would constitute an exception to this quietist generalization, since it singles out a quite small and fixed subset of contexts of moral terms as having a unique and controlling status in fixing the meanings of those terms. Would this be a *sui generis* exception to our generalization, or are there other exceptions which might to some extent provide models for this one?

One class of such exceptions might involve procedures consciously dedicated to ensuring stability and commonality of meanings of particular words or word-families, where meaning-variation would be particularly inconvenient. Cases might include the sciences (think of standard physical units, and of biological, chemical, and medical taxonomies), the law, and organized competitive games (as in "the Laws of Rugby"). These cases provide only a limited model for understanding RFMM. One reason for this is the fact that there is no good moral analogue (in current secular society) for the recognizable and empirically testable expertise of the scientific community, nor for the canonical rule-fixing status of the International Rugby Board. But though limited, the model has some relevance, since RFMM does assign a special place — and potentially, a recognized one — to fundamental moral principles as stabilizing moral meanings.

When thinking about how words are anchored and stabilized in their meanings, we must also acknowledge that for many words in our ordinary descriptive vocabulary this is done by straightforwardly referring to natural kinds, so that nature's own joints in effect constantly calibrate our usage. Just as water flows naturally into gullies, verbal distinctions, it might be said, gravitate naturally towards distinctions in nature. Indeed the development and systematization of canonical meaning-fixing stipulations and procedures in the sciences involve building on precisely that basis, as science reflexively improves our understanding of it.

1.2 Radical moral disagreements and the issue of semantic "anchoring" for moral terms

Can we exploit a version of this natural-kind-oriented, self-calibrating model in our portrayal of RFMM? I think not. Note first that this model certainly does not apply to all ordinary language terms and distinctions. (Consider for instance distinctions drawn in rules governing behavior around the scrum in rugby.) There is little plausibility in the view that moral terms and concepts are anchored to instances of natural kinds which, through their stable patterns of effects, may be identified as the kinds playing such and such causal roles.[4] One of several reasons to doubt such a view is the existence of an important subset of moral disagreements, those where dispute over uncontroversially descriptive facts is not in play.

A premise of what follows is that such fundamental moral disputes may be important, deeply intractable, and in principle irresolvable, in the following quasi-technical sense: *neither side can cite grounds for declaring the other side wrong, which could not reasonably be called question-begging in the context.* These are what — in line, pretty much, with contemporary usage, I am calling "radical moral disagreements".[5] These are not, it seems to me, situations where one might appropriately attempt to refute either party by seeking to backtrack along reference-transmitting causal chains, searching for a less than obvious "real essence", to determine the true content of the relevant terms of moral evaluation.[6] Moreover, an attempt to account in this kind of way for such stability of meaning as moral terms have, would make a poor fit with the narrow restriction of *canonical occurrences* of such terms to a relatively small number of Fundamental Moral Principles.

To further clarify the notion of radical moral disagreement, let's consider a sort of case which, though it blocks attempts at resolution, is *not* a case of this kind. Consider a dispute over the right way to dispose of a person's estate, involving different opinions of the probable contents of a lost will. Here the disagreement is not a product of ignorance about any hidden or inaccessible *moral* facts (and *a fortiori* it is not a case of ignorance about some moral "real essence" of which observable features of the situation might or might not be causal products). Unlike cases of radical moral disagreement, the case puts no pressure on any view about the nature of moral facts or of the semantics of moral terms. The contents of the lost document have a bearing on what it is right to do, nevertheless insofar as it turns on those contents, the dispute is

[4] See Robinson(2004) and (2009) for discussion of some pertinent issues about words, concepts, and semantics.

[5] Tersman(2006) provides a penetrating exploration of such disagreements and their significance.

[6] These issues surface again below, in part 1.3. A more nuanced discussion than I can supply here, would need to consider some well-known and important debates pitting "Cornell Realists" against Gilbert Harman. My views are closer to Harman's.

not properly speaking a moral one at all. Moral education is no more directed at learning how to find — or guess the probable contents of — lost documents, than it is directed at learning which observable syndromes probabilify an aetiology involving hidden real essences. *Radical* moral disagreements start when agreement has been reached about *all* such potentially relevant non-moral facts, but moral disagreement remains. (The disputants might abandon appeal to claims about the probable contents of the will and begin debating what it is right to do on the agreed assumption that those contents will never be known, but that would make the case a different one.)

Radical moral disagreements are doubly puzzling, on cognitivist and absolutist assumptions. On the one hand (we could call this "the primary puzzle") they present what, at least in another context, might seem like paradigmatic evidence of terms being used with different senses – albeit, in a relativist or indexical way, with contextually sensitive but importantly similar senses.[7] Thus these disagreements push cognitivists in the direction of relativism. On the other hand, these disagreements do not present the appearance we would expect if we viewed them in the light of a *simple* relativism (we could call this "the secondary puzzle"). For such a view would leave unexplained the appearance of a genuine disagreement: it would predict that the proponents would be "talking past each other", and not genuinely in disagreement, robbing the entire dispute of any point once the misunderstanding — mistaking relative terms for absolutes — is recognized. As I see it, this problem is to be rectified only by moving to a hybrid view which combines a kind of constrained cognitivist relativism with non-cognitivist elements.[8] (This two-step is what I earlier called "the semantic twist" on arguments from moral disagreement.)

[7] Here we basically need a distinction reminiscent of Kaplan's, between different tokens of indexical expressions which in one sense have a common type-meaning — the same "character" — yet in another sense have different, contextually-determined token-meanings — different "contents".

[8] The implications of radical disagreements are explored in Robinson(2004) and (2009). I suggest in Robinson(2004) that there are radical disagreements about personal identity which are, as it were, radical moral disagreements by proxy. There I call the kind of hybrid view I believe is needed, "quasi-relativism", and radical disagreements "quasi-disagreements". I believe my standpoint here is at least broadly consonant with Tersman's(2006).

I do not have an account to offer of the details of such a view. This much seems clear: it will not be a view of the "besires" kind which entails a kind of lock-step between belief, and some desire-like or desire-related, motivating attitude like "valuing". If there is anything remotely right in the philosophers' idea of the so-called "psychopath" or "amoralist" who can mirror the beliefs and moral claims of the person who genuinely judges something to be wrong, whilst having absolutely none of the accompanying motivation required by the so-called "internalism" constraint, it can't be that there is an attitude which of necessity involves a correct proportionality between degree of belief and degree of motivation. Indeed if (much less controversially) it is merely the case that two morally concerned agents can be equally convinced of something's being a clear-cut case of wrongdoing, but differ in the extent or force of the motivation which that judgment involves or generates in them — a situation entirely common in our world of morally concerned people who are less than saints — then we must surely think of ordinary moral judgments and their normal motivational accompaniments as packages of two logically orthogonal components, the force or degree of which can vary independently.

