Georges Rey, Philosophy, Univ of Maryland, College Park, MD 20742;
georey@earthlink.net; home: (202)-331-9267; cell: (202)-680-8420 Mon 2 Oct 06

6-7K expected; 7.8K w/fns and refs; **7**K w/fns but not refs; 6K w/out fns or refs.

Very Final Draft **Resisting Normativism in Psychology**

for *Current Debates in Philosophy of Mind*, ed. by J. Cohen and B. McLaughlin; Oxford: Blackwell

1. The Background Normativity Claims

"Intentional content," as I understand it, is whatever serves as the object of "propositional" attitude verbs, such as "think," "judge," "represent," "prefer" (whether or not these objects are "propositions"). These verbs are standardly used to pick out the intentional states invoked to explain the states and behavior of people and many animals. I shall take the "normativity of the intentional," or "Normativism," to be the claim that any adequate theory of intentional states involves considerations of *value* not essentially involved in the natural sciences. Thus, according to Normativism, whether or not someone *thinks* that fish sleep, or even can *represent* fish at all, depends upon making a judgment about the person's goodness or rationality, of a sort that would not be involved in merely determining whether or not fish in fact sleep.[1]

Normativism has influenced a great deal of philosophy of mind for at least the last fifty years, its roots stretching back even further. One source is a concern with the links between freedom, morality and rationality that can be traced at least back to Kant. Another is the distinction many have felt between natural scientific ("erklären") and historical "empathic" ("verstehen") explanation, as it emerged in the work of Dilthey (1894/1977), Weber (1913) and (for many) Wittgenstein (1953). Most importantly for purposes here, it was recruited by Quine, as "the principle of charity," as a way of accounting for intentional ascription in the wake of his scepticism about the analytic/synthetic distinction, and his related thesis of the indeterminacy of translation. This latter thesis was the basis on which Davidson (1970/80:222) claimed that, in ascribing propositional attitudes to someone:

> We must work out a theory of what he means, thus simultaneously giving content to his attitudes and to his words. In our need to make him make sense, we will try for a theory that finds him consistent, a believer of truths, and a lover of the good. ... the constitutive ideal of rationality partly controls each phase in the evolution of what must be an evolving theory. –(1970/80:222-3)

Davidson took these constraints to mark a fundamental difference between psychological and the physical sciences:

> It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental

1. I join Rosen (2001:611) in not trying to provide a general characterization of the normative, but merely relying on cases for the issues at hand. I shall not, though, discuss here the views of Brandom's that Rosen seems to me to adequately discuss along lines compatible with my discussion here.

> phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual. --(p222)

In keeping with his sympathy with Quine, Davidson (1978/82:301) goes on to claim that "these matters bear directly on the...question how scientific a science of the mental can be,"[2] a view echoed by Daniel Dennett:

> deciding on the basis of available evidence that something is (or may be treated as) an intentional system permits predictions having a normative or logical basis rather than an empirical one... Intentional theory is vacuous as psychology because it presupposes and does not explain rationality or intelligence. --Dennett (1978:13,15)

In earlier work (Rey 1994, 1997, 2002), I argued that these claims are unwarranted. They presuppose our enjoying a sufficiently adequate understanding of serious psychology –as well as of rational norms– to reasonably make them, but, even if they are restricted to ordinary "folk" psychology, they are based upon a small diet of examples that don't do justice to folk wisdom. I'll summarize some of these arguments in §2 of what follows, indicating how variable and probably indeterminate are existing conceptions of rational norms and their application to actual psychology (§2.1); how intentional explanation tolerates a great deal of irrationality and frequently doesn't involve "reason" at all (§2.2); and how, in any case, serious judgment about these matters is premature, based upon an implausible "superficialism" that would not be invoked in any other domain, and is in fact at odds with the folk psychology it tries to enshrine (§2.3).

In §3 I'll turn to Ralph Wedgwood's recent defenses of Normativism. Although he has admirably resisted some of the less convincing arguments for it, his new arguments seem to me open to some of the same objections. Ironically enough, I'll argue in §4, they bring us back to the very issues about the analytic that led philosophers like Quine to Normativism in the first place. At any rate, Wedgwood's frequent appeals to mere intuitions about counterfactual cases would, I argue, be better served by a substantive theory of analytic conceptual connections, instead of an effort to base psychology upon norms that are ill-defined, open to subjective variation, and risk an invidious distinction between psychology and the natural sciences.

