14 Moral Functionalism, Ethical Quasi-Relativism, and the Canberra Plan

Denis Robinson

1 Agenda

In this essay I pursue a double agenda. I believe the items on that agenda complement one another. One item is to illuminate some issues relating to the Canberra Plan. My main focus will not be on generic issues, but on matters specific to the distinctively evaluative and normative domain of ethics. The other item is to promote a view—more properly, a sketch of a view—I favor, which I dub *ethical quasi-relativism*.[1] I shall discuss, as primary representative of a "Canberra Plan" treatment of ethics, Frank Jackson's version of "moral functionalism," especially as it is set out in his *From Metaphysics to Ethics* (Jackson 1998a).[2] My discussion will follow a some-what zig-zag course as I compare my view with Jackson's, and consider various issues relevant to his views and their similarities and differences. My thought is that a view like mine emerges fairly naturally if we take a paradigmatically "Canberra Plan" view such as Jackson's, acknowledge certain problems for it, and revise our view accordingly. If that thought is even near the mark, each view stands to be illuminated by the comparison.[3]

The Canberra Plan typically operates within an implicit framework that views ordinary speakers of some problematic discourse as holding a straight-forwardly descriptive theory of the world and how it works, implicit in that discourse, and seeks to determine whether and under what constraints the world does, or even could, conform to such a theory. Clearly a Can-berra Plan treatment of ethics (much as with any descriptivist treatment of ethics) will confront the question of how to accommodate the evaluative and normative dimensions of ethical discourse. One main focus of debate must be the "Open Question" argument, and the difficulties of inferring 'ought' statements from 'is' statements. Another must be the "internalism"

constraint, which insists that there is a necessary connection between moral judgment and motivation. Although Jackson has interesting and illuminating things to say about these, I shall not say much about them. A third focus—which will be the principal focus of my discussion—is on the role of assertions and denials of moral claims, in giving expression to certain kinds of disagreement.

When we try to characterize moral disagreements, a special subset of them (at least) proves difficult and puzzling. I shall defend a view that attempts to make room explicitly for disagreements of this distinctive and puzzling kind. I will claim that to do so, we must take a view of how moral concepts are to be described and individuated that is more sophisticated than the view a standard Canberra Plan approach, such as Jackson's, makes available.

2   Relativism, Disagreement, and Failures to Disagree

I'll start by considering some issues about agreements and disagreements. What I'll call "simple" relativism about *X* says that (i) truths about *X* are relative to agents or standpoints, and (ii) that apparent disagreements about *X*, where due to different standpoints, are therefore really nondis-agreements. Quasi-relativism about *X* differs (in a way) about (ii): hence the prefix 'quasi'. Thus *ethical* quasi-relativism claims differing concepts of right, wrong, good, evil, permissible, impermissible, etc., may be legiti-mate—unfaithful neither (a) to facts nor (b) to established usage, nor (c) to other legitimate constraints, *at least by non-question-begging standards* (the *no-fault condition*).[4] But it claims that nevertheless, from such diver-gences, a kind of *bona fide* disagreement springs (the *bona fide disagreement condition*).

But there is a puzzle in how the no-fault and *bona fide* disagreement conditions can be met simultaneously. Indeed on too narrow a view of what counts as disagreement, I don't think they can be. To progress, we need to examine not only disagreements, but also what I'll call '*failures* to disagree'. To this end I'll start with some illustrative dialogues between characters named 'Argle' and 'Bargle'. (Think of these as surnames shared by members of two argumentative clans.)

First, then, some failures to disagree, or, as I'll call them, 'mere argie-bargie'. Frank Ramsey left us a benchmark case (see Ramsey 1990, 247) that sets some kind of relatively low point on a scale of perplexingness:

Ramsey's case

Argle: I went to Grantchester this afternoon.
Bargle: No I didn't.

This case hinges entirely and obviously on the phenomenon of indexicality, which most of us understand better than Bargle apparently does.

"Delicious" dialogue

Argle: Yum, this Vegemite is delicious.
Bargle: No it's not.
Argle: It is to me!
Bargle: Well not to me.

This case is broadly similar to the previous, though the indexical element is not initially overt.

Sofa and divan dialogue

Argle: You said they'd given us a room with a sofa, but there's only a divan!
Bargle: But a divan *is* a sofa!
Argle: Nonsense! A sofa has arms and a back.
Bargle: On the contrary, a divan is precisely, a sofa which lacks a back, and possibly also lacks arms. This one has arms though. Don't be fussy!

Argle and Bargle associate different application-conditions with the word 'sofa'. Understood Argle's way, 'the room has a sofa' is false; understood Bargle's way, it's true. They are each right in their own way and there is no real contradiction in their claims.

First football dialogue

Argle: In football, it is permissible to pass the ball by punching it.
Bargle: It is not!
God (thinks):   Poor mortals! Argle is thinking of Australian Football, Bargle of soccer, so they are invoking football-concepts with different application-criteria, hence expressing no genuine disagreement.

Bank dialogue

Argle: After Harry arrived in River City he went down to the bank.
Bargle: Well I was there, picnicking down by the river, and I didn't see him.
Argle: Of course not: I meant the savings bank, not the river bank.

H1

I see all these cases as straightforward failures to disagree. What moral can we draw?

When protagonists make claims having truth conditions that are mutually logically inconsistent, let's say they have a 'basic logical disagreement'. It's easy to think, *real disagreement* requires *basic logical disagreement:* and this seems to fit the above cases. Token utterances superficially appear in each case to be contraries—to have jointly unfulfillable truth conditions—but in the mouths of their utterers, they do not. Either difference of context (including speaker) makes for differing truth-conditions, or the same or superficially similar words are used with different application conditions (we could say, to express different concepts). So no basic logical disagreement; so no real disagreement.

But is it *always* true that where there is no basic logical disagreement, there is no "real" disagreement? My aim is to argue otherwise. At least, I'll argue that there can be cases where there is "a kind of disagreement" despite a lack of basic logical disagreement. (I mean here to suggest the colloquial answer "kind of" to the question 'is there a real disagreement?'. "Merely semantic" debates about proper use of the phrase 'real disagreement' are not what concern me here.)

Let's widen our range of examples. In particular, let's ask whether the following is a "failure to disagree".

Euthanasia dialogue

Argle: Assisted voluntary euthanasia violates my entire concept of what's right, which holds human life sacrosanct. It is always wrong.
Bargle: *Au contraire*, it's often right. According to my concept, what's right is what maximizes well-being and minimizes suffering, so long as certain core moral values are respected. Autonomy is one of them. Coupled with the imperative to minimize suffering, it trumps the so-called sanctity of human life.

Viewed as we viewed the previous examples, this too should be a case of mere argie-bargie. But intuitively it is not!

Of course a great many doctrines on the philosophical market offer to explain this fact. But let's for now attempt to simulate a kind of pretheoretical innocence and see where it leads us if we try to take these statements—*including the references to "concepts"*—at face value (trying as we do to begin also with a presumption in favor of some kind of ethical cognitivism). I will suggest that sometimes ethical disputes really do have a character of the kind reflected in this dialogue, so viewed. There is a discrepancy

H1

between moral concepts deployed by the disputants, so that there is no basic logical disagreement: yet they still count intuitively as a "kind of" disagreement.

I hold *in addition* that in *some* such cases, it is idle to ask which party has, in absolute terms, the correct view, as if this question can be answered from some neutral standpoint, since there is no suitably "neutral" standpoint. Any attempt must import a nonneutral standpoint and hence be essentially question-begging. For want of a better term, I'll stipulatively call these disputes 'irresolvable'; henceforth I'll use the term *quasi-disagreements* for *irresolvable* disputes that involve no basic logical disagreement but which stubbornly resist being viewed as nondisagreements or mere argie-bargie.

3   Moral Functionalism—Whose Theory? Which "Folk"?

Let's now consider a few more details of, and issues for, Jackson's moral functionalism. Moral functionalism is named for its analogy with analytic functionalism in the philosophy of mind. In a nutshell, the latter suggests that users of commonsense or "folk" psychological vocabulary tacitly subscribe to a theory, "folk psychology," seen as implicitly defining terms of that vocabulary in nonmentalistic, commonsense terms. Analytic functionalists hold that were such a theory to be made explicit, purged of inconsistencies, and Ramsified, it would be appropriate to regard the (folk) mental properties as the (joint, near enough) satisfiers of the resulting Ramsey sentence. *Moral* functionalism is similar except that the relevant theory is *folk morality,* seen as implicitly defining moral terms in nonmoral descriptive commonsense terms. Moral properties are thus seen as the best (joint, near-enough) satisfiers of the folk-moral roles specified by the relevant Ramsey sentence.