If we can sensibly talk of these two "components" of moral judgments, then it is at least possible to ask which of them is better thought of as essential for type-individuating particular moral judgments. The "latitude idea" expounded in Tersman(2006) implies, rightly I think, that reflection on radical moral disagreements shows that at least for some purposes it is the motivational element which is more fundamental. Cases of radical moral disagreement, on this view, are not mere cases of talking past one another, despite the expressed beliefs not being logical contraries, since there is genuine contrariety between the *practical* implications of the opposing claims due to the associated *motivational* differences they express. (None of this entails that the latitude allowed on the belief dimension of moral judgments is unconstrained: that is why a pure non-cognitivist view cannot serve here.)

Full development of a view of this genre would require carrying out the difficult task of saying how the *content* (not just the force) of the particular motivation-related part of the moral-judgment package relates to the proposition, event, state of affairs, or what have you, of which the appropriate property of rightness, wrongness, or goodness is predicated in the cognitive component of the package. *Prima facie,* the latter must change as we move from taking a prospective view, regarding a possible future action, where we can coherently want to perform it (or want someone else to), to a retrospective view of an already performed action. In the latter case we cannot coherently desire that anyone either perform or un-perform *that* particular action. In this change from prospective to retrospective evaluation, the object of our attitude changes from an abstract action-type to a concrete action-token. This is one of many difficulties in constructing a detailed hybrid account. [9]

Having briefly touched on the need for, and some challenges posed by, the aim of articulating a hybrid view, let's next examine some issues which bear on, or pave the way for, a critique of the suggested rebuttals, mentioned in my opening paragraph, to the style of argument which leads to it. These responses suggest to me a view according to which indeed there are statable rules determining the correct application of predicates of moral evaluation, and expressing a set of canonical moral principles, so that it is by teaching of and reference to this set of moral principles that the meaning of moral predicates is shared, communicated, and kept stable. The most extreme form of such a view — the view I dub "RFMM" — would hold that the truths belonging to this canonical set of principles somehow collectively distil and encode the foundations of all moral truth, and in so doing, fix, anchor and preserve the meaning of key moral terms.

1.3 Core/periphery approaches to meaning-fixing

That makes RFMM an instance of what I call the "core/periphery" strategy for attempting to rebut arguments from Radical Moral Disagreement, which have a semantic slant aimed at absolutist, cognitivist moral realism. This is the strategy — a thoroughly questionable one, in my opinion — of drawing (often tacitly) a "core/periphery" distinction within moral truths. It is as if Absolutist Cognitivists divide moral claims into two categories (aristocrats and commoners, as it were). Then two steps are taken. *The first of them* denies that disputants "really" disagree about morality provided they don't disagree about "core" (or "aristocratic") moral claims, even if they irrevocably disagree about "peripheral" (or "commoner") claims. *The second step* asserts that

[9] It must be remembered that although "cognitivism" is associated with a view of moral judgments as primarily akin to beliefs, motivational states may be intentional states — propositional (or de se) attitudes — also, and thus also have "cognitive content". Where a moral judgment is a hybrid combination of a desire-like and a belief-like state it is perfectly reasonable to ask what the relationship is between these two kinds of content, and how it is constrained.

Important and diverse suggestions for hybrid theories have been put forward by various authors; see for instance Campbell(2007), Copp(2001), and Ridge(2006). Further discussion of those views is beyond the scope of this paper.

radical moral disagreements only ever involve “peripheral” moral claims. (Perhaps they show, as it were, their lack of breeding, by getting involved in such disorderly disputes.) When it comes to defining, or evidencing grasp of the meaning of moral terms, the commoners are simply discounted. (Naturally these claims are not usually made in *quite* these terms.)

Neither step of this procedure seems credible to me, when it is used in an attempt to rebut arguments from radical moral disagreements. Both steps seem egregiously *ad hoc* and incongruent with the nature of such disagreements. The case is very different from one where predicates are vague. If Ned takes offence at Don walking on his hill, Don might assert that the slightly sloping region around the foot of the hill, where the slope of the ground is 15° or less, does not count as “on” the hill, and Ned may disagree. Merely relying on a mutual grasp of the ordinary word “hill” will not settle the issue. Here it is clear that a resolution must be, if not entirely arbitrary, at least based on some suitably agreed or imposed convention not plausibly deriving solely from the standard meaning of that word. Each party to a radical moral disagreement, on the other hand, is likely to insist that their assertions are based, precisely, on a correct understanding of basic moral terms like “right” and “wrong”, and there is no comparable sense in which adopting a precisifying convention could be taken as “resolving” what is at issue between them. As Solomon taught us, moral disagreements do not always allow us to “split the difference”.

The basic issue here is whether the disputants in radical moral disagreements can reasonably be held by absolutist cognitivists *not* to be using key moral terms with different senses, or at least relativized to different parameters.[10] The essential feature of a radical moral dispute is that *nothing crucial is hidden* from the disputants. Both know the large man was on the bridge, and was pushed to his death in front of the runaway trolley by a small but strong bystander, in order to save — only just — the lives of the seventeen deaf children standing on the track and admiring the view in the opposite direction. Let us set aside the possibility that the disputants disagree because they make different estimates of the probability that the large man would have gone on to find a cure for cancer, or that one of the children will do so. We know enough to know that disagreement in a case like this might persist even if they agree on all such matters. If an absolutist cognitivist view is

[10] In the Appendix I point out the broad range of views I am counting as “relativist”.

correct, and if properties predicated by predicates of moral evaluation do indeed supervene on uncontroversially descriptive properties, then in such cases the disputants, in agreement about the uncontroversially descriptive, display different understandings of the relevant supervenience relations.[11]

But unless moral predicates have a causal-referential semantics involving possibly unknown real essences, are Twin Earthable, and so on, in the way Putnam has taught us to see natural kind terms, the relevant supervenience relations are necessarily *and analytically* correlated with the meanings of the relevant terms. To understand moral terms we need to know more than, merely, that the moral supervenes on the descriptive: we need a suitable grasp of the form of the supervenience relation — and once causal-referential semantics is set aside, that grasp is not only necessary but sufficient to know the meanings. If we have *differing* understandings of the relevant supervenience relations we simply differ in our understandings of those terms — we attach different meanings to them. Modeling this situation in a way which is both consistent with absolutist cognitivism and which treats the disagreement as betraying a local and minor mistake, by one or both parties, in the use or import of moral terminology, rather than revealing a basic meaning-difference, seems far-fetched and, as I have said, *ad hoc.*

Here's one way to cling to absolutism while *accepting* the diagnosis that the disputants are *not* giving the same meaning or content to their moral terminology. Absolutists might hold that at least one party's claims are not moral claims at all, but *"schmoral"* claims – claims put forward as moral, or playing a similar role to moral claims, but which employ what purport to be

[11] What of "thick" terms of moral evaluation, like "courageous" and "heartless"? I shall rely on the assumptions a) that such terms are not themselves "uncontroversially purely descriptive", meaning that claims employing them, made by either party in a moral disagreement, "could reasonably be called question-begging in the context"; and b) that insofar as such a term has non-evaluative descriptive content, that part of its content is equivalent to that of a possible predicate whose application supervenes on the applicability of uncontroversially purely descriptive terms. If those assumptions are right, I believe, it does not affect my discussion whether parties to a radical moral disagreement are or are not using such "thick" evaluative terms in expressing their disagreement. Compare Jackson(1998), pp.135-6, on "centralism" vs. "non-centralism".

moral terms with what are actually *distorted meanings*.[12] At least one disputant is not expressing genuine moral judgments, but *merely schmoral* judgments. But this claim, if it is not backed by viable analyses of the ontology, semantics, and epistemology of moral judgment *which are capable in principle of distinguishing, in any given radical moral disagreement, the moral from the schmoral,* strikes me as unattractively reminiscent of a "pie in the sky" or act of faith rebuttal to the challenge such disagreements pose. Someone who accepts this view, and is *themselves* party to a radical moral disagreement, must either lay claim *hubristically* to an epistemically mysterious ability to distinguish the moral from the schmoral, not shared by their opponent, or agree that *they themselves* are completely in the dark as to the morality of a disputed action — *despite* all the effort they have put into making a judgment based on a knowledge of all seemingly relevant non-moral facts, and long and earnest reflection and debate, and even *despite* having been convinced by these exercises of the correctness of their judgment.