2. Norms and Psychology[3]

Normative reasons, of course, have their place in ethics, decision theory and epistemology: what one *ought* to believe or do, given a set of values and evidence of how things are. The question is whether they are essential to *psychology*. There seem to me a number of reasons to think they aren't.

---

2. Which is how Davidson is standardly understood. Jaegwon Kim (1985/93), for example, writes about how Davidson "joins a small but influential group of philosophers who have taken a dim view of the scientific prospects of psychology. ... [His] argument has far-reaching implications regarding some basic issues about the nature of mind, ... and points to a conception of the mental that I find both intriguing and appealing" (pp194-6).

3. Portions of this section have appeared previously in my (1997) and (2002).

2.1 Which Norms?

In the first place, there really isn't any serious agreement about what the relevant norms might be. To be sure, there's been considerable progress in understanding *deduction*, and, to a lesser extent, with respect to certain highly idealized theories of normative "decision." But the various problems of characterizing "good" inductions, abductions, *actual* decisions and the rational "updating" of one's store of beliefs in the light of further actions and experience (the so-called "frame" problem) are notoriously more difficult. As things stand here, there are merely hosts of intuitions about particular cases –some good, some questionable– that have yet to admit of a general characterization, much less one that is uniform across different people and intelligent animals.

Moreover, as Gilbert Harman (1986, 1995/99) has repeatedly stressed, it's crucial to distinguish an interest in "logic" as a theory of rational *relations among propositions* from a theory of rational *inference* as a theory about what it is rational for someone to *believe*: the former enjoys a stability and objectivity that may never be available for the latter. Why, after all, should one person's high standard of confirmation (in a particular context?) be any more or less rational than another's lower one? Is it better to optimize than to satisfice? How are we to assess differences in people's attitudes towards risk, e.g., towards being exactly right in a few cases or approximately right in many; or towards making up their mind under the pressure of time? Is it more rational to undertake risk (however that is to be defined) by maximizing average expected utility or by minimizing the likelihood of a worst outcome? What is the role of second- (and n–) order reasonings (thoughts about one's thoughts) to the usual first-order ones? Reflection on the difficulties of understanding the world has led many to tolerate the likely falsehoods of idealized models or promising research programs, and sometimes even outright contradictions (think of the standard philosophical paradoxes).[4]

Even should there emerge a unified and determinate account of rational norms, the question would arise as to how it should be applied to intentional creatures. Should it be applied only to their "whole" minds, or also to the specialized modules that seem to underlie, e.g., speech, perception, navigation and motor control? In the case of people, is it really *the beliefs* they are willing to *avow on reflection* that are the appropriate domain of rationality, or is it rather their "gut" reactions, or their underlying *judgments,* whether or not they can avow or are even remotely conscious of them?[5] Remembering the absent-minded, perhaps the norms should be applied only to what a person *notices* (consciously or unconsciously) in a sufficiently brief window of time. But then should our short-term memory and processing limitations be regarded as intrinsic to our reason, or should we nonetheless idealize ourselves as rational gods, closed under deduction (but then who would need any mental *processes*?)[6] Even if there turn out to be facts of the matter about all these difficult issues, it's hard to

---

4. See Slote (1989) regarding satisfycing, Elster (1979, 1983) and even Davidson (1978/82:305) himself regarding higher-order attitudes; Cartwright (1983) on the likely role of falsehoods in practicing science, and Pollock (1991), Nozick (1993), Harman (1995/99)), and Frisch (2005) regarding at least provisional toleration of inconsistencies.

5. See Stich (1983), Rey (1988), Moran (2001) for discussion of some of these complexities. Wilson (2002) discusses a variety of experimental data that suggest introspective reflection is often less reliable about revealing one's actual attitudes than are spontaneous "gut" reactions (which rather gives one pause about the status of "philosophical intuitions").

6. See, e.g., Cherniak (1986) and Harman (1986, 1995/99) for discussion.

see why someone whose psychology (appeared to?) not to satisfy them in the right way would *ipso facto* fail to have the intentional states the theorist might otherwise have good reason to ascribe.

