DISCUSSION

RED, BITTER, BEST

1. Introduction

I would start by saying that every analytic philosopher should buy this book (*From Metaphysics to Ethics. A Defence of Conceptual Analysis*, by Frank Jackson (Clarendon Press, 1998. xii + 174 pp. £25.00)), but I suspect that most of them already have. They got their money's worth. The book operates on two levels, advancing a grand-scale methodological hypothesis while trying the hypothesis out on some ground-level disputes. The grand-scale hypothesis is that conceptual analysis, properly understood, is just as central to analytic philosophy as in glorious days gone by. The ground-level disputes are in various areas but especially mind, colour, and morality. A theme throughout is that physicalists are committed to the a priori deducibility of truths about— well, anything you like—from physical premises.

A rough and selective guide to the book's contents: "Serious metaphysics" tries to show how "a limited number of . . . basic notions" (p. 4) can be woven together into a complete description of reality—a story that implicitly contains (read: *entails*) any truth you care to mention. Once the story is found, we have a hoop we can ask candidate features of reality to jump through; for

1. D. Raffman, 'Vagueness without Paradox', *Philosophical Review*, vol. ciii (1994), pp. 41–74.
2. See Ch. 7.
3. R.M. Sainsbury, *Paradoxes* (Cambridge University Press, 2nd edition, 1995); see Ch. 2.
4. See the editors' introduction.

13

nothing is real unless the story entails it. Physicalists think they can tell the story in physical terms, so they must hold that nothing is real unless its presence is demanded by what goes on physically.

Chapter 2 finds a place for conceptual analysis in all this. To get a would-be phenomenon through the hoop—to solve what Jackson calls the "location problem" for that phenomenon—we need to establish "entailment theses between matters described in some preferred vocabulary and matters described in other vocabularies" (p. 28). And what is conceptual analysis if not the study of inter-vocabulary entailments? Suppose that a physicalist wants to defend freedom of the will. Her job is to show that 'it was done freely' is entailed by true physical premises. To do that, she will need to get clear on the conditions under which an action is properly described as free.

Described by whom? Not just anyone; the topic is "free action according to our ordinary conception" (p. 31). That conception is obtained by ramsifying the community's 'folk theory' of freedom, a theory teased out by consideration of possible cases. (Like the empirical scientist, we seek "the hypothesis that best makes sense of [subjects'] responses" (p. 36).) An action is properly described as free—this is a conceptual truth for Jackson—iff it has the property playing the freedom-role in ramsified folk freedom-theory.

The exercise is not to be conceived as obnoxiously prescriptive. If a folk concept forces unwelcome conclusions on us, we may seek to tweak it a little—change what we mean by the word—as the compatibilist perhaps does with the folk conception of freedom. (This is called "pay[ing] due homage to Quine's critique of analyticity" (p. 44).) But that tweaking is necessary just underscores the fact that we are dealing with sentences that "that could only come false by virtue of meaning change" (p. 46)—sentences that are analytic in the sense that their meaning guarantees their truth, and a priori in the sense of being true on any hypothesis about which world is actual. So in another way Quine's critique is simply shrugged off.

Chapter 3 asks: How did we get from the idea that all truths have got to be entailed by 'basic' truths to a conclusion about what is knowable a priori? Your average physicalist would much rather see the entailment relations as necessary a posteriori. And given the work of Kripke, there would seem to be nothing to stop her.

Jackson would say that there is nothing to *start* her, if she takes care not to read *Naming and Necessity* in a mystery-mongering way. There is only one kind of necessity in that work; what there are two of is *propositions* expressed by a given sentence. One, the C-proposition, is the set of all worlds that *S*, as used by us here in actuality, truly describes. So, the C-proposition expressed by "water = H_2O" is the set of all worlds, or all containing water. (This is what Americans tend to think of as *the* proposition expressed by *S*.) The A-proposition is the set of all worlds *w* such that, if we suppose for argument's sake that the actual world is *w*, then *S* is true. So, the A-proposition expressed by "water = H_2O" contains the worlds in which H_2O is the clear drinkable stuff of our acquaintance. The alleged distinction between metaphysical and epistemic necessity is really no more than the distinction between the necessity (simpliciter) of the C-proposition and that of the A-proposition.

14

All right, but why can't our physicalist just restate her point in these terms? What she 'meant' is that physical claim *P* entails mental claim *M* in the C-sense—the C-proposition expressed by *P* entails the one expressed by *M*—but not in the A-sense.