The "pie in the sky" in the case of the large man and the runaway trolley might be thought of as a Platonic lexicon, spelling out canonically the true meanings of moral terms, sufficiently clearly to demonstrate whose judgment, if anyone's, is correct in this case. But it is no more plausible, based on how people enmeshed in stubborn moral disagreements behave, that the disputants are trying to judge, or dimly perceive through some spiritual exercise, how to apply to the case in hand semantic information to be found in the contents of a Platonic lexicon, than that they are debating the evidence for one or another hidden real essence to which their utterances of moral terms are causally linked. And if we don't believe in some such symmetry-breaker we will remain confronted by what appears, from a cognitivist standpoint, to

[12] Milo(1986) provides an interesting example. Milo uses both a "moral/schmoral" distinction, and a *distinct* "core/periphery" distinction, in a slightly unusual way. He classifies what I would call radical moral disagreements into several categories: "moral deadlock", "partial radical moral disagreement", and "total radical moral disagreement". These groups differ according to whether the disputants accept all, some, or none of the same basic moral criteria for rightness or wrongness. Milo argues by use of a core/periphery strategy that moral deadlocks are benign because they are "in-house", since they occur between people with the same basic moral values. That's his "core/periphery" strategy. He argues that the term "radical moral disagreements" is an oxymoron: candidates are merely mock-disagreements between moral and schmoral claims. That's his moral/schmoral strategy.

be evidence in favour of a relativist account of the meanings being expressed by moral terms, or of the contents of moral judgments, in such cases — pushing us, by way of the "secondary puzzle" about radical disagreement, towards the further consequences, including a sizeable concession to non-cognitivism, noted above.

RFMM, as a version of a core / periphery strategy, thus faces serious difficulties from the outset. Nevertheless I believe it is illuminating to explore possible details of such a position, and thus see how certain of its weaknesses emerge. The view may appear to have some *prima facie* strengths. It makes a concrete claim about the transmission of moral knowledge, including knowledge of the meanings of terms of moral evaluation; it represents morality — or at least, the language of moral evaluation — as a human institution maintained and transmitted throughout a community by way of a distinct vocabulary whose use is benchmarked to a finite set of claims. It offers to simplify the task of moral translation: identify the basic moral principles, and concentrate on finding apt translations of those into principles, hopefully already current, expressed in other languages.

Note that the "principles" envisaged here are quite different from the "master principles" in terms of which one might formulate some of the great "isms" of moral theory: Utilitarianism, Kantianism, Intuitionism, Rationalism, etc. Such accounts tend to throw up just one or two basic principles, and though they are often defended in *a priori* terms, they are not really candidates for uncontroversial commonplaces of folk ethics, whereas that is precisely how an RFMM kind of view would wish to see the principles — whether they number ten or a dozen, several dozen, or a few hundred — which such a view sees as canonical.

1.4 More holistic versions of core/periphery strategies?

RFMM contrasts with more holistic approaches to the fixing of moral meanings, where so far as consistently possible, the whole totality of moral claims (including claims of "folk" meta-ethics) — or a very large majority of them — are treated as implicitly term-defining for the distinctively moral terminology. Jackson's "moral functionalism" and the "analytic descriptivism" he bases on it, together provide an instance of such a "holistic" strategy, with an added twist of idealization: the totality of claims in question is drawn not from current "folk morality" but from an ideally improved and corrected, future or hypothetical version of the latter, called "mature folk morality".[13]

Many difficulties attend such an approach. One difficulty, in navigating the path to mature folk morality, is settling the issue of how to allocate weight to different claims of folk morality — a more nuanced relative of the problem of how to demarcate the core from the periphery.[14] Another — highlighted by such issues — is the need to presume that radical moral disagreements, containing claims in danger of being relegated to the periphery, can all be eliminated in the "maturing" process — the need, in other words, for a presumption of convergence. This presumption may itself be, as some have claimed, a part of folk morality, indeed the very part which justifies Jackson's account of the intended referents of moral terms: but it might be false, for all that. Jackson himself admits the possibility in principle of a failure of ideal convergence, and that something like relativism would be the consequence of that failure.[15]

1.5 A trace of RFMM in Jackson

We could describe the unwelcome fall-back outcome Jackson envisages, if moral "convergence" turns out to be a will-o-the-wisp, as one viewing moral discourse as divided into a family of *dialects*. In that case Jackson *might* think the proper goal of moral debate would be to arrive at "mature" versions of each of these "folk moral dialects".

But Jackson expresses optimism about convergence, seeing belief in convergence as itself a presumption of folk morality.[16] So he prefers to frame the agenda in ethics — including folk ethics! — as moving towards a single, unified "mature folk morality". However this view cannot give a credible account of *current* folk moral discourse unless there is *more* commonality in

[13] See Jackson(1998), Robinson(2009).

[14] Jackson says "It is no part of moral functionalism that all parts of the network that is folk morality are equal" — Jackson(1998), p. 134. See Papineau(1996) for a useful discussion of this issue and some of its implications, as it arises in the functionalist definition of theoretical terms in the sciences.

[15] For a critique of Jackson's discussion see Robinson(2009).

[16] Richard Joyce(2011) astutely points out that this puts Jackson in danger of being "an accidental error theorist": if the concept of morality essentially incorporates the idea of convergence, but in fact convergence cannot be guaranteed, morality in Jackson's sense is at risk of not existing.

current folk use of moral terms than simply an implicit meta-ethical standpoint which involves faith in ultimate convergence. Unless we view the folk as univocal in their use of moral terms *merely* in virtue of sharing implicit common knowledge of and commitment to Jacksonian meta-ethics, the problem of securing common meanings has not been dodged if *current* moral discourse at the normative level splits into a multiplicity of "folk moral dialects".

This I think is a good standpoint from which to view the following remarks about the role of "good enough" commonality in fixing shared meanings, from Jackson (1998), pp. 131-32:

> "The principles of folk morality are what we appeal to when we debate moral questions. They are the tenets we regard as settling our moral debates. '… It would be a betrayal of friendship not to testify on Jones's behalf, so I'll testify.' … The dispute-settling nature of such a tenet shows that *at the time in question and relative to the audience with whom we are debating,* the tenet is part of our folk morality. If there were not such benchmarks we could not hold a sensible moral discussion with our fellows. Nevertheless these benchmark tenets are far from immutable …
>
> "What is, though, true is that there is a considerable measure of agreement about the general principles *broadly stated."…*
>
> "We can think of the rather general principles that we share as the commonplaces or platitudes or constitutive principles that make up the core we need to share in order to count as speaking a common moral language…."