2.2 Non-Rationality

Whatever norms do emerge, it's also hard to see that they play the crucial role on which Normativists insist. While it's certainly true that we sometimes explain someone's action by citing reasons[7] that recommend it, as when we explain a good piece of reasoning or a successful bet, sometimes people's reasons are *terrible*, as, for example, when they persist in standard fallacies, e.g., expecting a coin to come up heads after a run of tails. Daniel Kahneman just got a Nobel prize for his and Amos Tversky's extensive research on the surprisingly extensive range of errors of this sort to which humans seem naturally prone.[8] And philosophers themselves make their living discussing incoherencies that arise for our unreflective beliefs about truth, knowledge, motion, freedom, and the existence of the external world. In general, people fail to see some of the most immediate consequences of their thought, are often inconsistent, say the most bizarre things regarding logic, knowledge, statistics, religion, history, psychology, biology, cosmology and physics; they want all sorts of things they don't need, have positively alarming views about the good, and regularly disregard it when even minor interests compete. Whatever the idealized norms of rationality turn out to be, they would seem to be only some among many of the determinants of those roles, alongside limitations on perception, memory, attention, the influence of tradition, habits and desires, the sway of charismatic authorities and the sudden impulses of a moment. Awareness of these complexities doesn't in the slightest undermine intentional ascription: to the contrary, it regularly refines it.[9]

There's also a lot more to intentionality than reason alone. Wittgenstein and Austin emphasized how philosophy often suffers from too restricted a diet of examples, and this seems strikingly true of the Normativist's standard examples. Again, while it's true that much human thought and action is explained "rationally," a lot of it is not: intending to wriggle a finger brings about its wriggling whether or not one has a good reason to do so. Some actions, what Kent Bach (1978:363) has called "minimal actions," such as scratching an itch, doodling or automatically tying a shoelace, don't involve rational deliberation, although they may still involve rich representations, full of intentional content (about the location of the itch, the position of the laces).

Even many fully intentional actions are often performed without reasons. Arguing against Davidson's (1978/82:292) claim that an intentional action must be caused by a rationalizing belief-desire pair, Rosalind Hursthouse (1991) discusses a significant range of what she calls "arational actions" e.g.,

---

7. I leave aside the perfectly good use of "reason" that's just a paraphrase of "cause" or "causally explain," as in "the reason that Humpty Dumpty fell is that he was pushed." Throughout this discussion, one needs to be constantly alert to non-normative uses of normative terms, as in, e.g., "According to the *ideal* gas laws, the *right* temperature should have been higher; indeed, it *ought* to have risen quite rapidly. See my (1997:ch 10) for further discussion.

8. See Kahneman and Tversky (2000), Kahneman (2002/3) and Stein (1996) for extensive discussion. One might also merely reflect on the standard fallacies catalogued in elementary logic books, which, of course, have nothing really in common other than being common human errors.

9. In a nice reply to Dennett's insistence on optimal rationality, Fodor (1981) pointed out that in playing chess, for example, part and parcel of one's "intentional stance" may be the hypothesis that one's opponent is "a sucker for a knight fork." (p108).

kicking a door in anger, jumping up and down in glee, “rumpling the hair of, or generally messing up the person or animal one loves” (p58). She reasonably argues that such actions are not ordinarily performed to achieve some *end*. Although they may in fact *express* one's feelings, they are typically not performed *in order* to do so --indeed, their being so performed can undermine their genuineness as actions of the relevant sort. She draws attention also to *symbolic* actions, such as gouging holes in the eyes of a photo of one’s rival in love (p59-60), which, again, seem not to involve any serious rational plan (or, if they did, would involve pretty irrational use/mention confusions!).

And, of course, in addition to all these cases, there are all sorts of effects of intentional states that are not *actions* at all: startle at the unexpected,[10] laughter at jokes, tears at bad news, ulcers due to stress, trembling at the thought of speaking in court. Again (cf. fn 7), the joke may be the *explanatory* “reason” for the laughter, the stress the “reason” for the ulcer, without a trace of normativity (it’s not *irrational* to fail to laugh at a joke or develop an ulcer!).

Bearing the full range of ordinary cases in mind, it should be plain that intentional states are not invoked merely to rationalize actions, but to explain an extremely wide range of rational, irrational, arational, and brute physical actions and events simply by subsuming them in the standard way under conceptions about how people ordinarily work. The evidential situation seems the same as for the other sciences: any of the explicanda could serve as evidence of the underlying states. Someone's belief that a friend has died not only explains her making funeral arrangements, but also her tears, grief, exclamations, beating her breast, inattention, sleeping late, placing the friend's photo in a special box, etc., from each of which we might infer that belief. In this way, we would certainly appear to (in Davidson's phrase) “triangulate” onto such internal states as the common causes of these events, but in a way no different from how a doctor triangulates upon some malady from a set of symptoms, or a geologist upon some subterranean process from the age and shape of a rock. In view of the wide diversity of especially non-rational evidence, it is hard to see why rational norms need play any special role.

