

# Modal Rationalism & Logical Empiricism: Some Similarities

Stephen Yablo  
MIT

## I. Introduction

You remember logical empiricism, otherwise known as logical positivism. It was the big thing for a while in the middle of this century. A number of ideas were involved: the verifiability criterion of meaningfulness, the analytic theory of the a priori, and so on. As is familiar, most of these ideas came to seem mistaken. And the mistake was a fairly specific one: they ran semantics and epistemology too closely together. Why should meaningfulness, a semantic notion, depend so heavily on verifiability, an epistemological one? Why should a priority, an epistemological notion, depend so heavily on analyticity, a semantical one?

Not only are the postulated connections puzzling in principle, but there look to be counterexamples in both cases. "God exists" strikes many people as meaningful even if unverifiable. " $2=2=4$ " strikes many of us as a priori though not analytic.

So much is ancient history. Why care about these old ideas today? Because, or at least this is my conjecture, the old ideas are still around. A theory of meaning advocated by David Chalmers, Frank Jackson, and David Lewis -- and used by the first now and the second until recently as a basis for rejecting physicalism -- is caught up in something very like the old positivist themes. It's caught up in particular with the reductionist idea that what a sentence means goes hand in hand with the lower-level circumstances that would verify it.

Now this may strike you as the farthest thing from positivism. It may strike you as just the truth-conditional theory of meaning, which is widely viewed nowadays as just common sense, and objectionable only for being so toothless. If you are right about that, then my objection is to the truth-conditional theory of meaning as well. I'll be saying that theory too has more in common with logical empiricism than is usually acknowledged.

## II. Carnap on Internal/External

The aspect of logical empiricism I want to focus on is best known from Carnap's ESO, and Quine's criticisms of ESO in "Two Dogmas." The ostensible topic here is Carnap's distinction between internal existence questions like "are there primes between 10 and 20?" and external existence questions like "are there numbers?" But as we'll see the issue quickly turns to reductionism.

Here's what Carnap says. Meaningful talk about Xs has got to proceed under the auspices of a linguistic framework, which lays down the "rules for forming statements [about Xs] and for testing, accepting, or rejecting them." An ontologist who respects this requirement by querying "the existence of [Xs] within the framework" is said by Carnap to be raising an internal existence-question.

Although it's existential statements that Carnap is focussing on here, the idea is meant to be perfectly general. Any claim or question raised within a framework, and answerable to that framework's rules, is in Carnap's sense internal.

The point about internal questions -- 'is it true that S?' -- is that they raise no difficulties of principle. It is just a matter of whether applicable rules authorize you to say that S. If they do, the answer to 'is it true that S?' is yes; otherwise no; end of story.

Once again, the application that interests Carnap is to internal existence-questions. Looking at what he says about these will help us to understand more about how frameworks work. 'Do numbers exist?' understood internally is easily answered: of course they do, this follows from the fact that there are prime numbers between 10 and 20.

That the answer comes so easily, however, shows that the internal existence-question is not the one the philosopher meant to be asking: it is not the "question of realism." A system of rules making "there are material objects" or "there are numbers" unproblematically assertible is a system of rules in need of external validation, or the opposite. Are the rules right to counsel acceptance of "there are Xs"? It is no good consulting the framework for the answer; we know what it says. No, the existence of Xs will have to be queried from a position outside the X-framework. The philosopher's question is an external question.

Now, Carnap respects the ambition to cast judgment on the framework from without. He just thinks philosophers have a wrong idea of what is coherently possible here. How can an external deployment of a statement about Xs mean anything, when by definition it floats free of the rules whence alone meaning comes?

There are meaningful questions in the vicinity. But these are questions that mention "X" rather than using it: e.g., the practical question "should we adopt a framework requiring us to use "X" like so?" If the philosopher protests that she meant to be asking about Xs, not the term "X," Carnap has a ready reply: "You also thought to be asking a meaningful question, and one external to the X-framework. And it turns out that these conditions cannot be reconciled.

### III. Quine's charge of reductionism

Now let's look at Quine's response to Carnap. The key to Carnap's position (as he sees it) is that "the statements commonly thought of as ontological are proper matters of contention only in the form of linguistic proposals." But now, Quine reasons, similar claims have been made about the statements commonly thought of as analytic; theoretical-sounding disputes about whether, say, the square root of -1 is a number are best understood as practical disputes about how to use "number." So, idea: the external existence-claims can be (re)conceived as the analytic ones. The objection thus looks to be one of guilt-by-association-with-the-first-dogma: "if there is no proper distinction between analytic and synthetic, then no basis at all remains for the contrast which Carnap urges between ontological statements and empirical statements of existence."