Although these brief remarks from Jackson are simply addressing what he sees as being — or at least hopes to be — the merely temporary lack of complete convergence in moral opinions, their relevance to the puzzles associated (from an absolutist cognitivist standpoint) with radical moral disagreements, should be evident. The remarks about "benchmarks" required in order for us to engage in "a sensible moral discussion", and "speaking a common moral language" appear made to order to address the problem of "failing to disagree" raised by such disagreements (namely, the problem that insofar as they provide evidence that the disputants are using moral terms with different cognitive contents, they provide evidence that the supposed disagreement is merely a case of talking past one another).

Notable points in what Jackson says include the following. Firstly, the principles are represented as "dispute-settling", suggesting they have some canonical status equipping them to over-ride particular judgments. Secondly, although Jackson, given his assumptions, says they are "mutable", he nevertheless – as if to correct any mistaken impression this may give – says that "what is though true is that there is a considerable measure of agreement about the general principles *broadly stated".* The overall picture is one according to which there is *enough* agreement to secure a common language, and to provide the wherewithal to settle moral disputes, by appeal to agreed "benchmarks".

These are the remarks from Jackson which are reminiscent, arguably, of the position I call "Rules Fix Moral Meanings". To briefly sum up the essential core of Part 1, I have claimed that a straightforward construal of cases of radical moral disagreement, consistent with cognitivism, is that there is variability in the cognitive content (i.e., in the truth conditions) assigned by the disputants to moral claims about which they disagree. If so, yet the disputants are making *bona fide* moral claims, then absolutism is in trouble. Since a consequence of retaining cognitivism under those circumstances is that people do not have a genuine disagreement, cognitivism is also in trouble. A common diagnosis is that the judgment that one party has and which the other party opposes is not, or not only, a belief, but is at least in part another kind of judgment containing expressivist or imperativist elements such as non-cognitivists invoke. The RFMM proposal for avoiding these consequences is to hold that the cognitive contents of the disputants' opposing claims do render them genuinely in contradiction with one another, because the meanings of all such claims are governed by a finite set of moral principles which canonically determine the meaning of moral terms they contain. Thus those terms have common meanings in the disputants' mouths regardless of how they actually view the truth or falsity, in the relevant circumstances, of the disputed claims.

Part 2 Putting Moral Principles in their Place

2.1 Arguing against RFMM

In what follows I shall present three arguments against RFMM, two of them brief, and one of them more extended. These arguments will in turn be followed up with some concluding remarks expressing skepticism about the special status of moral rules or principles generally.

The general implication of all these arguments will be that moral understanding and commitment are not well represented as grasp of and commitment to a set of rules or principles. This point is highlighted by illustrating some different ways in which members of a group might reasonably be said to be mutually committed to a common set of moral principles, yet be capable of radical moral disagreement. These disagreements might arise between members of such a group, simply because of differences in how they respond to cases in which distinct individual rules apply, and clash. Only by dismissing those differences as reflecting someone's misuse of moral terms by their own lights, thus not *bona fide* manifestations of moral understanding, could we retain the view that two people's commitment to the same moral principles suffices to demonstrate that their moral concepts, or the meanings they give their moral terms, are identical. But I shall argue that in such cases the differences clearly reflect differences in moral standpoint and are thus not well-represented as products of confusion. The deeper message here is that on any realistic account, even those who are committed to particular moral rules need to make moral judgments in the process of applying them to particular circumstances, so that the content of morality cannot realistically be encapsulated in a set of rules.

Perhaps RFMM itself is consistent with relativism. Speakers of different moral dialects might be committed to different sets of moral principles, governing the meanings of the moral terms in those dialects. But here we are considering the possibility of employing RFMM to defend *absolutist* cognitivism against threats posed by the possibility of radical moral disagreements. If RFMM cannot stand up as a doctrine it can render no such service.

2.2 Two brief arguments against RFMM

2.2a Argument 1: Rules can be challenged without incoherence

The first argument against such a use of RFMM is a very simple one. It is a near relative of the Open Question argument. The premise is that someone might coherently (even if mistakenly) raise a moral objection to the absolutist's chosen set of moral principles. If an absolutist version of RFMM were true, this would involve some sort of self-contradiction, similar to claiming the standard metre (by a current definition) not to be 1 metre in length. This quick argument seems good to me, but might not impress those who dislike the Open Question argument. For argument's sake let's set it aside and move on to further objections.

2.2b Argument 2: Axiomatic systems as an inept model for RFMM

So secondly, let's consider a model for moral error which, cast in this role, RFMM might seem to require of us. At least in the case of radical moral disagreements, absolutists must assert that at least one of the parties is *somehow* in error. We assume that they share a commitment to the same term-defining moral principles, so by the standards of RFMM they are equally competent in use of the same moral terms. By definition of radical moral disagreements, the disagreement cannot be consequent on a purely descriptive factual error. It must be a cognitive error in which moral terms are applied in a manner inconsistent with their actual meaning (a performance error, not a competence error, if you like).

One available model, or metaphor, which might be applied here, is provided by a formally axiomatized theory taken as term-defining for a set of theoretical terms figuring in the axioms. Given that the axioms define the meanings of the terms, one who grasps and accepts the axioms may be said to know the meanings of the terms. Nevertheless it may not always be clear – indeed it may not even be logically decidable – what exactly is entailed by those axioms, making it easy enough for different parties to a dispute to have a common commitment to the meanings of the relevant terms, yet differ in their judgments as to what that entails *vis a vis* particular cases.

Thus we have a proposed explanation for moral disagreement, and we have some kind of reason for thinking that further thought and deliberation might in principle lead to agreement, since that would seem to be the right way to overcome confusions about what entails what. Indeed insofar as we find evidence that we subscribe to common axioms, we will seem to have evidence for some optimism of that kind.

It seems to me that the implausibility of this model becomes evident as soon as we start to scrutinize it seriously. We are not generally taught an axiomatized theory of this kind in the course of our moral upbringing, nor do I believe our efforts to resolve moral disagreement typically take the form of running over complex patterns of deduction looking for new proofs, or slips in our logic. Nor should they. Such exercises would draw our attention away from the disputed cases and their evident morally relevant features. If someone who has seriously reflected on what is involved in a disputed and morally problematic case were to be shown that a contrary opinion to theirs could be deduced, by complicated and far from obvious logical reasoning, from some alleged set of moral axioms, an obvious candidate response for them would be to apply *modus tollens* and reject or qualify one or more of those axioms. It doesn't take exceptional *logical* acumen to appreciate what is at stake in cases like that of the large man on the bridge: radical moral disagreements do not generally hinge on unusual complexity or subtle deductive fallacies.