In support of this approach, Jackson begins by arguing for cognitivism and descriptivism. Indeed he presents his moral functionalism as the working out of what he calls 'analytic descriptivism'. He argues for the latter from what he claims to be the a priori, necessary, global supervenience of the ethical on the descriptive.

Let's look more closely at the question: how, on views like Jackson's, do we build (or contemplate building) a folk theory, and out of what? Appeal to intuitions about actual and possible cases (the "method of cases") is not in principle, nor historically, the only option. One option suggests that we take as raw materials for a folk theory a suitably varied and comprehensive assemblage of commonplaces and truisms involving the relevant

H1

problematic terminology.[5] A well-known problem for such an approach is the problem of "epiphenomenal" truisms—things commonly asserted as truistic by the folk, but which seem to play little practical role in guiding the application of the relevant terms in contexts other than "metalevel" reflexive examination of the terms themselves.[6] This is likely to stem, for instance, from the incorporation of bits of religion, pop philosophy, or pop science into ordinary discourse. For example, given the prevalence of popular dualist views, it is almost inevitable that analytic functionalists who are materialists will need to dismiss some dualistic tendencies in explicit "folk psychological" theorizing on some such grounds.

An alternative is to pay more attention to actual linguistic conceptual practice. When the folk are not asked to make explicit what they believe to be analytically or truistically the case about the problematic domain, but rather observed with an eye to how in practice relevant terms are applied, withheld, and apparently conceptually linked, what theory can reasonably be seen as implicit in that practice?

As described the task is potentially massive, fit for an army of ethno- and psycholinguists. Philosophers have a couple of ways of cutting it down to size. One consists in deploying a philosophical "nose" for those aspects of ordinary usage that are potentially especially revelatory of those aspects of implicit folk theory that bear on the major points of philosophical controversy about the problematic domain. For instance, in debates about analytic functionalism about the mental, little time is spent exploring the functional distinctions if any between anger and rage, boredom and ennui, frustration and disappointment. We know we can let such issues lie where they fall without significantly affecting the central ontological issues. The real task after all is not, typically, to articulate a complete folk theory, so much as to refute philosophical objections to the possibility of doing so in a way that will render it consistent with particular metaphysical claims.

This resource of selective attention is typically incorporated into the "method of cases." Rather than observing speakers' actual usage in real-life circumstances, one describes actual or possible situations—chosen for their capacity to isolate and clarify points of philosophical contention—and elicits *intuitive* judgments about what is or would be the correct thing to say in, or about, those situations. 'Intuitive' judgments here means, roughly, judgments that arise spontaneously as manifestations of a subject's ordinary, tacit linguistic knowledge, rather than being deduced from their explicitly held, amateur or professional philosophical, psychological, or linguistic theories.

H1

The method of cases has the virtue that a philosopher can claim *qua* competent speaker of ordinary language to be as good a representative of what it is appropriate to say about such cases as any other: the relevant "folk theory" is conceived as, near enough, implicit in the shared linguistic competence of normally competent members of a given speech community. But it is not surprising that there is less than complete agreement between philosophers regarding what should be said about actual or possible test cases that bear on live philosophical controversies. The claim that one's intuitions are completely unaffected by one's philosophical views or background is as implausible as the claim that the intuitions of the folk generally are completely unaffected by religion, superstition, pop psychology, and the like, and once one turns from observing actual unselfconscious practice to consulting intuitions under cross-examination, the risk must increase that answers are affected, even if unconsciously, by a theory that is "epiphenomenal" with respect to practice (see Johnston 1987 for a useful discussion of this problem).

Also crucial in what follows is the point that any such method will be hostage to the assumption that, for relevant philosophical purposes, the folk speak as one. If philosophers fail to agree then we can turn to such paradigm masters of unadulterated folk terms and concepts as passengers on the Swanston Street tram or the Bondi bus. But if even those representatives of the folk turn out inconsistent with one another at crucial points, the method of cases will still be in trouble.

There are of course various resources for "explaining away" some such discrepancies. One may hope to argue that there is a common core of agreement that demarcates a shared concept. Discrepant judgments are to be considered "peripheral" rather than "core," and at least some of those who make these judgments will be considered to be in error relative to some truth determined by the core, conceived as relatively a priori, combined with relevant facts. In one sort of case, the core amounts to a canonical definition of a kind of theoretical entity, by way of a definitive functional role it plays relative to the uncontested domain; the periphery of contested claims consists of true or false a posteriori claims about that theoretical category. In another sort of case, the core will be thought of as a set of definitive principles, and the periphery will be thought of as consisting of claims held, in some cases mistakenly, to be deducible from those principles—claims, in other words, resulting from attempts, in some cases mistaken, to apply those general principles to particular cases. (In cases of both these kinds, the idea is that a "correct" account of the peripheral cases in one way or another follows from the core of agreement, together with

H1

further matters, not of opinion, but of fact.)[7] As well, or instead, there is the option of discerning or positing a relativistic or indexical element in the concepts in question: as portrayed in rather stark simplicity in the "Vegemite" dialogue.

Each of these possibilities will be salient at points in what follows. For the moment let us note that in any application of the Canberra Plan, the relevant folk theory must pretty well inevitably be thought of as some cleaned-up, systematized abstraction from an extremely complex host of facts about assorted points of usage and intuition. Ramsification proceeds by way of generating a schema (an "open sentence") from a theory by substituting variables of quantification for relevant theoretical terms in that theory. Respective members of the unique tuple of entities, if there is one, which ("near enough") satisfies that schema, are thus identified as the referents of those theoretical terms. Since no tuple has inconsistent properties, the price of attributing an inconsistent theory to the folk will be arriving at an "error theory" of the problematic domain. Often it will seem more appropriate to remove inconsistencies from the folk theory than to construe the folk as, literally, talking about nothing. For similar reasons, it will sometimes seem more appropriate to drop some claims from the folk theory because they simply clash with well-established empirical findings. Thus a "revisionist" element may be hard to avoid, in the attempt to construct a folk theory that is to have any prospect of identifying the subject matter of folk discourse.[8] At the same time, the point of the exercise will be lost if the folk cannot plausibly be held to be in some sense, however implicitly, *committed to* the relevant folk theory. The latter point, I suggest, should count as an important constraint on Canberra Plan methodology.[9]

4   Can Moral Functionalism Avoid Relativism?

Focusing our attention, then, on the case of ethics, we can hardly avoid noting that moral disagreement is pervasive and robust. Moreover, people often give firmly held moral opinions "bedrock" status, refusing to be argued out of them by appeal to shared common principles (let alone by explicit appeal to some domain of moral theoretical entities such as "moral facts" or "moral properties" conceived as somehow causally or otherwise *bestowing* some or other moral status on the disputed cases). Note too that this is far from a matter of giving mere theoretical lip service to the relevant points of view: such opinions are frequently acted on in the face of considerable obstacles. Taking not the mere fact of moral disagreement, but

these kinds of recalcitrance, into account, different parties can appear, if using moral terms to predicate descriptive properties at all, to be using them to predicate *different* descriptive properties. Moral functionalism, in other words, can threaten to lead to moral relativism.

One kind of tactic to avoid this kind of consequence—one that Jackson invokes—is to conceive of ourselves as Ramsifying a suitable *future* or *counterfactual substitute* for *actual, present* folk morality. Jackson calls this "mature folk morality." He says:

We can think of this story (John Rawls's) as one story about how folk morality should evolve over time: we modify folk morality under the constraint of reconciling the most compelling general principles with particular judgments. In this way we hope to end up with some kind of consensus.

In any case, however we should characterize the way folk morality is evolving over time, it is useful to have a term for where folk morality will end up after it has been exposed to debate and critical reflection (or would end up, should we keep at it consistently and not become extinct too soon). I will call where folk morality will end up, mature folk morality. The idea is that mature folk morality is the best we will do by way of making good sense of the raft of sometimes conflicting intuitions about particular cases and general principles that make up current folk morality. (Jackson 1998a, 133)

This kind of tactic comes in several varieties, and Jackson, in the above passage, is not always clear which he advocates. The simplest version, which he does not advocate, simply assumes that there will be a time when moral debate and critical reflection have run their course. Moral disagreement not due to simple errors of fact or reasoning will be done and gone, and everyone not subject to such errors will subscribe to mature folk morality. The moral properties are those that would be picked out by way of Ramsifying that theory.