Unfortunately, the association thus elaborated doesn't look all that close. For one thing, analytic existence-claims show no particular tendency to be external. Quine appreciates this but pronounces himself unbothered:

there is in these terms no contrast between analytic statements of an ontological kind and other analytic statements of existence such as "There are prime numbers above a hundred"; but I don't see why he should care about this.

Quine's proposal also deviates from Carnap in the opposite way; existence-claims can fail to be analytic without (on that account) failing to be external. An example that Carnap himself might give is "there are material objects." Quine apparently considers it a foregone conclusion that we will have experiences making "there are material objects" assertible in the thing framework. How could it be? It is not analytic that experience even occurs.

And yet Carnap agrees that the distinctions are linked: "Quine does not acknowledge [my internal/external] distinction" because according to him "there are no sharp boundary lines between logical and factual truth, questions of meaning and questions of fact, between acceptance of a language structure and the acceptance of an assertion formulated in the language." The parallel drawn here between "logical truth," "meaning," and "acceptance of a language structure" suggests that analytic/synthetic may define internal/external not directly, by providing an outright equivalent, but indirectly through its role in the notion of a framework. The assertion rules that make up frameworks are not statements, and so there is no question of calling them analytically true. But they are the nearest thing to, namely, analytically valid or correct. The rules are what give X-sentences their meanings, hence they "cannot be wrong" as long as those meanings hold fixed.

Pulling these threads together, internal/external presupposes analytic/synthetic by presupposing frameworkhood; for frameworks are made up inter alia of analytic assertion rules. Someone might ask, "why should analytic rules be as objectionable as analytic truths?" But that is essentially to ask why Quine's second dogma -- the reductionism

that finds every statement to be linkable by fixed correspondence rules to a determinate range of confirming observations -- should be as objectionable to him as the first.

The objection is the same in both cases. Any observation can work for or against any statement in the right context. Hence no assertion or rule of assertion can lay claim to being indefeasibly correct, as it would have to be were it correct as a matter of meaning. Quine may be right that the two dogmas are at bottom one; still, charge narrowly drawn is one of guilt-by-association-with-the-second-dogma, the dogma of reductionism.

#### IV. Reductionism and defeasibility

Now, oddly enough, Carnap agrees that no assertion or rule of assertion is indefeasibly correct. He says already in the 1930s that "all rules are laid down with the reservation that they may be altered as soon as it seems expedient to do so" (Carnap 1937, 318). He agrees that any link between theory and observation can be broken, and any can in the right context be forged.<sup>1</sup>

It is just that he puts a different spin on these scenarios. There is indeed (thinks Carnap) a possibility that can never be foreclosed. But it is not the possibility of our correcting the rules to accommodate some new finding about the conditions under which X-statements are "really true";<sup>2</sup> it is that we should decide for practical reasons to trade the going framework for another, thereby imbuing "X" with a new and different meaning.<sup>3</sup>

The upshot is that while Quine and Carnap agree that the rules are up for grabs, Carnap thinks that the rules are meaning-giving; to change them is not to revise our doctrine but only to switch to an alternative (if similar-sounding) vocabulary. To the extent that Carnap limits this point to artificial languages and frameworks, Quine can agree. But when it comes to languages in use, Quine sees no basis for Carnap's meaning/doctrine distinction; and so he maintains that virtually any change in the rules can and should be seen as having a doctrinal element. It's not that we give "X" a different meaning, it's that we change our minds about where the Xs are.

#### V. Modal rationalism

So much by way of explaining the Carnap/Quine argument about reductionism. Briefly, Carnap thinks it's a semantical matter under which observational conditions it's to be asserted that S, whereas Quine thinks it's a doctrinal matter. It's Carnap's position that I say bears comparison with modal rationalism.

Here to a first approximation is the modal rationalist picture. Associated with every sentence S are two propositions or sets of worlds. The first, the primary intension, is

$|S|_1$  = the set of all worlds 'verifying' S  
= the set of all w such that if w is actual then S is true.