2.3 Argument 3: A Spectrum of Diverse Roles for Sets of Rules, and its implications

The slightly more elaborate argument to which I now turn, will, I hope, bring out more fully why this highly idealized, but perhaps initially tempting, axiomatic-system model, does not really give a credible account of radical moral disagreement. I want to highlight how much can be left *unsettled* by *acceptance of,* or *conformity to,* a set of moral principles. I shall suggest that spelling out the ambiguity in such ideas of "acceptance" or "conformity", highlights the extent to which making moral judgments cannot realistically be portrayed as essentially consisting of applications of rules to cases. The argument begins by listing some options on a spectrum. Each item on the spectrum (which by no means claims to be exhaustive) provides a different sketch of how one might understand what is involved in commitment to a set of moral principles, in relation to determining permissible and obligatory actions. I deliberately give very abstract formulations. I hope the list is of some interest in its own right.

What follows revolves centrally around the following claim. *Commitment to rules underdetermines moral judgment: shared commitment to rules underdetermines moral agreement.*

2.31 A spectrum of options

2.31A Spectrum, Category A: Rules without Degrees

In these cases a set of rules is taken as absolute in the sense of admitting no degree. A rule either applies to a case or it does not. Any action is either in violation of this set of rules, or not. But in all but one of the following options, individual rules may be "trumped" or "disabled" by others in the set, based on an ordering which is either fixed, or which undergoes variation according to the situation.

Ai: Strict Absolute Rules with Moral Traps

There are absolute moral principles, each saying either that a certain type of action is wrong without exception – impermissible – or saying that a certain type of action is (in specified circumstances) obligatory. Let's think of the rules as numbered 1, 2, etc., and the corresponding action-types as numbered A1, A2, etc. Suppose it is possible to find oneself in circumstances such that one is obliged to perform an action of type A1, but the only available way of doing so is to perform an action which is also of type A2, which is impermissible. Then it would be possible to find oneself faced with no alternative but to do wrong – either by failing to perform an obligatory act, or by performing an impermissible act. This view accepts that consequence: sometimes one can *only* do wrong. Avoiding wrongdoing may require substantial care and anticipation in steering clear of these moral traps!

Aii: Moral rules with fixed ranking

The next, perhaps more humane, point on our spectrum is one in which principles are absolute in only a slightly weaker sense. The enumeration of principles now represents a priority ranking. Actions must be rigidly in accord with all principles except that no principle is to be followed if doing so would violate a higher-ranking principle — unless the higher-ranking principle is already outranked by one still higher-ranking, which mandates the action. One never acts wrongly if one acts in accordance with the ranking of principles in this sense. Where rules conflict, lesser-ranked rules are, so to speak, "switched off".

Aiii: Moral rules with contextually variable rankings:

The next possibility is exactly like the previous *except* that the ranking of principles is not fixed, but varies in a specific, pre-determined way according to context. (To take a silly illustrative example: perhaps "never consort with murderers" usually overrides "do not spurn offers of hospitality", but the order is reversed when the hospitality is offered by a murderer who happens to be one's father-in-law.)

Aiv: Moral rules with contextual-feature and action-feature-dependent variable rankings:

This is like the previous case, with an added, perhaps artificial, degree of complexity. But I believe it deserves notice as a possibility. For brevity I combine two variations on the previous case. The first is simple. Rather than a finite set of context-types, each of which induces a different priority-ranking for the moral principles, we imagine a number of contextual features which can be present in varying degrees. It is the combination of such features which determines the appropriate ranking of moral rules in a given situation.

Perhaps this could be crudely modeled by imagining an algorithm which takes the degrees to which various features are present, yielding a priority-ranking, but it would be more realistic to imagine that actual moral agents weigh up and judge the relevance of different features in a manner which is not overtly quantitative.

Illustrative example: "This occasion is in part a religious function, it has political significance and public interest, but it is also to an important degree a celebratory social gathering. Am I required to observe the rule — prioritized at religious functions and public political events — forbidding the telling of risqué jokes in the presence of clergy, or is it in this context trumped by the rule requiring one to share in the fun on social occasions?".

The second variation on the previous case is slightly more complicated and more conjectural. There are features of actions themselves which strike me as somewhat like action "contexts", in that, plausibly, they are not reasonably thought of as essential to action types as such. Thus a chosen action may have morally relevant features, capable of being present in varying degrees, over and above what is minimally required if no applicable rule is to be broken. There are, as one says, many ways to skin a cat, and since purely generic actions are impossible, any action actually performed will have many features, intended or not, inessential to the primary description under which it is intended. Even after context has been taken into account, the degree to which these features are present in a chosen action may affect the ranking of principles governing retrospective moral evaluation of that action.

Illustrative example: "The doctor was morally obliged to calm down the elderly lady in the church who was having a panic attack — knowing it would be putting her weak heart at risk. He could have done this without telling a risqué joke (so as to break the tension), so he was under no rule-governed obligation to break the prohibition on telling such jokes in church. Nevertheless, he should not be counted as acting wrongly given that by doing so he drew attention away from the old lady and onto himself, thereby sparing her the further embarrassment which, had he taken more orthodox action, would otherwise have ensued. 'Minimize embarrassment to others' would not normally override 'do not tell risqué jokes in church', but given how mortified she would have been to have him make her breathe into a brown-paper bag in public, the considerate *manner* in which he carried out an action of a type which was in any case obligatory is morally permissible (as it would not have been had he muffed it and merely exacerbated her panic)."

That completes my short list of *Category A* options. Two features of them should be specially noted. Firstly, it is likely that in a great majority of cases the differences between these options would be of no account — invisible, in effect — because in those cases no applicable moral principles clash. So a population of people each of whom occupies some place on this spectrum with respect to the same set of rules might often present an appearance of strong moral consensus. Secondly, with the exception of the first ("moral traps") option, they all resolve clashes between principles, should they arise, by employing rankings, fixed or variable, to *disable* some of the clashing principles. They are all in that sense consistent with the view that moral principles, when not disabled, are absolute: they brook no exceptions and involve no relativity and no watering-down. A person adhering to any of these patterns might reasonably be said to be committed to, or at least to conform to, the rules in the relevant set.

2.31B Spectrum, Category B: Moral rules with weights and degrees

Morally relevant component features:

Another kind of possibility arises if, instead of taking it that, in a given case, a rule either applies totally or not at all, we think of moral principles as rules to which an action might conform *in varying degrees.* Instead of options like the above, one might see the moral merit of an action as determined by reference both to the context of action and the particular degree to which a given action (taking into account the manner of performing it) instantiates distinct morally relevant features corresponding to different principles. Evaluation would be based, roughly speaking, on whether an action, given its context, maximizes overall conformity to those principles. Once again, rankings of principles might be relevant, and once again they might be regarded as fixed or as variable for a range of quantifiable or unquantifiable reasons. But rather than determining which rules can "switch off" others they outrank, the rankings will affect how much weight the various moral principles carry, in judging an action's overall moral merit. On this view, one could say, correct evaluation of an action — its moral goodness — is determined by the contextually appropriate weight assigned to the morally relevant features, and the degree to which they are instantiated by the action. Once again, there is no reason to think there is any absolute correct way of quantifying diverse kinds of morally relevant features ("the action was unkind, but *how* unkind was it?"), nor to think that if there were, any particular way of aggregating such numbers would be uncontroversially mandatory.