She can, but this only buys her a little time. The reason has to do with the nature of understanding. To understand a sentence is to appreciate how its C-proposition depends on context. If a speaker who understands *P* and *M* fails to see that *P* entails *M*, that is because he doesn't know enough about context to figure out which C-propositions they express. When the gaps in his contextual information are filled—as according to the physicalist can be done with a physical sentence *P**—the speaker is out of excuses. Armed with *P**, he can figure out which C-propositions *P* and *M* express, whereupon he will see that every world in the one belongs also to the other. Hence the physicalist's commitment to the "in principle a priori deducibility of the psychological [of *M*] from the physical [from *P* & *P**]" (p. 83).

Chapter 4 takes up the location problem for colour. A "subject-determining platitude" about yellowness is that it is "the property of objects putatively presented to subjects when those objects look yellow" (p. 89). Since presentation requires causation, and dispositions are causally inert, dispositional theories of colour—including those that identify colours with reflectance-dispositions—are dead on arrival. It has been clear for a long time that "the only causes . . . of objects' looking yellow are complexes of physical properties" (p. 93). And so yellowness must be a complex of physical properties.

A perhaps surprising consequence is that "the way [colours] look does not reveal their essential nature" (p. 102). But 'revelation' was never part of our folk theory of colour anyway. It must be conceded too that, insofar as different physical properties make things look yellow on different occasions, there is no single property of yellowness shared by everything yellow. But, on the one hand, not-excessively-disjunctive properties can be causes, so a common physical cause can be found more often than you might think. And, on the other, it does no great violence to our folk conception of yellowness if it splits jadeishly into a small number of disparate types. If there were *no* "interesting distinctive distal commonalities underlying similarities of apparent colour" (p. 112), then colour would have to be regarded as an out and out illusion. But Jackson thinks that unlikely.

The last two chapters address the location problem for ethics. Jackson assumes cognitivism as against expressivism: ethical sentences are in the business of "saying how things are", even if some of them for reasons of vagueness (or whatever) lack truth-value. He concludes that "ethical properties are descriptive properties" (p. 117). The argument is simple enough. It's a subject-determining platitude that nothing can be called 'rightness' that fails to supervene on the descriptive. But then the disjunction of the total descriptive natures of all possible right actions is necessarily equivalent to rightness, and so, fussiness about property individuation aside, identical to rightness.

Which descriptive property is it? The one playing the rightness role in (ramsified) folk morality—with the twist that it is not *current* folk morality that calls the shots but a more "mature" version. Reference is pegged to the place

15

"where folk morality will [would] end up after it has been exposed to debate and critical reflection" (p. 133). Jackson agrees with the Cornell realists that ethical properties are descriptive properties, and agrees (or can) that the identities here are necessary a posteriori. But he rejects their scepticism about the possibility of analysing ethical language descriptively. It is analytic that an action is right iff it has whatever property D meets descriptive condition E, a condition extracted by the Ramsey-Lewis method from mature folk morality.

This might seem to make Jackson easy prey for the open question argument: 'I see that keeping my promise has the E-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" (p. 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" (p. 152).

2. Ethics

It says a lot for the connectedness of the book's themes that later chapters raise similar issues to earlier ones. Let me start with Jackson's reply to the open question argument and work backwards. (I'll be brief with the chapter on colour; see my 'Singling out Properties'.)

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 occasionally 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 the like. 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 should be 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, and so Jackson faces a choice. Either he maintains about reasonableness too—let it stand in for the rest—that it is analysable in descriptive terms, or he treats it as irreducibly evaluative.

Suppose that he sticks 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

16

lines' are the lines that would be identified as such by the theory lying at their terminus.[1]

Where does this leave us? One option is to say that 'reasonable' does in fact stand for a definite descriptive property R, but only someone who already knows which property R is is in a position to 'identify' R via Jackson's instructions. Some such concession seems called for anyway. What the reference-fixer

(#) the property playing the X-role as defined by (mature) folk X-theory

stands for depends on the referents of (#)'s component predicates and their analysanda; so how can (#) be expected to tell those not already in the know the referents of those same predicates? This is important because if we have to know which property R is before (#) can 'identify' R for us, then it is no longer true that R can be identified on the basis of descriptive information alone. And so we lose our former reason for thinking that reasonableness-claims are a priori deducible from descriptive ones.