Well, it’s argued, all these complexities and examples of non-rationality are supposed to be explained away by the fact that "disagreement and agreement alike are intelligible only against a back-ground of massive agreement" (Davidson 1973/1984:137), and that consequently the norms must hold on the whole: “to the extent that we fail to discover a coherent and plausible pattern in the attitudes and actions of others we simply forgo the chance of treating them as persons” (1970/80:221-2). But, especially in view of the diverse examples we’ve noted, why should we think so? The only argument of Davidson’s that I can find is his appeal to the “holistic character of the mental”:

> The meaning of a sentence...is not an item that can be attached to it in isolation from its fellows. We cannot intelligibly attribute the thought that a piece of ice is melting to someone who does not have many true beliefs about the nature of ice, its physical properties connected with water, cold, solidity, and so forth. ... The clarity and cogency of our attributions of attitude, motive and belief are proportionate, then, to the extent to which we find others consistent and correct.

---

10. Think of how much is presently being learnt about the intentional states of pre-linguistic infants from merely observation of their involuntary startle responses; see Mehler and Depoux 1994..

–(Davidson 1978/82:302-3)[11]

But Normativism hardly follows from or plausibly explains cases like that of attributing [ice]. Even if it were true that attributing a concept of x entails attributing "many" true beliefs about x (although didn't Kripke (1972/80) famously refute such a claim?) still this doesn't entail a proportionate truth and consistency on the whole: maybe there are many true beliefs –and maybe *even more* false and inconsistent ones! After all, who's counting? And, again, with respect to precisely which of the diversity of belief-like attitudes (avowing, noticing, unconsciously judging at a time) that a typical human mind enjoys?

2.2 Superficialism

One sometimes gets the impression that Normativists presume our relation to psychology to be like the relation Einstein enjoyed in 1905 to physics, permitting sweeping claims about the character of an explanatory realm and the determinacy of certain parameters within it.[12] But surely, made explicit, the comparison is embarrassing. The physics of 1900 was a well-worked out, well-confirmed (even if still mistaken) theory of the nomological structure of the world, on the basis of which Einstein could say with substantial empirical authority that certain distinctions had no basis in the world. There is nothing remotely approaching such a theory yet in psychology.

Of course, many think there is, in particular, that our "folk psychology" is really all one needs. I've discussed elsewhere what seems to me this "superficialism" that has dominated the philosophy of mind, not only in the form of overt behaviorism, but in more recent defenses of ordinary thought and talk.[13] Originally, it was a piece of the verificationist response to scepticism generally, but it oddly survives in the philosophy of mind in a way that would (at best) be found quaint in physics or biology. Thus, Dennett (1991:461) defends an "Urbane Verificationism" according to which all psychological distinctions should be available in ordinary behavior; Colin McGinn (1991:132-3) "doubt[s] that our naive psychological classifications could be overturned ... under pressure from any sort of scientific theory of the mental"; and Jennifer Hornsby (1997:3-4) writes that "we ought not to assume at the outset that the basis of our every-day understanding of one another is susceptible of correction and refinement by experts in some specialist field where empirical considerations of some non-commonsensical kind can be brought to bear." I suspect it is something like this sort of confidence in at least the *perspective* of our folk thought about mind that leads Davidson or Dennett to their sweeping pronouncements. Indeed, when Steve Stich (1983) marshals

---

11. Davidson no doubt is relying here on Quine's peculiar combination of confirmation holism and a verificationist theory of meaning (see 1969:80-1, 1986:185) --peculiar, since a more likely response to confirmation holism might well be to abandon verificationism altogether. It's hard to see what else Davidson could summon on behalf of a thesis so extreme, one that, n.b., entails that people –even different stages of the same person– who differ about *any* content differ about *all* (see Fodor and LePore (1992) and Devitt (1996) for discussion)! I think what he really wants here is nothing quite so global, but just the standard intuitions about some claims being analytic, or true "by virtue of meaning," to which Quine forbade him to appeal. I'll return to this point in §4.2.