Further uncertainties arise as soon as we contemplate evaluating choice of action by asking which available action had the highest moral value. “Choose the best available action” (where “best” is evaluated in this manner) is a dubious rule, since at any given moment, skill and ingenuity are not *entirely* under voluntary control (as professional sportspeople are well aware, since an “off day” can be very costly for them). But ingenuity may be required to perceive a possible action, and skill to carry it off — for instance, in saving a life. This makes it in various respects a *vague* rule — for instance, it does not say how to incorporate perceived risk of failure, in deciding which action is “available”. (It may useful here to reflect on the question of whether it is sometimes a vice to be a “klutz”. Someone who frequently rushes to help but most often muffs it – not merely failing to catch the falling sugar, but spilling the milk as well – needs to learn that the moral credit accruing to “helpful” people may not really be theirs to claim.)

For present purposes there is no special point in trying to enumerate variations of Category B approaches. It’s evident, I think, that there are countless ways in which people who respond similarly to similar morally relevant features, and who even mostly give them similar priority rankings, might respond differently to complex combinations of those features in various degrees. It’s also evident, surely, that the range of morally relevant features is quite large, so that one aspect of the *prima facie* appeal of RFMM — the simplicity of the picture it presents of moral evaluation — is considerably lessened once we start equating *conforming to moral principles* with *responding to morally relevant features.*[17]

2.32 Notable points

As will perhaps be obvious, the above list of options is designed to have certain features.

Firstly, every option is meant to represent a way in which correct moral evaluation of someone’s actions might be understood as governed by that action’s conformity to some set of principles.

Secondly, as we go down the list, an account of how the moral status of an action relates to the relevant set of moral principles must increasingly refer to other matters not captured in the principles themselves. Furthermore, those “other matters” – such things as the context of action, the fine-detailed

[17] Milo(1986) comes close to such an identification, seeming to equate *criteria of moral relevance* with *moral standards.*

description of the concrete action itself, and / or the degree to which the various principles apply to the action – appear to me, more or less increasingly as we go down the list, to have three related features. The first is that it is not generally plausible that they are all matters susceptible of being precisely measured or quantified. The second is that it is even less plausible that insofar as we take such matters into account in dealing with recalcitrant moral matters, we actually do measure or quantify them. The third is that the options seem to me to become increasingly similar (or decreasingly dissimilar) to how things work in real life — though this is an empirical claim which some might dispute, and I cannot speak for the experience of others.

Thirdly, if two people agree that such and such is a correct list of moral principles – or generally both seem to show conformity to such a list in their actions — we have as yet little reason to predict that they will never disagree about difficult cases. Neither we nor they themselves may know how, when the chips are down, they will resolve conflicts between the principles. Even on an option as simple as A(ii), all it will take to leave people who mostly seem to conform to exactly the same principles, potentially in disagreement over certain difficult cases will be a difference in their ranking of the same moral principles. The number of possible variations increases enormously as the number of rules increases. (Consider: "Yes, I accept that violent criminals should be apprehended and that we should help the police to do so. Yes, I accept that the criminal was also wrong to steal food from the store. Under most circumstances I would have called the police. But I also believe children should be kept well-fed. You believe that too. But I gave it top priority and I knew the food was destined for the criminal's hungry children. You don't think that overrides the other considerations; I do.")

Fourthly, if it's true that as we go down the list the options become more lifelike, then it seems to me we also move closer to a particularist vision of moral evaluation. According to most Category A positions on our spectrum, and with Category B positions even more so, what is required beyond simple matters of fact on the one hand, and moral principles on the other, in order not to leave indeterminate the moral evaluation appropriate to some particular action, is determination of a number of *issues which themselves involve various kinds of evaluation.* These issues may include the relative priority of principles, the degree to which various aspects of a situation or context affect that priority, the extent to which the particular details of a particular chosen action constitute it as an act falling under one or another principle, and to what degree or extent the act considered fully in its details and in its context, counts as maximizing the suitably weighted and combined overall satisfaction of the relevant principles.

It would be idle to pretend that these further matters of evaluation may themselves be settled in general by appeal to further principles. But what would be still more idle would be to pretend that we can optimistically hope that our typical community of moral agents agrees in their even implicit acceptance and application of a sufficient set of such second-order principles. Our ordinary ways of managing involve making summary evaluative judgments which at best only implicitly incorporate judgments about these second-order matters.

Thus we reach the conclusion of this third argument. By now the cat of moral meaning has got out of the bag of moral principles. *Any realistic criterion of assent or conformity to a set of moral rules or principles will leave far too much undetermined when it comes to how people will evaluate particular actions or situations, and what is left undetermined is clearly part and parcel of what evinces, in agents' choices or evaluations of actions, their understanding of the content of, and the requirements imposed by, moral notions.* If there is any such thing as the constitutive "core" of moral understanding, it cannot be characterized in terms of adherence to a small set of principles.

Hence, to repeat a central contention of this paper, it is not very plausible that arguments for relativism from moral disagreement may be rebutted by pointing to large numbers of general moral principles on which there is agreement. Insofar as we can see in this a demotion in the status of general moral principles, perhaps we can see it as at least a step also in the direction of particularism, whether or not that destination, strictly speaking, is reached.

2.4 Reflective equilibrium, particular judgments, and a demotion of rules

Before trying to sum up an attitude to moral rules I think is consistent with the above discussion, I need to touch on a couple of other matters. The first of these is the notion of reflective equilibrium — both because, ever since its introduction by Rawls[18], it has become a mainstay of discussion about moral disagreement and moral deliberation, and because, since moral rules figure in its characterization, it provides a useful context in which to raise further skeptical questions about moral rules which carry over to prevailing ideas of the nature and role of reflective equilibrium itself. I've argued that moral

[18] Rawls(1951).

rules or principles are not alone *sufficient* to fix the meanings of moral terms or capture the core of moral understanding. We can now ask whether, and in what sense, moral rules or principles are even *necessary* to moral understanding.

The notion of reflective equilibrium in part pictures a struggle in which particular moral evaluations are sometimes pitted against moral principles, adjustments being made on one side or the other (with whatever help moral theory has to lend) until equilibrium — signifying coherence — is achieved. If these two kinds of judgment may be pitted against one another, then one might wonder whether they have different origins or are grounded in different ways. Are particular moral judgments more or less firmly grounded than judgments about moral rules, and is the grounding of the same kind, or different, in the two kinds of case? Is accepting a moral rule itself a kind of moral judgment? Or is it just accepting a summary of actual or possible moral judgments? Is a moral rule a tool for producing moral judgments? Or is it, as is very commonly suggested (for instance by Michael Smith[19], who makes the idea of reflective equilibrium central to his notion of "normative rationality") a tool for *justifying* moral judgments? If someone was reliably capable of making particular moral judgments without invoking rules, what value would be added to their moral understanding were they to come in addition to accept various moral principles? Is unity of values truly a virtue?