This gives the a priori aspect of meaning; one knows just in virtue of understanding S what the actual world has to be like for S to be true. The secondary intension is

$|S|_2$  = the set of worlds 'satisfying' S  
= the set of all w that S truly describes.

This gives the 'what is said' aspect of S's meaning. To the extent that what is said depends on contextual factors, one doesn't know S's secondary intension just in virtue of understanding S.

---

1  
2  
3

Suppose for instance that S is "water is plentiful." Then what S in fact says, assuming that the watery appearances around here are appearances of H<sub>2</sub>O, is that H<sub>2</sub>O is plentiful. The secondary intension is thus the set of worlds with lots of H<sub>2</sub>O. But what S requires of this world is different; it requires that the watery stuff be plentiful. So the primary intension is the set of all worlds with lots of watery stuff. And now here are the fundamental 'predictive' theses:

(A) it is a priori that S iff  $|S|_1$  is the set of all worlds, that is, iff S is true in virtue of the a priori aspect of its meaning.

(N) it is necessary that S iff  $|S|_2$  is the set of all worlds, that is, iff what S says is the case must be the case.

(A) and (N) neatly explain how S can be necessary yet aposteriori, or contingent yet apriori. But the point that I want to focus on comes earlier, in the intuitive explanation of primary intension as the a priori aspect of meaning. To call primary intension the apriori aspect of meaning is to say in effect that

(P) whether a world w verifies S is determined by the apriori aspect of its meaning.

The word 'verifies' here means just: is such that S is true if w in fact obtains. No connection is intended with the logical empiricist doctrines of verificationism and reductionism. But that's not say to say there is no connection.

## VI. Initial comparisons

First let's clear away some apparent differences between modal rationalism and Carnap-style reductionism. Reductionism envisages analytic rules connecting theory to experience. MR doesn't envisage anything like that. Yes, people have to be told a priori whether S is true in a presented world. But Chalmers and Jackson don't say that the world is presented experientially. (Both seem to think that worlds are presented descriptively; more on this later.) There's no question then of associating MR with an experiential or phenomenalistic reductionism.

So far so good. But if one looks at Carnap's writings on "protocol sentences," it turns out that LE reductionism does not have to be terribly experiential either. Under the influence of Neurath, Carnap thinks it is somewhat of an open question which sentences ought to be counted as protocols. Sometimes a protocol-sentence is said to be any sentence "belonging to the physicalistic system-language" which we are prepared to accept without further tests (Ayer, 237). Usually it is said to be a matter of convention which sentences will count as protocols. The important point for our purposes is that Carnap thinks there are analytic rules connecting theoretical statements with protocols whatever protocol turn out to be.

Another seeming difference between (logical) empiricism and (modal) rationalism emerges from Quine's complaint that reductionism overlooks the holistic nature of confirmation. The complaint might be this: One never knows whether S is really correct until all the observational evidence is in, which it never is. Hence any rules portraying S as verifiable on the basis of limited courses of experience -- courses of experience small enough to be enjoyable by particular observers -- would be untrue to the way confirmation actually works. This complaint the (modal) rationalist can rightly claim to have avoided. The (m) rationalist never represents limited information about a world as enough to ensure that S; the rules he contemplates take as input complete information. Dave is very clear about this:

[Quine says that] purported conceptual truths are always subject to revision in the face of sufficient empirical evidence. For instance, if evidence forces us to revise various background statements in a theory it is possible that a statement that once appeared to be conceptually true might turn out to be false.

This is so for many purported conceptual truths, but it does not apply to the supervenience conditionals that we are considering, which have the form "If the low-level facts turn out like this, then the high-level facts will be like that." The facts specified in the antecedent of this conditional effectively include all relevant empirical factors-.The very comprehensiveness of the antecedent ensures that empirical evidence is irrelevant to the conditional's truth-value (TCM, 55).

But there is another way of construing Quine's complaint that makes it quite unrelated to the "total evidence" issue. Quine says that

the dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion, issuing essentially from Carnap's doctrine of the physical world in the Aufbau, is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body (TDE, 41).

Never mind the irony of Quine attacking Carnap for overlooking a point he himself gets from Carnap. Quine's complaint here is not that S's confirmational status is underdetermined until all the empirical evidence is in; it's that S's confirmational status is not fully determined even by the full corpus E of empirical evidence. The degree to which E confirms S, Quine thinks, is tied up with the extent to which E or aspects of E are deducible from S. But nothing of an observational nature is deducible from S except with the help of a background theory T. Hence the degree of support that E lends to S depends on which background theory we use.