Such a view is already one step away from a view that takes people's actual, *current,* moral concepts as a guide to what they are talking about in moral discourse (unless there is *actual current* unanimity among the folk— or evidence in their practice—that the content of actual future folk morality better expresses the content of their current moral concepts than their current moral views). But as Jackson reveals, we must also allow for the possibility that fundamental moral debate will last as long as the human race. For this reason if no other, Jackson must envisage a potentially merely counterfactual "mature folk morality"—what we will *or would* wind up with. Something further emerges in the requirement that we "keep at it *consistently*," and in the mention of Rawls's view as bearing on how folk morality *should* evolve over time: the risk of a "bogus" folk morality being

H1

what actually will or would emerge at the end of fundamental moral debate. If this is a coherent possibility, then, arguably, it is not how folk morality actually will or would, but how it *should* (given the chance) evolve over time, that really matters. And in that case moral functionalism not only needs a substantive account of the constraints on moral debate and reasoning required in principle to distinguish genuine from bogus mature folk morality: it must be reasonable to judge that the folk are sufficiently committed, explicitly or implicitly, to those constraints for it to be reasonable to construe whatever version of mature folk morality those constraints would mandate, as taking priority over their current fundamental moral commitments, in best expressing their actual, current moral concepts.

Before defending the claim that moral functionalism must meet such a challenge, we need to note the fact that Jackson *does* endorse the fall-back possibility of a relativist moral functionalism. It is worth quoting an extensive passage here:

I have spoken as if there will be, at the end of the day, some sort of convergence in moral opinion in the sense that mature folk morality will be a single network of input, output, and internal role clauses accepted by the community as a whole. In this case we can talk simply of mature folk morality without further qualification. Indeed, I take it that it is part of current folk morality that convergence will or would occur. We have some kind of commitment to the idea that moral disagreements can be resolved by sufficient critical reflection—which is why we bother to engage in current folk morality. But this may turn out to be, as a matter of fact, false. Indeed, some hold that we know enough as of now about moral disagreement to know that convergence will (would) not occur. In this case, there will not be a single mature folk morality but rather different mature folk moralities for different groups in the community; and, to the extent that they differ, the adherents of the different mature folk moralities will mean something different by the moral vocabulary. . . . I set this complication aside in what follows. I will assume what I hope and believe is the truth of the matter, namely, that there will (would) be convergence. But if this is a mistake, what I say in what follows should be read as having implicit relativization clauses built into it. The identifications of the ethical properties should all be read as accounts, not of rightness *simpliciter,* but of rightness for this, that, or the other moral community, where what defines a moral community is that it is a group of people who would converge on a single mature folk morality starting from current folk morality. (Jackson 1998a, 137)

Several things need noting in this passage. One is how little attention is paid to the question of what would count as the *right* kind of critical reflection, or the *right* kind of resolution. It is all very well to say that we share some kind of commitment to the idea that moral disagreements can be

H1

resolved by sufficient critical reflection, but that provides no basis for saying our ethical concepts are in essence held in common, if we have or would have differing conceptions of what would count as suitable "critical reflection."

Second, it is to be noted how Jackson slides into speaking of what the adherents of the different mature folk moralities *will* mean. If we wish to give an account of the subject matter of contemporary moral debate and discourse, the issue should not be what *they* mean, but what *we* mean: and every counterfactual antecedent, every distinction between their circumstances and ours, draws a prima facie wedge between the two.

Third, it's worth noting that the more moral questions left unresolved, and the less correlation between people's individual answers to each of them, at the end of moral debate, the smaller and more numerous the "moral communities" will become. If we think of a continuum between a radically particularist view of moral judgments, and an opposed extreme that holds that a very few very general moral rules suffice in principle to settle all moral issues, then Jackson, as an optimist about moral convergence, seems to be judging that the truth—at least as it would be revealed by suitable critical debate—lies nearer the latter end of the spectrum. But the less that is true, the worse, I suggest, his problems get.

Finally, and most importantly for our purposes, Jackson says nothing directly about the *residual* "disagreements" that might remain between distinct "moral communities." But a straightforward implication of what he says is that they will simply amount to "failures to disagree." Members of different communities will mean different things by their ethical terms, critical reflection will have removed all factual errors and mistakes in reasoning, and the residue will be a number of communities within which there is fundamentally perfect ethical agreement, and between whom there is mere "talking past one another."

In short, Jackson *needs* some help from quasi-relativism, if he is not (by my lights) to misrepresent this situation, given that he cannot rule it out. In fact, it's rather worse than that, if we imagine, as Jackson must, somehow bridging the prima facie gap between what such future or hypothetical communities would be talking about, and what we and our contemporaries are currently, actually talking about when we use the moral vocabulary. I now argue (roughly speaking) by dilemma. (There are further options we could consider: but plausibly they are mere intermediate cases that don't improve the available outcomes.)

First option: for each current user of moral vocabulary there is a determinate "moral community" they would, under suitable counterfactual

H1

assumptions, belong to such that the meanings of terms in their moral vocabulary should be seen as the same as the meanings of terms in the moral vocabulary of that moral community. *Bad consequence:* there are two types of apparent moral disagreement. There are moral disagreements between people affiliated with the same (hypothetical) moral communities, and these are genuine disagreements—disagreements, so to speak, that reflect disagreement or ignorance about which judgment the relevant moral community would make. Other apparent moral disagreements are really nondisagreements, since the disputants are affiliated with distinct moral communities; hence, meaning what members of those communities would mean, they are not genuinely in disagreement, merely at cross-purposes. *Worse consequence:* since none of us knows what comprises the full set of (future or counterfactual) moral communities, let alone which of us is affiliated with which community, none of us knows in the case of any particular apparent moral disagreement, whether it is by these lights real or bogus.

Second option: typically, we contemporary users of moral vocabulary are not, as individuals, determinately associated with particular future moral communities. In particular, there is no entailment from what members of any particular one of those moral communities will or would mean by terms in their ethical vocabulary, to what we mean by those terms in our vocabulary. *Bad consequence:* We either mean something entirely different, determined by our actual current moral judgments—vitiating the role of mature folk morality—or at best we mean something vague that is indeterminate with respect to matters over which those different communities differ. The latter alternative is unattractive precisely because the points over which those hypothetical moral communities differ are likely to include matters over which contemporary users of the ethical vocabulary are in fundamental, bedrock disagreement. If such bedrock disagreements suffice to make it true that the hypothetical communities differ in what they mean, it seems odd that they should not have similar consequences for those involved in contemporary moral debates.

Jackson does not really expect that this relativist outcome will or would come to pass. But if it merits even a modicum of credence, his view seems to leave a lot to be desired so far as addressing the topic of moral disagreement. Jackson's proposals raise a panoply of highly debatable questions about what the outcome of critical debate must be, how it may be arrived at, and what connection it has, exactly, with the semantic knowledge and intentions of contemporary folk. And if we take current speakers' actual moral judgments (as displayed both in speech and in action) as revelatory

H1

of their ethical concepts, we are in danger of having to describe as non-disagreements disputes that stubbornly resist such a characterization.

Interestingly Jackson makes one remark that seems to run counter to what ought to be the underlying assumption of his proposal of mature folk morality, namely, that there is a bridge between our current moral talk and the meanings of moral terms as they would figure in mature folk morality, such that the latter determine the common subject matter of our contemporary moral debates. He says:

> 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*. (Jackson 1998a, 131–132)

Apart from the optimistic focus on what happens when moral debates *do* get settled, Jackson here goes directly against the idea that the essential common ground in our debates is a commitment in principle to the deliverances of *mature* folk morality, whatever they may be. For he insists that it is *currently endorsed* but *mutable* tenets that provide the common ground that is essential if there are to be sensible debates. But in the light of these comments it is hard to see how mature folk morality can lay claim to settle the *subject matter* of our current debates. As for the "general principles *broadly stated*," insofar as there is robust disagreement about particular cases, it has to be doubted whether those who subscribe to such principles thereby interpret and apply them in such a way as to rule out the possibility that they are applying moral concepts that differ importantly in their application to those particular cases.[10]

5 Basic Logical Disagreement Despite Differing Application-Conditions?

Before going any further we need to dispose of an important and pressing objection to the whole way I have laid out the issues so far. Here's how it goes. In the above cases, the failure of basic logical disagreement, which makes the claim that there is "real" disagreement problematic, stems from differences in extension between words or concepts employed by the disputants, due to differing application-conditions for those words or

H1

concepts. But often the extensions of terms or concepts are *not* best seen as settled by associated application conditions, and then such situations don't arise. If that's how it is with all putative ethical quasi-disagreements, then there is no call for a quasi-relativist position. Let's look at some cases of that kind.