Smith's view is that the value added is indeed the value of *unity,* providing a kind of coherentist *justification* which he sees as important to normative rationality. Others may think that the very abstractness and generality of moral rules makes judgments about them more error-prone. After all, it is implicit in the above discussion of ways in which rules may figure in moral evaluation, that the complexity of particular real life situations is always likely to outrun what can be contained in a rule or even a set of rules. The devil is in the details. Therefore, if rules extrapolate from known to further particular situations which they help to evaluate, there is a real risk that they will lead us astray, due to a mistaken application of the human *penchant* for simplifying things in order to make them neat and orderly. If on the other hand rules merely summarize what we already know, it may be thought, they cannot "add value".

These are large and important questions, which probably have as many

[19] In Smith(1992). See for instance pp. 159-160.

answers as there are kinds of moral theory. Shortly, I will simply express my own attitude, which is hinted at in the last sentence of the preceding paragraph — namely that rules are best thought of as mere summaries which do not explain, justify, or "add value" — and try, not so much to defend it as to merely make it somewhat comprehensible, by the use of a couple of philosophical analogies.

But in order that my view should not seem altogether far-fetched, I need to make an important caveat. Naturally I accept that social life, with even a few basic shared goals, and particularly given any remotely egalitarian ideal, imposes heavy demands on people to conform, more or less, and to expect others to conform, more or less, to some mutually known set of rules. This is an important reason for putting up with much of what governments do, and in particular, for accepting the necessity for a legal system and for an apparatus of police, lawyers, judges and so forth, to operate it. That system, specially in societies which consider themselves not to be very corrupt, is often called "the Justice System". Such a system is usually imperfect in that, much like many an economic system, it does not always produce outcomes of which we are prone to morally approve. "The law is an ass", and "there is a moral as well as a legal code" are proverbial reminders of that fact. "The Justice System" is in fact not always just. Nevertheless, it might happen that, all things considered, it's possible that there should be a legal code, and ways of arriving at and administering it, which might seem to be reasonably fair (particularly if embedded in a fair economic system, relieving the "Justice System" of the task of defending material inequities).

It is a consequence of what I am suggesting that even the best such legal code and justice system, given the simple necessity of operating according to rules and principles which are, pretty much, common knowledge, may need to enforce legal judgments which have, by ordinary standards, ethically regrettable consequences (or at the very least, consequences which could only be viewed as not ethically regrettable after factoring in the desirability of having such a justice system at all). In short, it is likely that even the best system of law — even despite the human elements interpolated by judges and juries — will *inevitably*, at times, be "an ass".

But all that, surely, does not apply to "the moral code" itself? I think perhaps to some extent it *does*, and inevitably so, if we take the phrase "the moral code" too literally.[20] It is a frequent suggestion that "morality" is something shared which exists for the common good, rather like a more subtle and (insofar as not legally enforced) a more voluntaristic legal code. But if I am right about moral principles, there will be an inevitable possibility for internal tensions between the requirements of a shared and commonly known "moral code", conceived as encapsulated in a set of moral principles. Thus it might seem inevitable, given what I have said about moral principles, that even "the moral code" will at times also be "an ass".

But how could what is morally prescribed be morally wrong? Does this line of thought lead to a quick *reductio?*

There is a wrong way and a right way to respond to this challenge. Only the wrong way, which rests with the idea of a legal-system-like moral code, leads to a *reductio.* The right way, is to see that *ideally* what we would like to see in a shared social morality is a tendency to agree in moral judgments *even* where there is *no* deductive path from shared moral principles to a particular judgment. We should *not* think of morality, viewed communally, as literally a "code" similar to the legal code. A shared capacity for particularized and contextually appropriate moral judgment should be seen as part, ideally, of what morally binds a community.

Still and all, the message to be drawn from the possibility of radical moral disagreements is that we cannot reasonably hope to have a social morality which is so perfectly shared that like telepaths, or a certain stereotype of identical twins, we all always concur in our moral judgments, with or without the aid of rules. This is in part *good* news. One of the functions of moral language is to enable us to *express our moral differences.* We do not wish to have a blanket permission to rule moral critics out of court on the grounds that, tautologously, they are misusing moral language and making merely *schmoral* objections.[21] This is something we should not only learn, but welcome, from

[20] It seems to me that some do. Compare for instance the reference to "a set of principles" in this contractualist statement from Scanlon: "An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement." (Scanlon 1998, p. 153).

[21] A view which Milo(1986) seems to flirt with. I should emphasize that it is consistent with my view that *sometimes* it can be right to dismiss a proposed alternative moral view as a "merely schmoral" claim. I believe that is implicit in what

the perennial relevance of the Open Question argument. Losing the permanent possibility of moral disagreement would mean losing one of the central functions of moral language in expressing some of our most important differences.

Let me finish, then with a couple of metaphors and analogies, in which I shall attempt to bring to life the kind of view I am urging, of moral rules. First, a metaphor. Let's consider a sort of distant analogy with the Euthyphro issue. Suppose the gods say that we ought to do X. We might decide that it is true that we ought to do X, but not because the gods say we ought to do it, rather the reverse: the gods say we ought to do X, because we ought to do X. If the gods always get this right, it will be true for any X, that if the gods say that we ought to do X then we ought to do X. Yet this makes them moral authorities only in a demoted sense: they do not make moral truth, merely report it. Suppose we have some alternative account of what does make moral truth, which implies that we are in principle (if we work at it) well-placed to judge such matters for ourselves. Someone might say "I was confused about what to do and I sat down to make a considered judgment – but I was so tired: in the end I just took the gods' advice – they usually get it right". This would be to demote the gods to the status of something like a personal database or GPS system. For an idea of the view I am suggesting, substitute "moral principles" for "the gods", in the above.

A similar point can be made by reference to David Lewis's account of laws of nature, which goes roughly as follows. Given all the truths there are, expressed in a suitably naturalistic language, there might be a deductively-closed system containing only a subset of the truths, but which by suitable standards combines strength and simplicity better than any true rival system. The laws of nature are all the contingent generalizations belonging to that deductive system. On such a view, a law of nature is part of a neat summary of a large and highly patterned portion of the truth. But neither such a law, nor any of its consequences, is true in any different sense, or on any other grounds, from other contingent truths. Logically, all facts are on a par. If some fact conflicts with what would otherwise be a law, there is simply no contest: there is no such law. Lewis's view is quite compatible with the *existence* of laws of nature: they simply do no work in making anything true.

I have said about the ineliminable cognitive element in moral claims. But we don't want a view which makes this response *regularly* available.

Similarly, there may be true moral principles. Yet it may be the case that they provide no *independent* source of motivation or moral justification, and are best regarded as simply useful (and perhaps only roughly accurate) summaries of, or guides to, the truth of groups of particular moral judgments or evaluations. In this case, if a particular moral judgment comes into conflict with acceptance of a moral rule, the rule is what should always be discarded.

Finally, an example directed towards those who believe that moral principles must play an *indispensible* role if people are to make moral judgments, be taught how to make them, understand moral concepts, and generally to make communicative use of moral vocabulary, without being at cross-purposes.