This complaint would be easily evadable if there were an analytically guaranteed fact of the matter about which theory T is selected for by E. For in that case one could simply ask whether E supports S relative to the E-preferred theory whatever it might be.

One has to assume, then, that this is what Quine is really concerned to deny; he denies that there are analytic connections between the total corpi (?) E of empirical evidence and theories T of nature, and hence there are no analytic connections between E and particular statements S. A number of things suggest that these analytic confirmational relations are indeed the target:

I am impressed, apart from prefabricated examples of black and white balls in an urn, with how baffling the problem has always been of arriving at any explicit theory of the empirical confirmation of a synthetic statement (TDE, 49)

This could be taken to mean just that the sought-after theory of confirmation would have to be very very complicated. But Quine has something different in mind. He is aware after all of Carnap's attempts to work out a logic of confirmation which would tell us what to believe on the basis of such and such observations. He is aware too that the attempt failed even for the simplest sort of examples. Carnap came up with a bunch of confirmation functions (the m-functions) and was forced to regard the choice between them, and hence between the statements they recommend, as one that could not be made on logical or semantical grounds.

What then is Quine's point? His point is underdetermination: "total science, mathematical and natural and human, is underdetermined by experience" (TDE, 45). The version of underdetermination he needs is really a rather mild one. He needn't deny that there is an objectively best theory relative to a given body of evidence. He needn't even deny that there's a single most rational theory to adopt. All he need claim is that the choice between theories compatible with the evidence cannot be made on purely analytic grounds. Our choice of theory

turns on our vaguely pragmatic inclination to adjust one strand of the fabric of science rather than another. Conservatism figures in such choices, and so does the quest for simplicity.

This can be reconciled with the analytic view of confirmation relations only by supposing that meanings suffice to determine for us how conservative we should be, and what sort of simplicity to favor, and how to trade these sorts of desiderata off against each other. The idea would have to be that when people differed on what sorts of tradeoffs made sense, at most one could be considered to be speaking correctly, ie., in accordance with the meanings of her words. But that, it seems clear, is not how the science game is played.

## VII. How we operate

Oddly enough, Carnap agrees (at least some of the time) that that's not how the game is played. His goal as he usually describes is not to describe our actual practice, but to give us tools for making the practice more rational and more efficient. He thinks disputants should pick a common framework and then resolve their disagreements by reference to its assertion rules. One can argue about whether this would really be helpful. All I'm saying right now is that not even Carnap believes that it's how we really operate.

Is there anyone who does believe it's how we operate? It seems to me that the modal rationalist believes it, or at least that such a view is implied by modal rationalism. Here is a first stab at convincing you of this. The modal rationalist says that grasp of S's meaning, or at least the kind of grasp you need to count as understanding S, is knowing which worlds *w* are such that if this is *w*, then S. This applies not just to observation-level statements, but theoretical statements T as well. It would seem to follow that the rationalist has to see it as part and parcel of knowing T's meaning to know what the world has to be like for it to be the case that S. And that is not obviously different from Carnap's idea of analytic confirmation rules.

I say "not obviously different" because it's unclear to begin what form this "knowing which worlds are S-worlds" is supposed to take. Under what description exactly are we supposed to think of the relevant set of worlds?

One idea about this is clearly not intended. It's not supposed to be that my grasp of the meaning of "there are horses" lies in the fact that I know it to be true in the worlds with horses, or rather the ones such that if they are actual then there are horses. For suppose that a 'homophonic' grasp of the set of verifying worlds was all one needed. Then contrary to (A) above, there would be no reason at all to expect a sentence to be knowable a priori just because its primary intension contained all worlds.

The easiest way to explain this is with an example from Chalmers's discussion of physicalism. Consider the conditional "if things are physically like so, then there is pain." According to (N), the only way for this to be non-apriori is for there to be worlds not in its primary intension: there have got to be zombie-worlds. But if a homophonic grasp of the primary intension is all that's required, then we can explain the non-apriority of the conditional without bringing in zombie-worlds at all.

It'd be enough to point out that to understand the conditional it suffices to understand its antecedent A and consequent C, i.e., to know that their primary intensions are the A-worlds and C-worlds respectively. And to know that much about the primary intensions you don't need to know the set-theoretic relations between them. Even if the physically-like-so worlds were included in the pain-worlds, there'd no reason why the understander should realize this, and hence no reason to think she should be able to determine the conditional's truth value just on the basis of her understanding.