Water dialogue

Bargle: Surely it's a conceptual truth that water repels fire? Where there's water, there's no fire: isn't that why we put out fires with water?
Argle: Your water-concept is not merely mistaken, it's positively mediae-val! Don't you know that a burning reel of nitrate film will continue to burn, even underwater, since nitrate film generates oxygen as it burns? Magnesium also burns underwater!

5.1   Anchoring and Bypassing

Here Bargle's concept is *bypassed*: he is talking not of some mythical sub-stance that repels fire, but of the natural kind called 'water'. Despite the difference between current and previous theories of water, we *can* refer to the same stuff as those who held those contrary earlier theories, because of the existence of that natural kind to which, we reasonably say, it was always intended to refer. The kind *anchors* uses of the term, enabling com-monality of reference even when people's water-concepts are idiosyncratic, mistaken, or out of date. This is just like the familiar case of theory change with persistence of reference, as in evolving theories of the electron. Bar-gle's idiosyncratic concept is irrelevant and there is no threat of failing to disagree. Notice here that the circumstance enabling the concept to be bypassed also blocks the "no-fault" condition.

But what if water-as-natural-kind itself turned out mythical, with no real natural kind anchoring uses of the term? Insisting the term must refer to a natural kind or nothing would leave debates about water and water-con-cepts resembling debates about unicorns and unicorn-concepts. Quite likely we would revise our ideas about the semantics of 'water', and go on using the term for various substances more-or-less occupying something resembling the traditional water-role. But that would make it *vague*, so that people might sometimes legitimately say 'you may call this water, but I don't', just as they sometimes say 'you may call that color green, but I don't'. There are preconditions for "anchoring" a term without which a default assumption or even a stipulation, that all within the speech com-munity shall use that term with common reference, will be undermined.

H1

5.2   Social Kinds and Anchoring, Perfect and Imperfect

Not that a term has to be or even resemble a natural kind term to be reasonably firmly "anchored" and so "bypass" any associated concept. Consider:

Second football example

Argle: In football, it is permissible to pass the ball by punching it.
Bargle: It is not!
Argle: Oh, you must be thinking of soccer! I'm talking about Aussie rules.
Bargle: And so am I! I've known the game since childhood and my concept of it is very clear. Passing the ball by punching it is not permitted.

We must not conflate dispute over concepts with dispute over things those concepts apply to. This case involves genuine disagreement over concepts of Australian Football: but is it also a genuine dispute about games the rival concepts apply to?

The question is ambiguous. The concepts invoked have different application conditions, hence different extensions ('it's permissible to pass the ball by punching' being a necessary condition for a game's being in the extension of Argle's concept, but not Bargle's). In *this* sense of 'apply' there is *no* misapplication of concepts; thus we might say each is right in his own terms, and this is another case of failure to disagree.

But that would be *wrong.* We should more defensibly say Bargle is wrong and his concept (given his words) is *mistaken.* This is because we take Argle and Bargle to succeed in commonly referring to and disagreeing over a single well-defined football code, about which Bargle's claim is false. Bargle uses the words 'Aussie Rules', in a context which by default licenses assigning to the parties' words whatever reference they *normally* have in the relevant speech community. This gives another sense (though perhaps a stretched one) of 'what the concepts apply to', allowing a concept to be *mis*applied. Because an idiosyncratic concept like Bargle's can be "bypassed," Bargle achieves reference despite it, to the same code (and games) of football as Argle. Hence we have no quasi-disagreement here (in the sense of 'quasi-disagreement' introduced earlier).[11]

However, we must remember "Wittgenstein's lemma": *not all games are like Aussie rules.* Many games are not so clearly codified. More generally, there are many social kinds, games among them, such that our names for them are vague in respect of many details one might expect to be

H1

determined a priori by the relevant concepts: and thereby, disputes may arise. Think for instance of 'common law marriage', and 'gay marriage'. Are these "real" marriages? Opinions differ, and debates about the matter often take the form of debates about what 'marriage' really means.

Cultural variations in marriage customs complicate the semantic issues. When English-speakers speak of 'marriage', they don't speak of it only as it exists in their own English-speaking community. A colleague recently witnessed in Northern China a wedding ceremony in which two brothers married a single woman. In the preceding (true) sentence I apply an English-speaker's concept of marriage to a form of that institution that is not part of a typical English-speaking community, nor of any of the atypical English-speaking communities to which I belong.

In any case, there are many terms that allow limited variation in reference, tracking variation in concept. 'Sofa' appears to be one. Usage varies, so argie-bargie easily ensues—and in that case, as in many others, is almost as easily resolved.

The distinction between natural-kind terms and others is sharpish, but nonnatural, especially social, kinds form a continuum, from cases like 'Australian Football', via 'hopscotch' and 'sofa', to cases like 'marriage' or 'chief'. The further along this spectrum, the less "anchoring" is available to stabilize reference across time and speech-community, so the more room and need for individuals' concepts to play a complementary, disambiguating role in reference determination.

Thinking about translation helps to illustrate these issues. Consider the following dialogue:

Anthropologists' dialogue

Anthropo-Argle: The two most important figures in the socio-politico-religious hierarchy of the Sregnablay people are the namriahc and the reganam. For our purposes we can best translate those terms as 'king' and 'prime minister', *respectively*.
Anthropo-Bargle: All wrong as usual! The one you call 'king' is the high priest, and the one you call 'prime minister' is the king.

Here either choice of labels might be permissible, but one choice might fit Argle's usage better, the other Bargle's. Though Anglophones will agree on many paradigms, there may be insufficient anchoring of the terms for either disputant to be clearly misusing them in a case such as this. When we go from culture to culture trying to identify moral concepts, we are similarly likely to find ourselves at times inevitably wavering. Is this a very strange morality we encounter here? Or a comparably strange

H1

religion? Or an odd political tradition? Or some bizarrely overblown commitment to notions of "honor," or of "etiquette"? However clear the answers may be in some cases, in others there may simply be no clear answer.

I've suggested that for some terms, reference is "anchored" for a speech community, "bypassing" concepts individuals take as analytically expressive of their meaning: mechanisms for this can implement the default convention that terms shall be used with a common reference within that community. Sufficiently robust anchoring can block "no-fault" disagreements of the kind we're considering. For other terms, prerequisites for such "anchoring" mechanisms are missing or incomplete. Then an individual's concept cannot be wholly "bypassed," since it is required to substitute for, or complement, those mechanisms: these cases allow for "no-fault" disagreements.[12]

6 Contestable Normative Notions in Jackson's Account—"Critical Reflection" and "Rationality"

We can now make a further comment on the problem Jackson faces in "bridging" the prima facie gap between current and hypothetical future moral concepts, seen as implicit in current and "mature" folk moralities. The Water dialogue and the Second football dialogue remind us that (in old-fashioned terminology) differences or changes in concepts or "sense" need not entail discrepancies or variations in reference or extension. But this is possible only to the extent that there exist determinate relations of natural-kind similarity and difference—or clear and canonical social conventions—capable of trumping or replacing concepts in determining or "anchoring" reference or extension, in a way that keeps reference or extension constant across the relevant differences or changes. Jackson's extensive contributions to the theory of reference notwithstanding, I claim that we do not have (and cannot find in *From Metaphysics to Ethics*) what would be needed to tell us that moral terms in contemporary use, and their descendants that would figure in mature folk morality, are "anchored" in such a way as to refer respectively to the same moral properties. There are various remarks about our shared implicit commitment to accepting the results of critical reflection, coupled with descriptions of mature folk morality as the future or hypothetical product of critical reflection, but in the light of the foregoing discussion, it seems to me that these simply do not establish the required commonality of anchoring. Jackson himself in fact says: "critical reflection is, by definition, what any theory should be

H1

subjected to. Of course, precisely what critical reflection on current folk morality comes to in detail is a matter of considerable debate" (Jackson 1998a, 139). He says also: "What is a priori according to moral functionalism is not that rightness is such-and-such a descriptive property, but rather that A is right if and only if A has whatever property it is that plays the rightness role in mature folk morality, and it is an *a posteriori* matter what that property is" (ibid., 150–151).

But it is not clearly a priori that this commitment to the canonical status of mature folk morality can be given the requisite degree of determinacy while itself surviving critical reflection, given the possibility of debate about the proper kinds, methods, and circumstances of critical reflection, about which individuals, and in what psychological states, are appropriate participants, and so on, and consequent vagueness in the idea of mature folk morality. Nor is it clear that it is a commitment capable of trumping current other commitments, in giving an account of what moral terms *currently* mean. Compare: I may preface a book with the words 'I am sure some claims in this book are false'. In no way does this abnegate my current commitment to any individual claim in the book. I cannot get "off the hook" for false opinion by saying 'it was always my intention to tell the truth'.