Consider beauty. We make judgments of beauty, learn to employ the concept, learn to communicate about beauty, agree about some instances, disagree about others, and so on. Manifestly, none of this requires any of us to formulate "rules of beauty". Nor would we be likely to feel tempted to revise our judgments of beauty merely because someone pointed out that they were in breach of some supposed rule. *Even* if rules were devised by careful observation of the causal regularities governing our prior judgments of beauty and their relationship to (for instance) patterns of visual stimuli, we would feel under no particular compulsion to make further judgments in conformity with them.

It's important to note that when we judge an art-work, a person, or a scene to be beautiful, we may point out features which we take to make an important contribution to the beauty we see. Obviously we often do this. But it would be mistaken to infer that we thereby justify our judgment by appeal to some principle to the effect that such and such a kind of feature contributes to beauty. It is not like that at all, and it is our knowledge of this which explains why (if we do) we find the idea of rules of beauty ludicrous.

I don't suggest that there is a complete analogy between the concept of the beautiful, and central moral concepts. Moral judgments have a practical importance which judgments of beauty mostly do not. Perhaps that difference is crucial – perhaps the practical importance of moral judgments requires us to try to formulate and act in accordance with moral principles. But that would need to be argued. It is true that a presumption of objectivity seems to be an element of much folk moral discourse, whereas it is proverbial that "beauty is in the eye of the beholder". Nevertheless I think the point stands that there can be no quick and easy argument from the learnability of an evaluative concept and our capacity to communicate by use of a term expressing it, to the claim that we must at some level know and appeal to general principles stating necessary or sufficient conditions for its application. The crucial claim the analogy is meant to support is that not only are rules, contrary to RFMM, insufficient to fix agreement in principle about the extensions (and hence, to fix agreement about the meanings) of terms of moral evaluation: even in those many cases where agreement about the extension of those terms is to be found, rules have no essential role in bringing this about.

So, I conclude, there may be true and useful moral generalizations. Perhaps there are fundamental component determinants of goodness and badness, or of rightness and wrongness, which, contra some versions of particularism, never change their valency, so that for each of them, there is a kind of moral generalization one might formulate, to that effect. For all that, I say, we should *not* think of moral principles as having some sort of special and distinctive warrant, ground, or authority, which makes it appropriate for them to act as a corrective to particular moral judgments.[22]

Terminological Appendix.

I see arguments from radical moral disagreement as challenging in the first instance a conjunction of interwoven doctrines, including moral cognitivism, realism, descriptivism, and absolutism, where the latter is taken to exclude both relativism and subjectivism. But because it highlights the doctrines I *most* wish to challenge, I use the short phrase “absolutist cognitivism” to refer to this complex conjunction of doctrines, though other conjuncts (specially realism!) should not be forgotten.

[22] This paper began life (in a very different form) as a contribution to a Workshop on Normativity and Mentality in a World of Causality, at the Swedish Collegium for Advanced Study, Uppsala, September 25-26, 2008. I am grateful to Folke Tersman and to the Swedish Collegium for Advanced Study, for making it possible for me to visit Uppsala for the Workshop. I am also indebted to the Philosophy Department at Victoria University (Wellington) for an invitation to deliver a version of this paper to their Departmental Research seminar in June, 2009. Philosophically I am indebted to the participants in the Uppsala Workshop and at the Victoria University seminar, and I am specially indebted for philosophical input and support to John Bishop, Bronwyn Finnigan, Fred Kroon, and Folke Tersman. Thanks to John Davis for helpful comments on a penultimate draft.

My use of some of the terminology in the above paragraph needs briefly clarifying.

"Realism" here includes both the denial of semantic anti-realism for moral statements, and the denial of an "error-theoretic" assignment of falsehood to all substantial moral claims.

I follow Jackson (1998) in using the term *"descriptivism"* as a substitute for Moore's "naturalism" (the latter term is currently over-worked and has inappropriate contemporary connotations relating to the role of natural science, and the like). The crucial claim for present purposes is that cognitivists should accept as a priori that moral properties necessarily supervene on "purely descriptive" properties. Jackson concludes — and rightly so, on an appropriate conception of properties — that given this supervenience, moral properties must be identified with purely descriptive properties.

Whether "response-dependent" accounts of the meanings of moral terms are consistent with *"absolutism"* will hinge on their success in articulating a viable notion of "ideal rationality", "counterfactually ideal selves", or some such, adequate to sustain the doctrine that those making moral claims may be viewed as committed to understanding their truth-conditions as jointly tracking a hypothetical consensus of idealized responses, rather than merely individual responses. I'm pessimistic about this idea of a hypothetical "factoring out" of the subjectivity inherent in the idea of response-dependence.

Throughout, I am supposing that according to *subjectivism,* the truth (or propriety) of people's moral judgments supervenes on their mental states, and that according to *relativism* people's moral judgments are true (or appropriate) only relative to some parameter, or some index (some element of indexicality) which might or might not be individual, mental, or reflexive. (Thus I am using it in a far more inclusive sense than, for instance, "cultural relativism".)

I shall not try to give more finely-tuned accounts of these meta-ethical doctrines. A useful discussion of their complex relationships is to be found in Chapter One of Tersman(2006).

Denis Robinson

University of Auckland

Revised February 4, 2012.

References:

Campbell, R. 2007. "What is Moral Judgment?". *The Journal of Philosophy* 104: pp. 321-349.

Copp, David. 2001. "Realist Expressivism — A Neglected Option for Moral Realism". *Social Philosophy and Policy*, 18: pp. 1–43.

Jackson, Frank. 1998. *From Metaphysics to Ethics.* Oxford: Clarendon Press.

Joyce, Richard. 2011. "The Accidental Error Theorist", in Shafer-Landau, R. *Oxford Studies in Metaethics*, Volume 6, Oxford: Oxford University Press, 2011.

Milo, Ronald D. 1986. "Moral Deadlock." *Philosophy,* vol. 61 no. 238, pp. 453-471.

Papineau, David. 1996. "Theory-Dependent Terms." *Philosophy of Science,* Vol. 63, no. 1, pp. 1-20.

Rawls, John. 1951. "Outline for a Decision Procedure for Ethics." *Philosophical Review,* vol. 60, no. 2, pp. 177-197.

Ridge, Michael. 2006. "Ecumenical Expressivism: The Best of Both Worlds?" in Shafer-Landau, R. *Oxford Studies in Metaethics*, Volume 2, Oxford: Oxford University Press, 2006.

Robinson, Denis. 2004. "Failing to Agree or Failing to Disagree? — Personal Identity Quasi-Relativism", *The Monist,* vol. 87, no. 4, pp. 512-536.

Robinson, Denis. 2009. "Moral Functionalism, Ethical Quasi-Relativism, and the Canberra Plan." In *Conceptual Analysis and Philosophical Naturalism,* ed. David Braddon-Mitchell and Robert Nola, Cambridge, MA: MIT Press, 2009.

Scanlon, Tim. 1998. *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.

Smith, Michael. 1994. *The Moral Problem.* Oxford: Blackwell.

Tersman, Folke. 2006. *Moral Disagreement.* Cambridge: Cambridge University Press.