What the modal rationalist intends then is that to understand S one should know what lower-level facts have to obtain in a world if it is to verify S. And this is indeed what Chalmers says:

The supervenience conditionals that we are considering, which have the form "If the low-level facts turn out like this, then the high-level facts will be like that." The facts specified in the antecedent of this conditional effectively include all relevant empirical factors...

These conditionals are for the modal rationalist analytic in the sense of being true in virtue of meaning; indeed they are true in virtue of the aspect of meaning to which we have a priori access. This is why I say that the modal rationalist is committed to something prima facie analogous to the analytic confirmation relations advocated by Carnap and rejected by Quine. The modal rationalist who wants to escape Quine's criticisms has got to do one of two things:

- (a) show that the criticisms don't work even against logical empiricism;
- (b) show that the cases are relevantly different.

To accomplish (a) would be to find a mistake in Quine's reasoning. Maybe, for example, it's just untrue that theory is underdetermined. To accomplish (b) would be to show that what the modal rationalist says is different enough from what the logical empiricist says that the Quinean critique doesn't generalize. Maybe, for example, the lower-level facts on the basis of which we can tell a priori whether S are quite unlike the "empirical" facts on the basis of which we can't tell a priori whether S. The rest of the paper explores in a roundabout way the prospects for doing either of these things, and thereby escaping Quine's critique.

### VIII. Reference-fixing

I've been saying, basically, that the rationalist's analytic supervenience conditionals are hard to square with the Quinean doctrine of 'underdetermination of theory.' The version of underdetermination we need is a mild one. All we need claim is that the choice between theories compatible with the evidence cannot be made on purely analytic grounds: considerations about which theory is simpler or clearer or more natural inevitably come into play. This alone makes it hard to see how the supervenience conditionals can be analytic. Because the issue of which supervenience conditionals are correct cannot be separated from the issue of which theory is correct given the facts in the conditional's antecedent. If the one issue is not analytically resolvable, then neither is the other.

One way that the rationalist tries to weaken our resistance to the analyticity claim is by bringing in the notion of a reference-fixing condition or description. Suppose for argument's sake that water is stipulated to be the local clear drinkable stuff. (This is an idealization but it is said a harmless one.) Then just by virtue of understanding 'water,' I know what its extension is relative to any assumption about the lower-level facts.

I wonder though whether the idealization is really so harmless. To call it harmless suggests that even an unidealized understanding of 'water' would determine the word's extension relative to any hypothesis about the lower-level facts -- so that if someone disagreed with me about where the water was in a world specified in lower-level terms, that would show she meant something different by water. Is this plausible?

It seems to be agreed all around that the extension of 'water' in a world is at least somewhat a function of what the right scientific theory is. But then if meanings aren't enough to single out the right scientific theory, they are unlikely to single out the right extension for 'water' either. It would seem to follow that the reference-fixer for 'water,' to the extent that it's true in virtue of (an a priori accessible aspect of meaning), cannot possibly give explicit necessary and sufficient conditions in lower-level terms; there just aren't any lower-level conditions that are analytically guaranteed to deal correctly with all scenarios, no matter how outre. A reference-fixer guaranteed to give the right results cannot avoid appealing at some point to notions like naturalness, simplicity, nonarbitrariness, etc. -- in a word reasonableness. But then the supervenience conditionals are not analytic unless it's analytic that such and such is the most reasonable thing to say about an imagined scenario.

Now there are places where Dave seems to agree that considerations of reasonableness will have to come in:

In certain cases, the decision about what a concept refers to in the actual world involves a large amount of reflection about what is the most reasonable thing to say; as, for example, with questions about the reference of 'mass' when the actual world turned out to be one in which general relativity is true, or perhaps with questions about what qualifies as 'belief' in the actual world-consideration of just what the primary intension picks out in various actual-world candidates may involve a corresponding amount of reflection. But this is not to say that the matter is not a priori: we have the ability to engage in this reasoning independently of how the world turns out. (58).

This sounds right; we do have some ability to engage in this reasoning independently of how the world turns out. The question is whether the results of our reasoning deserve to be counted into the meanings of words like 'mass' and 'belief.'