There are two strands of discussion that raise somewhat similar issues in Jackson's ethics chapters—chapters 5 and 6—in *From Metaphysics to Ethics*. In principle they are distinct but at the end of the day they are closely interrelated, and for me, similar doubts arise about them. First—mostly in chapter 5, "The Location Problem for Ethics"—there is a defense of descriptivism and the promotion of "moral functionalism" as the right way to implement it. Jackson stresses the "ecumenical nature" of moral functionalism, and mentions a number of different lines along which mature folk morality might develop, maintaining on the whole an agnostic stance as to which of those possible outcomes will, would, or should, come to pass. My critical comments so far have essentially been directed to the generic account there given of "mature folk morality," and the role assigned it.

But in chapter 6, "Analytical Descriptivism," and especially in the closing sections of the chapter, where Jackson deals with the Open Question argument, and the connection between moral judgment and motivation, he needs to put a few more of his own cards on the table, and express some opinions about some of the content that mature folk morality will have. (Actually, some of these commitments already surface in chapter 5.) He continues to allow for a range of options, but they are all variations on a theme. The theme is that a correct account of the subject matter of moral

H1

judgment must be a "response-dependent" one. This notion is first mentioned as an attractive option, early in chapter 5: "Thesis (S) [the global supervenience of the ethical on the descriptive] is compatible with the idea that ethical nature . . . is in part determined by facts about our responses and attitudes, with the appealing idea that, in Mark Johnston's terminology, value is response-dependent" (Jackson 1998a, 120).

There are further remarks in chapter 5, which reveal Jackson's approval for this option, but at this point let us note an important passage from chapter 6. He says:

Moral functionalism sees the meanings of the moral terms as given by their place in a network. Part of that network is certain output clauses that tell us how beliefs about ethical properties connect with facts about motivation. The details of these output clauses are highly controversial. But, to fix the discussion, let's suppose that the connection with motivation goes roughly as follows . . . it seems to me to be on essentially right lines . . . (a) right act is one that has properties of value to an extent that exceeds that of the various alternatives to it, and a property's value depends on its being rational for us to desire it. The *moral* rightness of action is, then, a matter of its having properties of value . . . that pertain to morality. . . . It is important to this sort of proposal that we can give an account of what makes it *rational* to desire a property that does not reduce, uselessly, to its being a rightness-making property. (Ibid., 156)

After listing some recent proposals for identifying properties it is rational to desire, Jackson continues:

The details will not matter for what follows. What will matter, though, is something I take to be widely agreed, perhaps under the heading of the rejection of Platonism about value, perhaps under the heading of the response-dependence of value. It is that what confers value on a property ultimately comes down to facts about desires. . . . Accordingly, this much is right in subjectivism about value: what gives value whatever objectivity it has comes down, somehow or other, to some combination of facts about the convergence, the stability, the coherence between first-order and higher-order desires, the desires of idealizations of ourselves, the desires of our community, and the like. (Ibid., 157)

A central tenet of my own view is that there will always be the possibility of bedrock disagreements about what counts in this sort of context as *rational.* This is inevitable given persisting bedrock moral disagreements, one ethicist's modus ponens being another's modus tollens. This is why I claim that there may be moral quasi-disagreements in my strong sense: *irresolvable* disputes in which there is no uncontroversially "neutral" standard by which at least one party may be judged to be mistaken. We have no better standard than ideally rational judgment under ideal circum-

H1

stances: but the notion of "ideally rational" judgment lacks sufficiently determinate anchoring to eliminate bedrock disagreements, in a similar way, and for similar reasons, to moral concepts themselves. This is a main reason for preferring ethical quasi-relativism. And in his rejection of Platonism, his critique of "Cornell realism," and his other doctrines, Jackson makes clear that he has cut himself adrift from other potential dispute-settling anchors for moral concepts.[13]

These two strands of Jackson's argument—the general idea of mature folk morality as implicitly specifying what the moral properties are of which we speak, and the particular adherence to a response-dependent account of value, revolving around the idea of rationality in desire—are in principle distinct. But in practice it appears to me that these operate as two sides of a single coin, in the view Jackson is outlining, and some similar criticisms apply to both. What counts as ideal "rationality" in desire, and its import for the identification of moral properties, and what counts as the right kind and context of critical reflection, and the general role of the mature folk morality it will, would, or should arrive at, are all matters that must be hammered out in arriving at mature folk morality, and the determination of these different matters will be part of a single process. Even setting aside the "bridging" issue, these interwoven processes cannot do their jobs unless subject to appropriate norms, and the content of those norms must at the end of the day itself be a matter for quasi-disagreement, not subject to resolution from any uncontestedly "neutral" standpoint.

7   Words versus Concepts

Thus my view is, roughly, that for words like 'right' and 'good' (in the moral sense)—and I could add, 'person'—we can say that "anchoring" to *some* degree *fails,* leaving room and need for individuals' concepts to play a role in reference determination, and no impartial way to adjudicate between speakers so as to secure a perfect match. I say 'roughly' partly because this formulation suggests that the view is more strongly tied than I intend to claims about particular linguistic expressions. So far I've mentioned the application conditions of words, and of concepts, more or less as if these are interchangeable ways of speaking. Now I need to put in a correction. Compare Jackson:

I use the word 'concept' partly . . . to emphasize that though our subject is the elucidation of the various situations covered by bits of language according to one or another language-user, or by the folk in general, it is divorced from considerations local to any particular language. When we ask English users in English for their

H1

intuitive responses to whether certain cases are or are not cases of knowledge, we get information (fallible information . . .) about the cases they do and do not count as covered by the English word 'knowledge'. But our focus is on getting clear about the cases covered rather than on what does the covering, the word *per se*. We mark this by talking of conceptual analysis rather than word or sentence analysis. (Jackson 1998a, 33–34)

This is an unsatisfactory and ambiguous passage. Just what does "getting clear about the cases covered" amount to? We might think of conceptual analysis as combining a part where suitably varied possible situations are divided into those ordinary English speakers do and those they don't count as (in this example) "knowledge"—here an English word is definitely central to the inquiry—and another part attempting to discern principles or patterns governing the resulting division.[14] The latter part makes no reference to language as such, but that hardly "divorces" the whole process from considerations local to English, whatever we say "our focus" is on. Perhaps no other language uses a word that divides cases in just the way 'knowledge' does? Epistemologists might, after all, investigate truth, justification, belief, reliability of belief-generating processes, flukes, hidden defeaters, and all the rest of it, without logically compounding them in the precise manner associated with that English word: there is no a priori guarantee that any language but English does that, and the more convoluted the twists and turns of Gettierology, the more likely it may seem to be an enterprise entirely local to English.

The issue of "anchoring" is with us again. If conceptual analysis of an English word like 'water' points to its naming a natural kind, we can happily turn the job of further investigation over to the chemists, thereby pushing away the ladder of English language, and considerations "local" to it. But when we speak of abstractions—knowledge, virtue, causality, probability—we cannot anchor our conversation by planting the relevant thing on the table or decanting the relevant stuff into a test tube, saying *'This* is what we're talking about'. We must, rather, use some language or other to identify our subject matter. Whether this makes our inquiry permanently parochial to that language is a question to be answered on a case by case basis. In principle, the answer should in part depend on how translation into other tongues fares.[15] We may have some shrewd hunches based on our knowledge of human nature and other cultures, but a hunch is a starting point for inquiry, nothing more.

Consider another of Jackson's remarks: "I said, following Humpty Dumpty, that we can mean what we like by our words. But if we want to speak to the concerns of our fellows, we had better mean by our words what they mean.

H1

If we are interested in which property the word 'right' in the mouths of the folk picks out, we had better give a central place to folk opinion on the subject" (Jackson 1998a, 129). Ethical quasi-relativism takes seriously the possibility that there is no one such property as *the* property that is picked out by 'right' in the mouths of the folk. But regardless, if we wish our inquiry into folk opinion on ethical matters to transcend a local concern with English, we will need to arm ourselves with something more than a resolve not to gratuitously deviate, Humpty-Dumpty-wise, from ordinary usage of the English word 'right'. Whichever words we use, if we are to find a way to communicate with the polyglot folk about ethical matters, we need to find or create linguistic means, simple or complex, for discussing ethical concepts with them. Roughly speaking, we need to come up with a suitable *translation* of the word 'right' into *their* language. So we need, roughly, to figure out how they express a concept the same as (or sufficiently similar to) the one we are expressing: then—but perhaps not till then—we may have some confidence of investigating considerations "not local to any particular language."