A second option is to deny that R exists—to deny, that is, that *any* descriptive property is such that 'reasonable' stands for *it* as opposed to some other descriptive property. This might look like an impossible position. Don't the same supervenience considerations that led Jackson to his descriptivism about the ethical apply here too? But the most that comes out of supervenience is that any decent *candidate* for the reasonableness-role had better be descriptive in nature. It's compatible with this that there are and will continue to be many candidates, so that there is no particular descriptive property that deserves to be called 'the' property of reasonableness.[2]

And now someone might say: Why should it make our blood race if 'reasonable'—and by implication 'right'—is determined to mark some descriptive line or other, if the question of the word's application to particular actions remains as contestable as ever? A finding of truth-aptitude begins to seem like a pretty superficial victory for the cognitivist. He is right about the technical semantical point: each evaluative predicate stand for a descriptive property. But when it comes to the larger issue of whether there is a *truth* of the matter about what is right (reasonable), matters are more or less as described by the non-cognitivist.

1. Could it be that Jackson thinks our existing folk-theory of reasons is good enough? His one comment on the issue sounds a (to me) surprisingly optimistic note: "it is certainly true that we cannot, as of now, write down in a natural language necessary and sufficient conditions for being rational. (Though we can say something useful and to the point—whatever the defects of the inductive logic sections of textbooks and extant discussions of experimental design, they are very far from useless.) What would be incredible . . . would be if there were no story to be told constructible from our folk-classificatory practice: we are finite beings; we do not work by magic; we give useful information to each other by means of the word "rational". There must, therefore, be a story to be told (extracted). And when it is told (extracted), rationality will have been codified" (p. 67).
2. This would not necessarily mean that 'reasonable' was vague or ambiguous. It picks out a definite *evaluative* property, viz. reasonableness. Why should it matter that there's no agreed-on way to map the property out descriptively? We can't map Kiliminjaro out atomically, either; that doesn't stop 'Kiliminjaro' from picking out a definite mountain. See also note 2, p. 19.

What would Jackson say about this mitigated cognitivism? He maintains 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 moral debate" (p. 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; the further claim that we *expect* agreement seems like a hypothesis running far ahead of the data.

A different reason for imputing a commitment to convergence has to do with moral debate. What if we were to arrive not at "a single mature folk morality but rather different mature folk moralities for different groups in the community" (p. 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" (p. 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 in his discussion of colour. 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'" (p. 108).[1] The presupposition would be "risky" because it would put the legitimacy of our practice with 'red' at the mercy of developments that could for all we know turn out badly. 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 is to be written into our folk theory, it's not that risky conditions ABC are met—conditions 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. Using evaluative language together in figuring out how to act is such an overwhelmingly important business that we do not allow people 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.

1. I have my doubts that they presupposed it for 'water', as opposed to wanting the extension to be as natural-kindy as possible. If for 'water', why not also 'earth' and 'air'?

18

The commitment comes at a price: a certain kind of anti-dogmatism. The readier we are to claim conceptual authority for our own moral 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, far from 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 disturbing. 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. (p. 152)

I would have thought it was *part* of 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 the term. 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.[1] 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.[2]

3. Colour

A couple of words about Jackson's defence of physicalism against "revelation"—which says that "colour experience is transparent in the sense of revealing the essential nature of colour"—and "unity"—which says that "redness (e.g.) is the property common to all red things".

The objection from unity is that redness-qua-common-to-red-things will have to be a *disjunction* of microphysical surface properties; but disjunctive properties can't be causes; so, given the role of causation in representation, redness is not the property presented in red experience. Jackson replies that disjunctive properties *can* be causes provided they're not *excessively* disjunctive.

Someone might ask: what led us to admit that redness was disjunctive in the first place? That redness is (up to necessary equivalence) a disjunction of other properties? No, for that much can be said of *every* property. That it is a disjunction of *natural* properties? No, because natural (and so presumably

---

1. Tyler Burge, 'Intellectual Norms and the Foundations of Mind', *Journal of Philosophy*, vol. lxxxiii (1986), pp. 697–720.
2. I wish I had more to say about how the envisaged descriptive openness relates to garden-variety vagueness. They are certainly not the same, since vagueness is tolerant of brute, 'no-fault', disagreement, whereas disagreements about what is right are felt—or at least hoped—to reflect a mistake on one side or the other.

non-disjunctive) properties can be that too: charge is the disjunction of posit-ive charge with negative. A better idea (borrowed from Lewis) is that

> a property is disjunctive to the extent that it is a disjunction of properties than which it is much less natural.