Of course one is free to define 'meaning,' or 'a priori component of meaning,' however one wants. But we need to be clear that this is somewhat of a novel use. The usual view is that differences about the application of a word in specified circumstances sometimes reflect differences in meaning, sometimes due to differences in doctrine:

differences about what it's reasonable on balance to believe. The effect of modal rationalism is to eliminate this second category, so that people who have different ideas about what (it's reasonable to think) various people in fact believe in a given situation come out meaning different things by the word 'belief.' This makes an enormous difference to the significance of the rationalist's claim that the supervenience conditionals are analytic or true in virtue of meaning.

### **IX. The meaning of 'reasonable'**

But we are getting ahead of ourselves. The question we need to be asking is, how far should differences about when to apply a word be taken to reflect differences in meaning? It helps to focus on the word 'reasonable' since it is conflicting judgments about reasonableness that underlie disagreements about the extensions of other words.

So: is it the case that anyone who means what I do by 'reasonable' has got to agree with me on the reasonableness or not of proposed courses of action? Should I regard it as an analytic consequence of the lower-level descriptive facts that this course of action is the proper one to pursue?

I can't regard it as an analytic consequence, it would seem, without falling afoul of the open question argument. Imagine someone who says, 'that this alternative is descriptively so and so entails that it is the alternative that makes the most rational sense.' This person can be accused of not getting it. It flows from the concept of reasonableness that it's always in order to wonder whether the proposed option is really the most reasonable one. The question is, as they say, always open. And an open question, whatever else may be true of it, is a question whose correct answer is not analytically true.

So: the fact that one can't avoid appeals to reasonableness in deciding extensions shows that the supervenience conditionals are not analytic. Or rather, it shows this if we agree that the open question argument prevents there from being an analytically determined fact of the matter about which option is rationally warranted.

### **X. The open question argument**

Given our reliance here on the open question argument, it has to be counted as a marvelous stroke of luck that at least one modal rationalist, Frank Jackson, has considered this argument and rejected it. Jackson is responding to the argument as it concerns 'right' rather than 'reasonable.' He thinks that it's not an open question what is the right thing to do in a given set of descriptive circumstances. Which actions are right is analytically determined by the lower-level descriptive facts. First let's see how it's determined; then we can ask how the open question is supposed to be closed.

Take just about any predicate F, Jackson thinks, and the following is analytic: Fness is the property playing the F-role in folk F-theory. This holds in particular for the property of rightness -- with the twist that it is not current folk morality that calls the shots but a "maturer" version. Reference is pegged to the place "where folk morality will [would] end up after it has been exposed to debate and critical reflection" (133). It is analytic he thinks that an action is right iff it has whatever property D meets descriptive condition M, a condition extracted by the Ramsey-Lewis method from mature folk morality.

This would seem to make Jackson easy prey for the open question argument: "I see that keeping my promise has the M-property, but is it right?" He has two main replies.

(1) Since "what matters is the nature of mature folk morality, there will, here and now, inevitably be a substantial degree of 'openness' induced by the very fact that the rightness role is currently under investigation" (151).

(2) If the question still seems open when "all the negotiation is over and we have arrived at mature folk morality, we are entitled to dig in our heels and insist that the idea that what fits the bill that well might fail to be rightness is nothing more than a hangover from the platonist conception that the meaning of 'right' is somehow a matter of its being mysteriously attached to the form of the right" (152).

Suppose we consider these in turn. If we accept reply (1), then we ground the openness of "is that right?" in the openness (for us today) of "is that mature folk morality?" Jackson talks as though this latter openness were a matter of ordinary descriptive ignorance about the future; we haven't got there yet, so we don't know.

But of course morality could take a direction that we would regard as quite misguided, and for good and specific reasons; tomorrow's moralists might end up resting a lot on tendentious analogies, or giving in to self-interest, or etc. The point is just that it is no part of current folk morality to defer to whatever comes along. If deference must be paid, it is owed to those (probably hypothetical) future populations who had thought things through carefully, reasonably, and with due concern for all. So the best (1) can hope to accomplish is to ground the openness of "is it right" in that of "is that the most careful, reasonable, etc. way of developing the theory from here?"

But "careful," "reasonable," 'due concern' and so on are themselves evaluative terms. So Jackson faces a choice. Either he maintains about reasonableness too -- let it stand in for the rest -- that it is analyzable in descriptive terms, or he treats it as irreducibly evaluative.