12. Dennett (1995:532) is quite explicit about the comparison, and it sometimes is cited on behalf of Quine's indeterminacy and inscrutability theses (see, e.g., Quine 1969:48-9).

13. See my (1995) and (1997:197-201, 275-80) for extended discussion. Note that Dennett (1995) cheerfully embraces superficialism, quoting with approval Quine's quotation of the motto of the Sherwin-Williams paint company, "Save the surface and you save all" (p530).

scientific evidence against Dennett's (1978) Normativist claims ("a system's beliefs [and]...desires are those it ought to have given its biological needs," 1987:48-9), Dennett simply replies:

> I would insist, however, that all this empirically obtained lore is laid over a fundamental generative and normative framework that has the features I have described. –Dennett (1987:54)

After all:

> No other view of folk psychology...can explain the fact that we do so well explaining each other's behavior. --(p51).

But I don't see that what success folk psychology enjoys remotely warrants such confidence. Although we are sometimes subtle and insightful, a moment's reflection reveals we know next to nothing about even such basic activities as perception, thinking, reasoning, language, decision making, motor control, not to mention consciousness, creativity, scientific insight or morally responsible action. And, *pace* Dennett, Hornsby and McGinn, we *know* we don't: much of the folk all the time presume that specialized knowledge may well cast light on all those issues, as well as on the nature of personality, intelligence, psychological development, supposed racial and gender differences, and so forth. It seems increasingly apparent that our relation to our minds is pretty much on a par with our relation to our bodies. More specifically, along the lines of present research, our understanding of the minds of people and animals seems like the understanding clever children have of their computers: they know a good deal about their *interaction* with them --they can play games, and get them to do various things– but with only the sketchiest ideas about their internal causal/computational structure.

Given our massive ignorance, Normativism would require some pretty powerful, relatively *a priori* arguments. These are what Ralph Wedgwood tries to provide.

3. Wedgwood's Arguments

3.1 Wedgwood's Claims

In a number of articles and a forthcoming book, Wedgwood has defended the claim that "intentional facts are partially constituted by normative facts" (NI:1). Unlike the superficialists, he adds that he understands the thesis as a "metaphysical claim concerning the essence of intentional states...not a semantic thesis about the meaning of intentional terms or a conceptual thesis about what is built into our concepts of intentional states" (NI:2; see also Ch7:10). As states of real spatio-temporal beings, of course, the crucial claim is that the agent of an intentional state have a *disposition* to accord with rational norms:

> The sort of disposition that a thinker must have, if she is to possess these concepts or be capable of these sorts of attitude, is a disposition to use the concept in ways that the principle in question specifies as rational. (Ch7:18)

Thus, a *belief* is "correct if and only if the content of the belief is true," and it is "rational (in relation to a body of information *I*) just in case *I* makes it sufficiently likely that the content of the belief is true" (Ch7:15). A person actually enjoys the relevant state only if she possesses *ceteris paribus* dispositions to

conform to these objective constraints. For example, someone has a belief with the content [yellow][14] iff she is

> disposed to form a judgment applying the concept "...is yellow" to a perceptually presented object whenever she has a visual experience that represents an object in the relevant way. --(Ch7:18)

And she possesses the concept of a conditional, only if she has "the disposition for rationally accepting instances of inferences like *modus ponens*" (CH7:26).

Of course, as merely *ceteris paribus* dispositions, these proposals are free of some of the implausible commitments of Davidson and Dennett I discussed in §2. Dispositions may persist even if *never* manifested. The force of Wedgwood's constraints is that

> when we do not conform to these very basic requirements of rationality, then the situation was in some way abnormal (the *cetera* were not *paria*), and so there must be some special explanation of what went awry. (Ch7:19)

He claims (fn 16) that so construing the dispositions makes them invulnerable to the kinds of objections raised by Edward Stein (1996) regarding the standard evidence of human foibles in reasoning mentioned earlier.