The question now is: why think that redness (qua common to red things) *is* much less natural than its microphysical disjuncts? One reason would be that they are physical, and redness is not physical. (This would seem to follow from the discussion of physicality on pp. 6–7.) I don't know whether this is the reason Jackson would give, but if so he may be assuming a stronger form of physicalism than he officially espouses. This stronger physicalism maintains not just that the physical story is *complete*—all other truths are entailed by it— but that it is (far and away) the most *natural*—all other taxonomies introduce an element of the arbitrary.

I don't say that this stronger physicalism is wrong. What I do want to point out is that someone who disagrees with it—call him the taxonomical pluralist— can avoid the unity problem entirely; he can deny that redness is much less natural than its disjuncts, thus denying that it's in any important sense disjunctive.

There is a connection with revelation as well. That colour experience doesn't reveal anything microphysical about redness may or may not mean that it does not reveal the 'essential nature' of redness. It may be of the essence of Jackson's *disjunctive* redness to be built on microphysical disjuncts, but redness conceived as *non*-disjunctive is no more microphysical than, say, the property of being (approximately) round. Both can be implemented in lots of microphysical ways, but since neither is a disjunction—in the relevant sense—of these imple-mentations, it seems gratuitous to write them into the essence. Of course, if *all* necessary properties are written into essence, then colour experience does indeed become 'non-transparent'. But then so does the experience of roundness.

What about secondary-seeming qualities other than colour? These raise an interesting question for Jackson's account. Colours are intrinsic; in particular, a red thing cannot cease to be red just through changes in the typical human experience of it. But, as Jonathan Bennett long ago argued, phenol-thio-urea *would* cease to be bitter if the typical human experience of *it* were to change. Conclusion: bitterness is extrinsic and so not in the relevant sense primary. Strangely, though, it seems just as much a subject-determining platitude about bitterness as redness that it is "the property of objects putatively presented . . . when those objects [taste bitter]" (p. 89). If the platitude is enough to make redness primary, how can bitterness fit the platitude while remaining non-primary?

4. Necessity

Not many eyebrows will be raised by Jackson's view that metaphysics is committed to 'entry by entailment' theses. The hubbub concerns his claim that (if physicalism is true), physicalistic statement P necessitates mentalistic statement M only if M is *a priori* necessitated by P & P*. He argues as follows:

20

contextual information—by virtue of telling us in principle the [C-]pro-positions expressed by [P and M]—enables us to move a priori from the [physical] way things are to the [mental] way they are. But if physicalism is true, all the information needed to yield the propositions being expressed about what the actual world is like in various physical sentences can be given in physical terms [P*], for the actual context is givable in physical terms according to physicalism. (p. 83)

An immediate puzzle is that Jackson is attempting here to ground an a priori entailment in *C*-propositions, when his official account has it that *X* a priori entails *Y* iff a certain *A*-proposition holds in all worlds, viz. the A-proposition expressed by 'if *X* then *Y*'. To square his apriority claim with the official account, he needs to show that the result the speaker allegedly *can* establish,

(i) assuming that this is a *P**-world, all worlds that are *P* are also *M*

can be parlayed into the result that (officially) makes for an priori entailment, viz.

(ii) whatever world this may be, if *P*&*P** then *M*.

But this is easily done. From (i) we see that if this is a *P**-world, then if it is *P* it is *M*. Clearly then if this is a *P**&*P*-world, it is also an *M*-world—which is just what (ii) says. The problem lies more with (i). How exactly does the speaker establish it?

If Jackson is right about what it is to understand, then the speaker knows (based on her knowledge of context) what it takes for a world to be *P*, what to be *M*. Why must she therefore know that the one set of worlds is a subset of the other? The result would indeed follow if

(*) to know what an *F* was, and what a *G*, you had to know the inclusion relations among the *F*s and the *G*s.

I don't deny that there's a sense of 'knowing-which' on which (*) is true. The problem is that it's a 'high' sense; it sets the bar higher than ordinary users set it. Speaking in accordance with (*), even trained mathematicians don't know what an even number is, and/or a sum of two primes. (If they did, they'd be able to figure out the truth-value of Goldbach's conjecture.) The question for us is whether understanding a sentence *S*, and knowing the context, confers knowledge in a *high* sense of what an *S*-world is, or knowledge in a more *ordinary* sense?