(As Quine has taught us, issues about translation usually have analogues within a single language community. Mark Twain well understood this. When Huck Finn said "I got aboard the raft, feeling bad and low, because I knowed very well I had done wrong, and I see it warn't no use for me to try to learn to do right," he was in many ways giving a central place to folk opinion on the application conditions of 'right' and 'wrong': but we know that he did what he felt was right, though neither he nor members of his linguistic community called it that. Thus his word 'right' did not express his concept of what was right. It's important to note that, and how, we can know this.)

For all the vagueness of concept-talk, then, I think we do better to accept a practical need to talk of concepts—and to find good methods of analyzing and individuating them—in ways that do not easily reduce to a coy way of talking about particular words and word-meanings.

Generally, given due allowance for two-dimensionality, semantic externalism, and such, I take it to be perfectly appropriate to talk of 'the concept expressed by the English (or French, Chinese, or Etruscan) word *w*'. But only a subset of concepts corresponds neatly to word meanings, and many important ones do not: I include 'identity', 'person', and 'right' among them. The words I just used are English words (or philosophers' variants), but those concepts could in fact figure centrally in the lives of people whose language had no such words, and I maintain that analyzing them need not be best understood as doing linguistic analysis of the words associated with them.[16] People's person-concepts, for instance, are mostly *not*

H1

deployed by using *words* like 'person'. It would be as misguided to take analysis of that word's meaning as a guide to the structure of those concepts, as to question Leibniz's law by citing usage of the word 'identical'.

8 Modeling Quasi-disagreement

How can the quasi-relativist best represent the "kind of" disagreement involved in "quasi-disagreements"? For this I borrow from Michael Dummett's work on concepts.

Dummett is well known for suggesting that logical constants may be characterized either by natural deduction introduction rules, or elimination rules, and that these rules should be subject to a mutual "harmony" constraint, so that adding a logical constant to a theory results only in a conservative extension of that theory (e.g., in Dummett 1991, 210–218). Dummett also, in his discussions of truth and assertion, draws parallels with the concept 'a win in chess'. One ignorant of competitive games would need to learn that the goal is winning, not merely which positions count as "wins" (and for which player), in order to acquire this concept. Dummett (1981, 296–297) makes analogous claims about the difference between merely knowing sentences' truth conditions, and knowing how to make assertions using them. These are all instances of a general distinction he draws between *conditions for* a concept to apply, and the *consequences* of applying it: he holds that *each of* these distinct aspects is to be considered *constitutive of* a given concept. This allows a richer and more nuanced account of aspects of concept-individuation than simple application conditions accounts can provide. For present purposes, the consequences to be taken into account may be in our theories, in our discourse generally, or be *practical* consequences in our lives, including specially, our evaluative and affective lives.

We could see this distinction as being between, broadly speaking, two (*very* different) aspects of 'content'. It may not be applicable to every concept, but it seems to me useful for making sense of quasi-disagreement. Let's stipulatively label the set of those consequences of applying a concept *that are partly constitutive (or individuative) of it,* the *C-role* of that concept. This is to distinguish the "role" in question, both from the concept's application conditions (which could also be called part of its "role"), and from the *functional roles* specified by (for instance) folk psychology and folk morality. I emphasize yet again: we are *not* talking merely of *inferential* consequences, but of consequences in general, including *practical* consequences—attitudes, emotions, and actions.

H1

The point is not that consequences that are constitutive or individuative of a concept should *always* follow from its application: but if those consequences *never* followed, for an individual or a population, that individual or population could not be said to possess or apply that concept. (This mirrors the thought that an extreme psychopath might be said to lack the concepts of right and wrong, even if he or she could apply the *words* 'right' and 'wrong' in a manner that exactly mimicked the usage of someone highly moral. We would not on the other hand be inclined to describe people as temporarily ceasing to possess those concepts every time they fall into episodes of apathetic or narcissistic amorality.)

The idea is that we can view some concepts as having some kind of commonality—as suitably "the same"—*despite* having different application conditions. Sometimes, I think, we in effect use 'same concept' so as to give priority to application conditions, sometimes so as to give priority to a concept's C-role. Generally (though it's context dependent) the more evaluative, and especially the more normative, a concept, the more we prioritize C-role in concept individuation; the more descriptive, the more we prioritize application conditions. But the general idea is that in quasi-disagreements, the relevant concepts differ in their application conditions, but are competing to play the same C-role. The difference in application conditions accounts for the lack of basic logical disagreement, but competition for the same C-role—the clash of competing criteria for the same consequences—accounts for there being, nonetheless, a kind of "real" disagreement. In cases that interest us this may be not theoretical (or "basic logical") disagreement but *practical* disagreement, in a social context: this is why there often *needs* to be competition rather than peaceful coexistence between moral concepts.

Commonality of C-role can also, of course, account for other aspects of apparently successful communication between speakers who use the same word to express concepts with different application conditions. Indeed, it provides an extra resource for dealing with the difficulty I raised for Jackson, of establishing a plausible bridge between current actual and hypothetical "mature" folk morality, capable of underwriting the claim that in spite of the differences, in some important respect terms in those two theories "mean the same." Not that this provides any solution Jackson can use, since it in no way establishes that the same properties are referred to by the relevant terms of the two theories.

Likewise, we are now in a better position to understand the enterprise of translating the moral language of another culture with whom we are in moral disagreement. Difference in application conditions need not obstruct

H1

us in translating a word of theirs into English 'right', if it is clear that it plays something like the same C-role for them that 'right' plays for us. Not that identity of C-role *alone* should suffice to identify a word of theirs as to be translated by our 'right'. Too large a gulf between the application conditions for their word, and for ours, would rule out such a translation, just as would a gulf between the respective C-roles. Still, approximate commonality of C-role should carry substantial weight in identifying suitable translations. Conversely, a good way to get a grip on what I am here calling the relevant "C-role" is by reflecting on what we would give weight to in such a translation situation.

It would be misleading, however tempting, to say that arming ourselves with these borrowings from Dummett will enable us to individuate concepts with increased precision. All this talk of "concepts," of extracting a coherent folk morality from the intuitive judgments of the folk, and so forth, concerns quite heroic abstractions from an extraordinarily complex host of events involving individual utterances, actions, and contexts (presumably each day on Earth brings with it billions of events of relevant kinds, though each of us encounters only a few of them). And compared to application conditions, the notion of a C-role is a relatively blunt instrument. But it is I think essential to understanding quasi-disagreements and related issues.

8.1 C-Role versus Functional Role

We might wonder how C-roles for moral concepts relate to the *output clauses* of folk morality, and the functional roles they specify. In fact the application conditions/consequences distinction for concepts, though analogous to the distinction between those aspects of functional role specified by input and output clauses, is importantly different from it. Let's call a functional role specified by Ramsifying folk morality an FM-role, and let's call the part of that role specified by output clauses an FMO-role. Here are some observations.

(1) An FMO-role, like the complete FM-role of which it is a part, is a role for a *descriptive property*, whereas a C-role is a role for a *concept*.

We could in addition say: *FMO-role-specifying output-clauses* specify certain (noncausal)[17] consequences of *things instantiating the properties* of goodness, rightness, and so on. An illustrative example from Jackson: properties that make something good are "*the properties we typically have some sort of pro-attitude towards*" (Jackson 1998a, 131; italics added). (This is an *output* clause because it licenses an inference from a claim explicitly about moral properties to a claim not explicitly about them.) As I read Jackson here, it

H1

is the generality of instances of the relevant properties toward which we typically have some sort of pro-attitude, not some especially local subset of those instances.

C-roles, however, selectively relate individuals in possession of the relevant concepts, and their applications of those concepts, to particular cognitive, affective and behavioral events, in the lives of those very individuals: there is an inherent reflexivity in the notion. (In addition, pending a contrary ruling from some more substantive theory of concepts, I am taking the consequences in question to be *causal* consequences.)

(2)   There's a reason Jackson's account could easily lead to confusion over these issues (and which makes what I said under point (1) above something of an oversimplification). If we were to Ramsify physical theory en route to clarifying the semantics and ontology of physics, we wouldn't expect the output clauses dealing with 'electron' to state relationships between electron-*beliefs* and other phenomena: we would expect them to state relationships between *electrons* and other phenomena. Yet when we look at what Jackson says about "output clauses" for folk morality, virtually without exception, he mentions relationships between moral *judgments* and other phenomena, rather than relationships between moral *properties* and other phenomena. The example quoted in point (1) above is the only exception. But even that example comes from the following passage (which, notably, comes fairly early in the general presentation of moral functionalism in chapter 5): "The output clauses of folk morality take us from ethical judgments to facts about motivation and thus behavior: the judgment that an act is right is normally accompanied by at least some desire to perform the act in question; the realization that an act would be dishonest typically dissuades an agent from performing it; properties that make something good are the properties we typically have some kind of pro-attitude towards, and so on" (Jackson 1998a, 131).