Confidence regarding this latter issue, however, requires more discussion than Wedgwood provides. Appeals to *ceteris paribus* clauses are well and good, but only so long as there's *independent evidence* that *cetera* and are not *paria*. Specifically in the case of rationality, what would need to be shown is that deviations from rationality are "performance" errors, concealing an underlying competence. But, as Stein (1996:chps 3, 8) is at pains to point out, it's far from obvious that they are. The *persistence* of many foibles despite correction and reflection suggests that some of them are built into at least some sub-systems of the mind. Kahneman (2002/3) himself regards much of his and own and others recent work as supporting the hypothesis that humans work with at least two systems of reasoning, one "intuitive," fast and automatic, the other slow, deliberate and reflective, the second occasionally (although, alas, not reliably) correcting the errors of the first. In order to sustain his Normativism, Wedgwood would need to show that the first system is essentially an interference with the workings of the second, and could not be understood without it. None of the evidence so far seems to me to suggest that this is so. To the contrary, the "intuitive" system may be, evolutionarily, a more primitive one, present without the slower one in many animals (see Wilson (2002:44-45), who cites Reber (1992)).

Wedgwood, himself (see NI:§3, pp4-7), is also unconvinced by many of the standard arguments for Normativism, and relies instead on two further arguments, which I will discuss in turn: an argument from asymmetry (§3.3), and one from defeasibility (§3.4). Before considering them, however, I want to consider an argument that both he and I reject, but is so easily suggested by the terms in which he and many others set up the discussion that it deserves explicit airing from the start.

3.2 The Argument from (In)Correctness

---

14. I refer to concepts by enclosing in square brackets the words that express them, properties by placing the words for them in italics. I would ordinarily distinguish concepts from the natural (or mental) language symbols that express them, but, since Wedgwood doesn't in his discussions, I won't do so here.

The usual discussion begins, as does Wedgwood in NI, "with the assumption that the paradigmatic normative terms are "ought," "should," "right," "wrong,"... "correct," "incorrect," "rational," and `irrational'" (NI:1). If one combines this assumption with the (let us grant) independently plausible assumption that intentional ascription necessarily involves such concepts, then, of course, Normativism would immediately follow.

But, of course, one trouble with so facile an argument is the starting assumption. Although all these quoted terms have a variety of normative uses, they also have non-normative ones (cf. fn 7 above). Thus, one can say such things as, "given how long he was in the water, he ought to be dead by now," or "the planets should move in ellipses," meaning merely that these claims follow from, say, the truth of various laws of physiology and motion, without any serious normative implications. Moreover, as Wedgwood himself acknowledges "some philosophers might argue that "correct" applied to beliefs is just a synonym for "true", and that "rationality" is merely a logical, not a normative concept" (ch7:6).

In any case, as Wedgwood (NI:5) notes Paul Horwich noted, normative implications by themselves don't make a concept itself normative; "killing is *prima facie* wrong; nonetheless one can presumably characterize `x kills y' in entirely non-normative terms" (Horwich 1998:188). We need, that is, to distinguish properties, like *moral* or *beautiful* that are *intrinsically* or *constitutively* normative, from ones, such as *killing*, which are normative due to some *further*, *extrinsic* normative theory. So the question raised by Normativism is whether intentional notions are intrinsically normative. Wedgwood proceeds to provide two arguments that they are.

3.3 The Asymmetry Argument

Wedgwood rightly notes that intentional states cannot "float completely free of our dispositions" (NI:20), where I take him to mean *all* of a person's dispositions to deploy a concept, both internally in relation to other concepts and in relation to the external world (any more restrictive claim would require further argument). He then reasonably argues that, among the dispositions that fix the content of a concept, some must be *basic,* determining which concept it is (NI:26), on which all the other dispositions depend (NI:25). These basic dispositions could, let us suppose, be either rational or irrational. "However," Wedgwood writes:

> it seems to me doubtful that one's possession of a concept can rest on an irrational disposition... The possession of a concept is a cognitive *power* or *ability* –not a cognitive defect or liability. –(NI:25)

He provides as an example, the case of "if," in which "it seems plausible that it is the disposition for rationally accepting instances of inferences like *modus ponens* that is essential to possessing the concept, whereas the dispositions to make fallacious inferences is not" (Ch7:26).

One can certainly agree with the plausibility of the case. But I don't see how it begins to establish Normativism. First of all, it does rather beg the question to suppose that concept possession must be the kind of "power" or "ability" that couldn't be a "defect" or "liability". The question is why we should think it's a power that needs *intrinsically* to be described in any normative terms at all.

Secondly, a doctrine as controversial as Normativism needs to be supported by something more substantial than a seeming doubt. In a footnote supporting his doubt (Ch7:p25fn21), Wedgwood does cite Gareth Evans' (1982:331) remark that "there can be no truth which it requires acceptance of a falsehood to appreciate." But this won't be enough. Perhaps possessing certain concepts requires having certain *conditional* "analytic" beliefs, e.g., that if something's a square it's four-sided, and, arguably, all such beliefs would perforce be true. However, this doesn't entail any claims about the believer's adherence to norms *generally* (I'll return to this point in §4).