It is true that to know what an *S*-world is, I need a way of picking out the *S*-worlds in thought; and not any old way will do. But this is a far cry from (*). I know what a 'there is pain'-world is by knowing that it is a world in which there is pain. I know what a 'things are physically thus-and-so'-world is by knowing that it's a world in which matters are physically thus-and-so—and here I might be able to reel off some specific physical requirements. Obviously though to know in *these* sorts of ways what the *P*- and *M*-worlds are does not put me in a position to tell whether *M* is true in every *P*-world, even if in fact it is.

One response would be to insist that understanding a sentence is a matter of knowing which set of worlds it expresses in a *special canonical way*: a way that better responds to what worlds in their innermost nature are. Since the physicalist

21

thinks that worlds are in their innermost nature physical, she will presumably insist on a *physical* specification. But then it can't be that speakers 'miss' the fact that any world physically like ours is a pain-world simply through failing to think of the pain-worlds in physical terms. Thinking of them in physical terms is a condition of understanding, and we are talking about speakers who understand.

The suggestion is that *if physicalism is true*, then to understand S one must be able to decide (i) on the basis of *physical* information (ii) how to make the cut between S- and non-S-worlds in *physical* terms. (If physicalism is true, then understanding is 'physical' understanding.) This plugs the gap in Jackson's argument, and his conclusion is reinstated. Whatever physical premises necessitate at all, an expanded set of physical premises conceptually necessitates. Merely to understand the sentences is to appreciate their truth-relations.

But this is a result the physicalist can happily accept. It would be one thing if a *normal* understanding of 'things are physically like so' and of 'there is pain' sufficed for knowledge of the conditional; this really would support the claim that physics a priori entails pain. That a *physical* understanding of the same sentences should have this result is not surprising at all. A physical understanding of 'there is pain' is *by definition* an ability to tell whether worlds presented in physical terms do or do not contain pain. The only physicalist who should be bothered by the refurbished argument is the one who thinks ordinary understanding is *physical* understanding as defined by (i) and (ii).

Everything here goes back to the idea that the physicalist will insist on a physical specification of the verifying worlds. Why should she? Physicalism was supposed to be an ontological theory, not a theory of understanding. This distinction is trampled on when understanding is equated with canonical grasp of truth conditions. It now becomes a 'consequence' of physicalism that typical speakers, to the extent that they can't deduce pain from physics, don't understand 'there is pain'! The physicalist presumably finds this as bizarre as anyone else.

But suppose that worry is waived. Suppose we have for each possible world w a complete physical specification P_w in physical terms; and suppose that 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* the case, then it *would* have been that S had it been that P_w.

The problem is as follows. One cannot decide the truth-value of arbitrary statements—'there is pain', for example—with respect to *physically* specified worlds without asking what it would be *reasonable to say* about these worlds. What is the best explanation, for example, of the fact that there is weeping and wailing and gnashing of teeth? That the weepers are in *pain*, or that they are in such and such brain states but feel nothing? If you and I have different ideas about this—because of our larger theoretical and practical projects, how anxious we are to avoid multiplying entities, how physicalistic we are, etc.— we may be expected to assign different truth-values to S_{vw}. But then, just because of background methodological/doctrinal differences, we wind up meaning different things by 'there is pain'!

22

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-fibre-firings*. You decide for ontological economy reasons that 'pain' and 'c-fibre 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 modal intuition and psychological autonomy considerations 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, 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—I don't say it isn't *true*, just that it isn't *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.

5. Summing up

Quine speaks somewhere of the tendency among philosophers "to seek the gist of every statement in objects that it is about". Jackson does something related; he seeks the gist of every (truth-apt) statement in worlds that we would/should regard as verifying it. Add to this that understanding is 'getting' the gist, and Jackson's main claim—that, physicalism granted, merely to *understand* a statement S is to be in a position to deduce it from physical statements —more or less falls out.

The question I have been pressing is: why should we chalk it up to *conceptual* competence that a speaker regards 'there is pain' ('that was a decent thing to do', 'he's being illogical') as an appropriate thing to say about a canonically described world? Some of the credit surely goes to the speaker's sense of what is or is not a sensible thing to *think* in light of the available evidence. If that is right, then the physical/mental (etc.) conditionals that Jackson would call conceptual truths, or truths of reason, might better be described as truths of *reasonableness*. A truth of reason would compel the assent of every rational person who speaks the language; the conditionals just record the epistemic intuitions of people with epistemic intuitions like mine.[1]

THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY   STEPHEN YABLO

REPLY TO YABLO: WHAT DO WE COMMUNICATE WHEN WE USE ETHICAL TERMS?