Since these output clauses concern not ethical facts but ethical *judgments,* one could almost take Jackson here to be talking of folk *psychology* rather than folk *morality*! There has to be a prima facie worry here that Jackson's version of moral functionalism has gone astray. The idea is that the moral properties are in principle identifiable as those that (as Jackson writes) 'will/would' play a certain complex set of interrelated (noncausal) roles. How do the (causal) roles played by ethical *judgments* help us to identify ethical *properties?* Judging that someone is dangerous will normally lead me to treat him with caution, *whether or not* he is *actually* dangerous: similarly, wouldn't my judgment that something is right normally lead me to be inclined to do it, *whether or not* it is actually right?

H1

In fact Jackson here is violating a constraint Lewis lays down in "How to Define Theoretical Terms": "We must assume that all T-terms in the postulate of T are purely referential, open to existential generalization and to substitution by Leibniz's Law" (Lewis 1970, 429). Fully to explore this issue might require an investigation of the notion of de re desire (or de re "pro-attitudes") including such uneasy questions as whether and how one can have de re attitudes to properties as such.

This is not the place to explore such technicalities. What should be noted, though, is that these passages actually reflect Jackson's commitment to a response-dependent account of moral facts, noted earlier.[18] It is typical of response-dependent accounts to underwrite unusually tight connections between judgments and what they concern. But there is a complex array of considerations lurking here to do with the ontology and epistemology of "response-dependent properties." For starters, the notion of response-dependent *properties* is much more problematic than the notion of response-dependent *concepts*. Arguably the relation of concepts to properties is a many–one relation; and arguably response-dependent properties are not so inherently, but merely *qua* how suitable individuals respond to them under suitable conditions. In any case, response-dependent concepts will track determinate properties only to the extent that the relevant conditions are determinate. There is a great variety in the kinds of conditions that may be relevant—some response-dependent concepts relate to responses under conditions conceived of as "normal," some relate to responses conceived of as "ideal," and either way, the nature and state of the respondents as well as of their environment at the time of responding will typically need to be included among the relevant "conditions." Thus the epistemology of response-dependent properties is inextricably interwoven with the epistemology of relevant (in this case, "ideal") conditions.[19] We may desire to say that a proper specification of those conditions should be geared to the requirement that the relevant responses should "track" the properties we are interested in, but, as is well known, it is all too easy to fall into circularity in attempting to meet this demand. In the case of Jackson's approach to moral functionalism, the "ideality" of the relevant conditions incorporates the notion of what would be desired or valued were one ideally rational, so the circularity that (as Jackson says) must be avoided is the one we would generate by defining ideal rationality as what we exhibit when our desires track genuine values.

Thus there is a complex web of issues to be untangled in evaluating the acceptability of Jackson's inclusion among moral functionalism's "output clauses," of clauses relating moral judgments, rather than moral properties,

H1

to motivation. Though we cannot deal with them all here, the issues, and the relevant passages, are important: for one thing, they reveal how deeply Jackson has woven the idea of a response-dependent account's being endorsed by mature folk morality, into his whole account of moral functionalism. Once again, I am skeptical on at least two counts. One is skepticism about whether "mature folk morality" can be expected to find an agreed-on, noncircular, and determinate account of "ideal rationality." Another, at least as central, is whether a reasonable account of the primary conceptual commitments of the folk can permit the content of some future or counterfactual best-case scenario for the outcome of critical reflection and debate—one in which the subtleties of response-dependence have all been successfully negotiated and incorporated—to be taken as trumping their salient, current, and firmly held (albeit philosophically less sophisticated) judgments about what is right or wrong, good or bad, in determining the content of those current judgments. What deepens the latter concern is the observation that just as on this account the folk's actual moral *judgments* take second place, in delineating their *actual* moral concepts, to their *merely counterfactual* judgments, so too their actual moral *motivations* are similarly trumped by their *merely counterfactual moral motivations*.

That said, these elements in Jackson's position do also amount to some kind of lessening of the prima face gap between my view and Jackson's. For instance, 'the judgment that an act is right' is not much different from 'an application of the concept 'right''. But this should not be exaggerated. Jackson is saying that it is implicitly a priori in mature folk morality that there is (or will or would be) a connection of this kind between moral judgments and various pro-attitudes. This is to say that the existence of such connections is criterial for the *applicability* of moral concepts: it is *not* to say that *possession* of moral concepts is *responsible* for such connections. It is not apparent in what sense purely descriptive concepts could carry such responsibility.

(3) Jackson perhaps comes closest to what I say about C-roles—including the aspect just mentioned—in his discussion of the "Open Question" argument against ethical descriptivism, in particular where he moves from discussing the "content" strategy for explaining "the directed nature of the belief that A is right, within a purely descriptive framework," to discussing what he calls the "content-possession" strategy for "adding motivation to direction." Jackson says, inter alia: "on the moral functionalist story, to believe that something is right is to believe in part that it is what we would

in ideal circumstances desire. . . . Now this fact will typically manifest itself in our feeling to some degree the 'tug' of A" (Jackson 1998a, 159). ('A' is something believed right.)

Jackson continues in a postscript, considering a protest from "noncognitivists," which is similar to (though stronger than) my own insistence on the importance of C-role. (I consider my own position, incidentally, a hybrid of cognitivist and noncognitivist elements, rather than straightforwardly cognitivist or noncognitivist). The protest "insists that when [I take this to mean 'whenever'] one judges, *really* judges, that A is right, one must have a current, first-order pro-attitude towards A" (Jackson 1998a, 160), to which Jackson replies: "we can accommodate this view by refusing to *call* something a *moral* belief unless it is accompanied by the relevant pro-attitude" (ibid., 161).

This is one of the passages that lead me to say that "some may feel that little more than matters of emphasis separate us." But there is still a significant gap. Note the difference between saying 'a concept-tokening is not *moral* unless linked appropriately to motivation of its possessor' (which is closer to what Jackson says) and 'a moral concept is not *possessed* unless it is linked appropriately to motivation (etc.) of the possessor' (which is what I say).[20]

8.2 "Harmony" for Ethical Concepts?
Returning to ethical quasi-relativism. How, it may be asked, do application conditions and C-role mutually constrain one another? Even in the case of a logical constant, what it takes for the analogues of these aspects (introduction rules and elimination rules) to dovetail suitably is debated—but at least the question lends itself to precise answers (see Read 2000). But that kind of "harmony" is very different from anything that could sensibly be proposed in ethical cases.

At this point I can do no better than quote Dummett himself. He is talking about personal identity, but what he says is mutatis mutandis pretty much applicable to ethics also, in my view. He says:

not to say that the character of the harmony demanded is always easy to explain . . . the most difficult case is probably the vexed problem of personal identity. An assertion of personal identity has consequences both for responsibility for past events and motives in regard to future ones. We can imagine people who employ different criteria for personal identity, but attach the same consequences to its ascription: what is difficult is to say where their mistake lies. (Dummett 1981, 455n)

H1

And

> If there were no connection between truth-grounds and consequences, then the disagreement between us would lie merely in a preference for different concepts, and there would be no right or wrong in the matter at all. Because we feel sure that there must be some connection, but we cannot give an account of it or say how tight it is, we are pulled in both directions: we feel that the disagreement is not a straightforward factual one, but that, at the same time, it is not merely a matter of applying different concepts. (Ibid., 358)

These are just the kinds of considerations that can lead one to a quasi-relativist view. If we *could* establish an objective, neutral, canonical, and precise criterion for an analogue of "harmony" between application conditions, and consequences (i.e., "C-role"), for moral concepts, then of course it would provide a basis for declaring at most one of those who agree in the C-roles for moral concepts, but disagree in the application conditions, correct, others being mistaken. But to ask what is the right "harmony" relation between application conditions and C-role, if I am right, is just a trivial variation on asking to know which moral judgments are correct. We should apply Dummett's remark, that "we feel that the disagreement is not a straightforward factual one, but at the same time, it is not merely a matter of applying different concepts," to the moral case: the disagreement, we should say, is a *moral* one, and such disagreements are *not* straightforward factual disagreements; but *nor* are they *merely* a matter of applying different concepts. We cannot sensibly say these things unless we view moral concepts as *not entirely* constituted or individuated by their application conditions, or their purely cognitive role, but as combining two aspects that cannot vary quite independently, though there can be a significant amount of "slippage" between them.