One assumes that he will stick with his earlier descriptivism. "Reasonable" picks out whatever property plays the reasonableness-role in mature folk reason-theory, the theory we would arrive at if existing folk reason-theory were developed along the most reasonable lines. But this is circular; "the most reasonable lines" are the lines that would be identified as such by the theory lying at their terminus. So it is hard to see reply (1) as helping matters; if anything it confirms the impression that the meanings of 'right' and 'reasonable' leave their extensions a matter for debate.

Of course it's agreed that their extensions are a matter for debate now. But Jackson thinks that our current moral practice is predicated on the idea that the debate continued along present lines will eventually reach a conclusion. He says 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 moral debate" (137).

But commitment to moral debate is one thing, commitment to convergence is another. It would sufficiently explain the first commitment that we always hope to find a basis for agreement; that we expect that agreement will (or would) come seems like a hypothesis running far ahead of the data.

## **XI. Communication**

A different reason for imputing a commitment to convergence has to do with moral communication. Suppose that we were to arrive not at "a single mature folk morality but rather different mature folk moralities for different groups in the community" (137). Then, Jackson thinks, "the adherents of the different mature folk moralities will mean something different by the moral vocabulary because [their] moral terms...will be located in significantly different networks" (137). One assumes that current speakers too, to the extent that they belong to precursors of Jackson's "different groups in the community," will have meant something different by their moral vocabulary.

But then, insofar as we see today's factions as in danger of evolving into "different groups in the community," we see them as in danger of not really communicating, on account of their similar-sounding words having different semantic values. Contraposing, we have got to see ourselves as converging if we want to see ourselves as communicating.

I don't know whether Jackson is attracted to this reasoning or not. But it looks to be in tension with a point he makes elsewhere in his Locke Lectures, in his discussion of color. He says that the folk might well take it for granted that there is "a kind, indeed a natural kind, distinctive of the exemplars of water and gold. [But] the folk are too sensible to have presupposed something as risky as that there is a distinctive kind in common to things we call 'red'" (108). The presupposition would be "risky" because it would put the legitimacy of our practice with "red" at the mercy of developments that we're not in a position to prejudge. Presupposing moral convergence would be risky in the same sense. If the folk don't do it with "red," it stands to reason that they wouldn't do it with "right" either.

Then what does entitle moral disputants to regard themselves as communicating? Not enough attention has been paid to the cognitivist who answers like this: If something ought to be written into our folk theory, it's not that risky conditions ABC are met -- given which such and such a semantic mechanism sees to it that we attribute the same property by "right" -- but simply that you must mean something different by "right" is a diagnosis of the very last resort. The project of using evaluative language together in figuring out how to act is so overwhelmingly important that we do not allow others to opt out with this facile semantic excuse. Our primary commitment in other words is that "right" and similar action-guiding terms should stand for the same or similar properties in all our mouths.

The commitment comes at a price: a certain kind of anti-dogmatism. The readier we are to claim conceptual authority for our own evaluative views -- to say they follow from what "right" means in our mouths -- the harder it becomes to hold onto the idea of coreference as between disputing parties. But then, rather than its being a condition of moral communication that we expect to arrive at a single moral truth, the proper condition is that nothing will ever be regarded as the point of arrival: the point at which reference is finally fixed and moral theory acquires a conceptual imprimatur. This is what makes Jackson's second response to the open question so counterintuitive. He says that

the idea that what fits the bill that well might fail to be rightness is nothing more than a hangover from the platonist conception that the meaning of "right" is somehow a matter of its being mysteriously attached to the form of the right (152).

I would have thought it was part of the the bill that what satisfied it might still fail to be rightness. This is not because of platonism but the opposite: we refuse to attach "right" to any descriptive property so tightly that moral dissidents, even hypothetical ones, come out as simply misusing "right." Similarly, it seems part of what we have in mind by "reasonable" that the door is left open to the brilliant iconoclast who gets us to see that we have all along been acting contrary to reason. To dismiss such a person as meaning something else by "reasonable" strikes us as dogmatic. I doubt that as folk reason-theory matures it will take a kinder view of dogmatism than we do today.