1. One way to respond to an interesting review is to chase down as many points as possible in the space allowed. Another way is to address an underlying difference in approach between author and reviewer. In this case, I decided that the second course was the best. It seems to me that a number of Stephen Yablo's more particular disagreements stem from a big, if largely unstated, difference between us over language. The difference surfaces most in his discussion of what I say about ethics, so I will set my discussion in that context. My discussion draws on many conversations with Michael Smith and Philip Pettit, but I have to take responsibility.

2. What is it to understand a language and, in particular, to understand certain words in a language? Part of the answer is that understanding allows you to know what people who use the language and the words are saying about how things are. Languages like English or Russian or Japanese are, in large part, systems of representation that those who understand them use to represent to others who also understand them how things are. That is why understanding a language can be very useful—as we are all forcefully reminded when we travel to countries whose language we do not understand. In particular, someone who understands the word 'right' in English knows what is being said about how things are—how things are being represented as being—by the use of the word by speakers of English when they say, for example, that 'X is right'.

   In putting the matter this way, I am assuming cognitivism: the view that ethical language makes claims about how things are, and that an ethical sentence is true just if things are as the sentence represents or claims them to be. I will assume cognitivism in what follows, as I did in the book. In fact, I think that much of what I say below concerning various possible answers to what is being claimed when X is said to be right, could be said equally to expressivists concerning various possible answers to what is being expressed about certain attitudes. Expressivists allow that we understand ethical language and, presumably, hold that to understand it is to know what is being expressed by its use. But I will not pursue the issue here. (Nor I will pursue the question as to whether expressivists have told us what expressing an attitude comes to in a way that suitably separates it from reporting the attitude.)

   I now take a critical look at some answers cognitivists might give to what we are saying about how things are when we say that X is right, along with some answers perhaps better thought of as critical comments on the asking of the question. At the end of my survey, I will be able to say why I am unmoved by Yablo's main worry about my reductionist position in ethics. He is far from alone in having this worry, so I hope that what I have to say will be found useful by a number of readers.

3. ''X is right' ascribes an indefinable, sui generis property to X. Hence, what we are saying about how things are when we use the sentence is that X has this sui generis property.'

24

I take it very few nowadays accept this answer. The problems for Moore's style of non-naturalism are widely appreciated. It is metaphysically implausible; it makes it obscure why we should care about goodness; and it is very hard to square with the a priori supervenience of the ethical on the descriptive. I put it on the table partly to set it aside, and partly to allow us to note later that certain answers to our question that some do accept are uncomfortably like this answer.

4. 'We have no idea what property we ascribe to X when we say that X is right.'

But, then, why do we, by and large, seek to promote that which we are told is right, or judge for ourselves to be right? If we have no idea what property we are ascribing, our response to hearing someone say that X is right should be the same as hearing them say that it is flub.

5. 'We do not know exactly what property we ascribe to X when we say that X is right. We do, though, know that we are ascribing some property or other that ought to be promoted or ought to induce such and such attitudes or. . . .'

But to know that we are ascribing some property or other that ought to be promoted or . . . *is* to know what property is being ascribed. The property would, on the suggestion mooted, be that of having some property that ought to be promoted or. . . . The same point applies to any suggestion along the lines that we do not know what property we are ascribing to X but do know that the property has such and such characteristics. The property being ascribed in this case is having some property which has such and such characteristics.

6. 'What we learn from Kripke and Putnam is that the question of what a sentence means is an a posteriori matter. It is a mistake to assume that we should be able to answer the question as to what 'X is right' ascribes here and now simply on the basis of the fact that we understand the sentence.'

No doubt there is some meaning we might give to the word 'meaning' such that understanding a word does not go along with grasping its meaning. Indeed, if what I say in the book is right, a word's C-intension is a case in point. Nevertheless, we should resist the suggestion that we do not know what we are saying when we say that something is right. To embrace this suggestion would be to make a nonsense of the role of assertions that something is right in moral discussions; it would mean that we literally would not know what was being talked about. Why then do we care so much? Somehow or other, we understand the language of ethics sufficiently to know what is up for debate when we ask about what is good, bad, right, ought to be done, etc. Otherwise, ethical debate would be a nonsense. Of course, some have held something like that. But they are ethical anarchists or extreme noncognitivists. Expressivists like Allan Gibbard or Simon Blackburn do not think that moral debate is a nonsense. And J.L. Mackie's view is best thought of as that ethical debate has a subject matter; it is just that it is nowhere instantiated. If Mackie thought that we had no idea of what we are claiming about how things are when we say that X is right, he could hardly have argued that what we are claiming is never in fact the case.