One classic feature of moral debates that makes this possible is a pervasive phenomenon in matters, especially, of judgment and evaluation. *Even if* there is a measure of agreement about many relevant parameters of, or considerations relevant to, evaluation, in such situations there is typically no nonarbitrary way of measuring the relative strength or importance of those considerations, and no nontendentious way of combining them so as to *always* arrive at an agreed overall conclusion. Saying the proper way to do things is to use "critical reflection," or to value things "in accordance with rationality," seems to me to provide no real way of progressing in such circumstances.[21]

If moral truth consists only in what judgments or evaluations would be converged on under ideal circumstances, and if which circumstances count as ideal is underdetermined, or irresolvably contested, so there is no single

H1

definitive and canonical notion of ideal convergence, then the truth about corresponding moral disputes must itself be vague, indeterminate, or irresolvable. Quasi-disagreements are here to stay.

There are two motives one may have for thinking about idealized moral consensus. One is to secure the idea of non-agent- or community-relative moral objectivity or moral fact, in the context of something like a response-dependent account of what constitutes such facts, despite the familiar and pervasive actuality of moral disagreement. The other, less noted, but logically prior to the previous, is to secure a common *subject matter* for moral disagreement to be disagreement *about.* But there is something very odd for the latter purpose about substituting for people's actual working moral concepts, an account of concepts they don't but supposedly would all share in some perhaps impossible counterfactual circumstance. At best this would seem to leave us merely with failures to disagree combined with some in-principle commitment to negotiating common concepts in terms of which people might counterfactually express as genuine the disagreements they would then no longer have.[22]

Acknowledgments

Notes

1. This label is *not* intended to connote any particular relationship to what Simon Blackburn calls "quasi-realism."

2. All page references will be to Jackson 1998a unless otherwise indicated. For present purposes I am assuming that this work distils whatever is relevant in Jackson's earlier discussions of moral functionalism, including Jackson 1992, and Jackson and Pettit 1995, 1996.

H1

3. I am under no illusion that anything I say here could be construed as the "last word" on positions or issues here discussed. I have left unanswered many obvious questions about ethical quasi-relativism, which is why I call it a "sketch" of a position. Nor have I tried to compare this view with other views springing in part from similar concerns, and I have not discussed the content or relevance of Jackson's subsequent work on related topics, such as Jackson 1999.

4. Given what follows, it is important to be clear here that a "no-fault disagreement" need *not* be one in which neither party can find grounds to object to the other (that they may do is a plank of quasi-relativism). It is a disagreement in which, symmetrically, no grounds exist on which either party can object to the other, apart from grounds that are, in the context, question-begging.

5. The favored technical term is 'platitudes'. See, e.g., Lewis 1972. In his 1994, Lewis renounces the appeal to platitudes, describing folk theory (in particular, folk psychology) as "tacit" common knowledge, akin to grammatical knowledge.

6. Suppose I have a friend who insists that at least for her, it is analytic that someone is running only if in rapid bipedal motion in which their heels fail to touch the ground. I take her to a sports field and ask her to point out those who are running. Several of those pointed out are repeatedly making contact with the ground with their heels. My friend's supposed concept of running has been shown to be "epiphenomenal" with respect to her practice.

7. The propriety of sometimes adjusting peripheral claims to better fit with core claims will be relatively straightforward in the first sort of case, provided a "realist" rather than a "reductionist" attitude is taken to the theoretical entities mentioned in the "core." Things may be less clear in the second kind of case, since there a real "direction of fit" issue may arise regarding the relationship of supposed "core" to "periphery." If relevant general principles are best seen simply as summarizing the more particular claims that fall under them—whether in the manner of exceptionless generalizations, or in the manner of 'as a rule', 'for the most part' generalities—then it will be inappropriate to modify the more particular claims so as to better fit the principles belonging to the putative "core." That would be proper only to the extent that we had some degree of an independent epistemic handle on the generalizations as such. Otherwise the particular claims would be the "core." Though I have little space to explore it, this is a highly contestable point in ethics, one that has substantial relevance to what follows.

8. Nothing in this section should be interpreted as expressing wholesale skepticism about the method of cases, which I hold to be well-nigh indispensable for many philosophical purposes. See Robinson 2004.

9. Consider folk belief in witches. A physicalistic philosopher might produce a "cleaned up" version of folk witch theory that omits the parts about casting spells and performing black magic, on the grounds that such activities are known to be

H1

impossible. The resulting theory might be satisfied by various harmless if feisty women with an interest in herbalism. It would not be well, however, to declare those women to be witches. A good part of the reason for this is that it would be unreasonable to attribute a commitment to the cleaned up theory being a correct theory of *witches,* to paradigm folk users of the word 'witch'.

10. See also note 7 above.

11. I don't mean the example to be one where Bargle once played a version of the game with deviant rules. He is simply, let us say, easily confused, and has consistently misinterpreted a long series of questionable umpires' rulings.

12. I have put these points in simple generic terms—since any good theory of reference should accommodate them. Readers may wonder how well they gibe with the "two-dimensionalist" semantic framework that Jackson has espoused and promoted. The answer I believe is "quite well," and it is I think interesting, relevant, and illuminating to see how. Space prohibits saying more here, but what I have to say can be found in Robinson 2004, 526–528.

13. Among other implications, this means that an issue raised by note 7—what is the proper "direction of fit" between intuitively endorsed particular moral judgments, and intuitively endorsed general moral principles?—remains very much on the table given Jackson's views. Major questions, not here able to be addressed, about the epistemology of various kinds of response-dependent notion, arise in this connection. But I have a hunch that the risk of intuitions being a product of "epiphenomenal" theorizing is if anything higher in the case of relatively concise general moral principles than in regard to particular case-by-case moral judgments.

14. In practice these two parts are mingled in a test-and-correct process, oscillating between refining the division of cases, and testing and improving hypotheses about principles governing the division.

15. Not to omit the possibility that a relevant "other tongue" might be or include some purpose-built product of science, philosophy, or mathematics, rather than being one of the world's "natural languages" (think for example of "probability").

16. Compare Lewis: " 'Qualia' isn't a term of ordinary language. Neither is 'phenomenal character' nor 'raw feel' nor 'subjective quality'. . . . Yet despite the lack of a folksy word or phrase, I still say that the concept of qualia is somehow built into folk psychology" (Lewis 1995, 140).

17. Jackson (1998a, 131) says moral functionalism's principles "are not causal principles. . . . The principles of folk morality tell us which properties typically go together, but not by virtue of causing each other."

18. What Jackson most explicitly endorses is a response-dependent account of values, but it is clear that he sees moral facts as crucially involving suitable values;

H1

I base what I say here on this fact together with his choice of sample "output clauses."

19. This is very tricky. Jackson (1998a, 131) gives as an example "properties that make something good are the properties we *typically* have some kind of pro-attitude towards" (my emphasis). But 'typically' seems wrong here given the role of "mature" folk morality, in Jackson's account. If current folk morality, by the lights of mature folk morality, currently gets it wrong about some properties, then these will be properties that do not after all make something *morally* good, though current folk morality may mistakenly think so. Furthermore, if current (or mature?) folk morality is to consistently invoke the notion of rationality in the way Jackson suggests, we would expect the clause to take a form more like 'properties that make something good are the properties *those who are ideally rational* typically have some kind of pro-attitude towards'. But perhaps 'good' in this context does not mean only 'morally good'.

20. This is a prima facie difference, premised on the view I am taking of concept individuation. Imagine once again the psychopath who can mimic the application of the predicates 'right' and 'wrong' but is devoid of moral motivation. I say he or she lacks moral concepts. But someone who possesses moral concepts may still in my view count as making moral judgments in applying them, even if on some of those occasions, moral motivation is absent. Jackson's talk of "current, first-order pro-attitudes," however, suggests that a person employing the same concept (descriptively individuated) on different occasions may sometimes make moral judgments in applying it, other times not, as his or her motivation waxes and wanes.

21. One of the most obvious—and difficult—questions I have left outstanding about ethical quasi-relativism is what explains the *constraints* on "slippage" between application conditions and C-roles for moral concepts. Something less superficial—and less local to a particular language—is needed than a mere fact about correct usage of words like 'morality'. Perhaps the model of a set of agreed relevant considerations, but lacking quantifiability or uncontentious algorithms for combination, might form part of the story in accounting for the constraints. But that's a mere conjecture.

22. Much of what I have to say about what I call "quasi-disagreements" has points of similarity with views discussed (in much more detail), and in some cases endorsed, in the excellent book Tersman 2006. Folke Tersman calls such disagreements "radical disagreements." This essay was completed and submitted before Tersman's book, from which this essay might otherwise have benefited, came out.

H1