## **XII. Meanings as reason-laden**

I have been saying that facts about what it's reasonable to say or do are one thing, facts about meaning another. As we've just been seeing, at least one modal rationalist denies this. He thinks that the meaning of 'reasonable' is enough to determine (with the help of lower level descriptive facts) its extension: what it is that's in fact reasonable. Now let's look at a second way in which the semantic and the epistemic get mixed up on Jackson's picture. Not only does the meaning of 'reasonable' close off all questions about its extension, its extension has an unexpected influence on the meanings of other words. The meanings of other words in my mouth are skewed by my sense of what's rational. Let me explain..

Suppose we have for each possible world  $v$  a complete physical specification  $P_v$  in physical terms; and suppose that (as Jackson maintains) my understanding of a sentence  $S$  is given by the truth values I assign to instances of the schema:

$(S_{vw})$  if  $P_v$  is true, then  $S$  would have been true had it been that  $P_w$ .

This applies in particular to statements  $S$  about mental states, including phenomenal mental states.

It seems clear, though, that one cannot determine the truth-value of phenomenal statements with respect to physically given worlds without asking what it would make sense to say about these worlds. What is the the best explanation of the fact that there is weeping and wailing and gnashing of teeth? That the weepers are in in pain, or that they are in such and such brain states but feel nothing? If you and I have sufficiently different ideas about this -- because of our larger theoretical and practical projects, how anxious we are to avoid multiplying entities, etc. -- we may be expected to assign different truth-values to  $S_{vw}$ . But then, just because of our methodological orientations, we wind up meaning different things by "there is pain"!

Another example: Suppose that you and I are confronted with a world in which events with the characteristic physical manifestations of pain occur on all the same occasions as events that we agree deserve to be called c-fiber-firings. You decide for ontological economy reasons that "pain" and "c-fiber firing" pick out one and the same type in this world. I am not too worried about ontological economy; I decide on the basis of my attachment to psychological autonomy and modal intuition that they pick out different though de facto correlated types. What does the Jacksonian physicalist say about this case? Again, that you and I mean different things by "pain."

That is not how ordinary speakers see it. Remember the great identity debates of the 1950s, when it was assumed that mental/physical correlations would soon be found and the question was what ontological conclusions to draw. The disputants didn't think of these debates as driven by differences about the meaning of "pain"; they thought they were arguing about the metaphysics of pain.

Of course, to repeat the obvious, everyone is entitled to use words however they like. But if Jackson is using "conceptual" in a special sense, to describe differences which others would classify as doctrinal, then that bears on the interpretation of his claim that physicalists are committed to the 'conceptual' entailment of the psychological by the physical. It's not clear that the claim comes to more than this: those who find it on balance reasonable to apply mentalistic description to a physically given world can portray those who disagree with them as "meaning" something different by the mentalistic description.

### **XIII. Physicalism**

Now finally a word about the Jackson-Chalmers objection to physicalism. This objection turns in large part on the idea that there is no analytic or a priori entailment from our world's physical condition to its phenomenal condition. By thesis (A) in the first section, the failure of analytic entailment testifies to the existence of a world physically like ours but lacking in consciousness -- a so-called zombie world.

The modal rationalist chalks it up to my semantic competence with a word how I find it reasonable to apply that word in given circumstances. So it would be then with 'pain' as well. Now suppose that I find it reasonable to say of a world physically just like ours that there is lots of pain in it, and count the world into the primary intension of 'there is pain' as a result. Then it comes out analytic in my idiolect that if things are physically like so, there is pain. No failure of analytic entailment means no failure of analytic entailment to 'explain' by postulating a zombie world.

So to the extent that I was worried that I might have to give up my physicalism on account of the lack of an a priori link between physics and pain, I should stop worrying. The missing apriori link has been restored.

Am I saying that physics a priori necessitates pain, full stop? Not at all. The claim is rather that the modal rationalist faces a dilemma. If meanings are individuated so that people who disagree about what to think thereby mean different things, then for at least some of us, "if the world is physically like so, then there is pain" is true in virtue of (the a priori aspect) of meaning.

If on the other hand meanings are not to be individuated that finely, then the rationalist's claim that lower-level physics analytically necessitates everything but consciousness becomes hard to believe. It's directly at odds with a version of Quine's underdetermination thesis. And not a particularly controversial version. Positivism aside, it seems generally agreed that the methods we use to get from lower level data to high theory – methods like inference to the best explanation, Ockham's razor, and so on – are although perhaps rational not analytically valid. This is what the modal rationalist who takes on the second horn of the dilemma must deny.