7. 'What we are saying when we say that X is right is that the word 'right' applies to X.'

We do not seek to do what is right because a word applies to it. We seek to do what is right because of the property (in the wide sense that counts relations, including relations to our attitudes, and moderately disjunctive ways things might be, as properties) the word applies in virtue of.

8. 'There is no answer to the question of what property the word 'right' ascribes other than that it ascribes the property it ascribes. This is not to say that the property is sui generis. It is not to say that it is the property of having 'right' applying to it. It is not to say that we do not know what the property is. It is to say that we cannot say in words any more than that it is the property we ascribe with the word 'right'. No doubt we can introduce a synonym, or use a word in another language that means the same, but we cannot say anything interesting in other words.'

I grant that we should sometimes allow that we have reached the end of the road as far as what can be said about the property ascribed by some word, in the sense of what can be said which is of interest. However, as I emphasise in the book, this is a difficult position to maintain in the case of the ethical terms in English.

The ethical a priori supervenes on the descriptive. It seems to be part of what is involved in understanding ethical terms that one grasps that they mark distinctions among the descriptive. This is why the supervenience is a priori. But then coming to master the ethical vocabulary must be a matter of coming to latch onto the right distinctions among the descriptive, a matter of spotting the patterns in the descriptive.

Indeed, it is hard to see how else we learned to use moral terms. Obviously, we do not distribute them at random among descriptively specified situations, and, barring Moorean properties to be the 'unifiers' for the ethical terms, there seems no choice but to allow that we distribute them by latching onto a pattern, in the sense of that which is projectible, among the descriptively given. Of course, the patterns need not be solely in the situations as such; they may, in part, reside in the way we relate to the situations. But note that this relation itself must be descriptively specifiable, otherwise we are playing 'pass the parcel'.

But now the question is whether or not we know, implicitly, the descriptive patterns we are ascribing. If we do not, this simply raises, over again, the issue as to whether we know what we are saying when we use the ethical vocabulary; and if we don't know what we are saying, why do we care so much whether something is right or bad or . . . ? But if we do know, how come it is impossible to say what the pattern is?

In discussion, some object at this point that we can latch onto a pattern simply as the pattern that induces a certain response in ideal cases; we have a kind of Gestalt response. But that is all we can say. However, if this were the case, if it really were impossible to say more, then we have the pattern—it is delivering the Gestalt response in ideal cases. But in fact it is very implausible that we cannot dissect out from the situations that generate the responses, the pattern which underlies them. (This is the sense in which the knowledge is implicit.)

26

When we debate the rightness or wrongness of various situations, it is false that we have no idea of which features matter, and which do not, to their moral status. The situation is much more like judgements of grammaticality. Some sentences 'look wrong' and often we lack the terms needed to say why they are ungrammatical, but we know more or less where the trouble is. You do not have to know about verb agreement *under that name* to know that 'She are happy' is ungrammatical, and that the trouble can be fixed by replacing 'are' by 'is', or 'She' by 'They'.

9. I say essentially this about the term 'rational' in the book. There is a pattern to our judgements of rationality—we do not distribute the term 'rational' capriciously or at random—and to analyse rationality is to articulate the implicitly grasped pattern. Yablo finds my view 'surprisingly optimistic'. But I think (or I guess) that he has not considered the unpleasant nature of the alternatives, as outlined above.

Also, I think that the case of grammaticality is an encouragement to optimism. Opponents of conceptual analysis often make much of the fact that we fans of conceptual analysis have so few convincing offerings, in the sense of completed articles, on display. Our writings are full of sketches, with the exception of some nice simple cases, often taken from mathematics. (I surmise this is what Yablo had in mind when he penned 'optimistic'.) But no-one can complete 'X is grammatical iff X is . . .' for every natural language. It is not even clear that anyone can do it for English; and even if some can, the vast majority of English speakers cannot. But no-one takes seriously the proposition that there is no pattern to the sentences we judge grammatical. Moreover, this pattern is constitutive of what it is to be grammatical; grammaticality can be reductively analysed. No one is a Moorean about grammaticality, or thinks that all we can say is that grammatical sentences have in common that the word 'grammatical' applies to them. Moreover, no-one thinks that it is impossible in principle, however hard it may be in practice, to write down the pattern. In fact, we know that existing grammar books contain pretty good fragments of the full analysis, and maybe, for all I know, some have completed the job.

The case of grammaticality shows that it can be rational to be confident in the existence of a reductive analysis even when you are not yourself able to write it down.

10. Of course, it is more than fair to ask a supporter of conceptual analysis to give a clear indication of how we might approach the task of finding an analysis in any particular case. After all, we know roughly how to approach the task of writing down the conditions under which a sentence is grammatical —do what writers of grammar texts do. My answer in the book as to how we might approach the task of analysing 'X is right' is, of course, moral functionalism: the account that analyses 'X is right' in terms of its place in completed mature folk morality. That, according to me, is the pattern we learn to latch onto when we master the word 'right'.

27

11. Now, finally, I can address Yablo's concern that I make it too easy for moral disagreement to turn out to be talking past one another, that I fail to heed sufficiently the open question argument. Here is what he says: "I would have thought it was *part* of 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."

However, it is not a priori that we all mean the same by words like 'right' and 'reasonable'. If we refuse to attach 'right' to any property, what makes it true that we all mean the same by the word, and so count as potentially in genuine disagreement? Would it be better to attach 'right' to no property at all? But then *what* would we be disagreeing about? We certainly could not be disagreeing about the distribution of properties. This is why, in this context, I find Yablo's earlier claim that "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" so puzzling. He argues that we need to make this commitment to ensure that ethical disagreement (and agreement) is genuine, and not an example of talking past each other. But then it follows that if it we are using the terms for different properties, we *are* talking past each other. It cannot simultaneously be the case that the properties we use the ethical terms for need to be the same for ethical agreement and disagreement to be real, together with its being the case that any amount of difference over the properties we use the words for is consistent with the agreements and disagreements being real. Perhaps Yablo will reply that our primary commitment is that 'right' stands for the same property in our mouths but not the same descriptive property. I can see how Moore could say this but not how anyone who has turned their back on Mooreanism in ethics and who has taken to heart the lesson of the a priori supervenience of the ethical on the descriptive could.

12. What has gone wrong? I speculate, first, that there may be a confusion over what is meant by the property attached to 'right', and, second, that Yablo (and the many who have pressed a similar worry to me in discussion) may be forgetting that what our words attach to are the products of our agreements to use words in certain ways. On the moral functionalist picture, there is both the property that plays the rightness role (in mature folk morality) and the property of having the property that plays the role. It is the second, and not the first, which is the property ascribed by the use of the word 'right'. This is why it is the possibility that different people locate the moral terms in different networks in the idealised limit, which I describe as cases where there is a difference in meaning.

Secondly, while I grant the intuitive appeal of the open question argument, I think it is vital to remember that our words are our words. Somehow or other, an interaction between our agreed-upon usage patterns and our world

determines what our words stand for. Moreover, it is a priori that the correct usage patterns supervene on descriptive nature. These two points together mean that it cannot be the case that we can keep the meaning and reference of the terms of rational and moral appraisal fixed, no matter how we vary the descriptively specified facts (remembering that the latter include facts to do with places in networks, as well as what fills those places) they are supposed to apply to. In the book, as Yablo reports, I suggest that the contrary view may be a hangover of platonism. I would now add that it may also be a hangover of a 'super glue' view of meaning and reference.

13. Finally, let me say for the record that Yablo is right that I regard the a priori supervenience of the ethical on the descriptive as applying equally to the supervenience of the rational and reasonable on the descriptive. Hence, for me, the theory of ethical properties he mentions in passing that identifies them with descriptive properties that occupy certain roles specified in terms of what is reasonable, where the latter is not identified with any descriptive property, is the worst of both worlds. Better to go reductive across the board, or nowhere. In the same way, I think you should go cognitivist or noncognitivist across the board. I do not accept the arguments for noncognitivism in ethics, but if I did, I cannot see how I could resist them about rationality. The arguments in the two cases are so similar (as, for instance, Gibbard and Blackburn emphasise).

THE AUSTRALIAN NATIONAL UNIVERSITY   FRANK JACKSON