Thirdly, there are at least some cases that ought at least give Wedgwood pause. Put aside concepts that can be *truly* applied, and consider many of the concepts that I mentioned earlier are routinely found problematic by philosophers, for example, [freedom]. As Peter Strawson (1962/68) vividly pointed out, it's a concept that plays a significant role in our moral thought, feelings and interpersonal relations, and we may have little choice but to deploy it; but as Galen Strawson (1987) and others have also pointed out, it's a concept that may well be incoherent. Suppose this latter claim were true. Then, arguably someone wouldn't possess the concept unless they had the disposition to the incoherent ideas it implies. Similarly, one arguably doesn't have the traditional concept of [soul] if one doesn't feel the pull of a kind of personal identity, transcending memory, character and bodily continuity that Hume and Parfit have plausibly shown to be unintelligible; or of [angel] if one weren't confused about the status of their "bodies"; or of [cause] if one doesn't feel at least the pull of the idea of a "necessary connection" that seems, well, at least problematic.[15]

Lastly, the same "asymmetry" conception of meaning to which Wedgwood is appealing is advocated (in different ways) by both Jerry Fodor (1987, 1990) and Paul Horwich (1998), the meaning constitutive "laws" or "uses" being the ones on which all other uses asymmetrically depend. Why does Wedgwood insist, unlike Fodor and Horwich, that the basic ones are normative?

3.4 The Defeasibility Argument

I take Wedgwood's main argument to be the following:

> There is a reason for thinking that [the meaning constituting] dispositions cannot in fact be specified without mentioning normative properties or relations. ... They are dispositions to engage in certain forms of *rational reasoning*. For example, one sort of rational reasoning might lead one from having visual experience that presents an object in a certain distinctive way (in the absence of any positive reason to believe one's experiences to be unreliable) to one's forming a belief that predicates the concept "... is yellow" of the object in question. A disposition for this sort of reasoning might be essential to mastering the basic rule of rationality that applies to the concept `...is yellow' and so also to possessing that concept. –(RW:8bot)

15. See Slote (1975) for interesting discussion of some of these and other what he calls "inapplicable concepts," such as [magic] and [miracle].

Now, of course, this *"might"* be true; but what reason is there to insist that it *must* be? Wedgwood goes on to call attention to his proviso "in the absence of any *positive reason* to believe one's experiences to be unreliable" (NI:9, Ch7:28-9), which, he points out, is a normative condition. Indeed:

> These sorts of reasoning are *defeasible*. ... [and] the nature of defeating conditions is precisely that they are those conditions that make it *irrational* for one to regard it as reliable in the circumstances to form the belief or intention in question in response to the relevant input conditions. –(NI:9, Ch7:28-9)

And he adds a few pages later that "the only *simple* way" of specifying what those defeasible conditions all have in common "is in normative terms –as conditions that make it irrational to reason in certain ways"(Ch7:31).

I find these claims puzzling. If Wedgwood is merely explaining, as a matter of *epistemology* in *abstraction* from psychology, what it is to *be* a reason for believing something , then of course, what he says is perfectly plausible. But Normativism is an issue not about epistemology, but about psychology. And the question for psychology is whether people could possess a particular content and not be disposed to apply it appropriately, or to respond to defeaters. Of course they could. A person might have all sorts of patently *bad* reasons to withhold a concept –superstitions, silly theories, blind prejudice– or perhaps a holier than thou confirmation metric ("You call *that* confirming that something is --or even appears-- yellow?!" he cries, worrying about all manner of wild deceptions); or he might apply it in the face of genuine defeaters, failing simply to appreciate them, or mistakenly thinking they in turn have been defeated. Moreover, entirely non-rational, brute physical interferences in the normal operation of the brain could block dispositions here as anywhere, without dissolving a specific content.[16] There may well be *some* necessary connection between a concept and its deployment, but it's not clear how it can be captured by overt dispositions to apply it, or by insisting on a person's appreciating genuinely defeating conditions. These latter are, indeed, a matter in part of a person's rationality –as of her overall psychology– but a rationality that is *additional* to conceptual competence, not *constitutive* of it.

4. General Qualms

In addition to the specific objections I've raised to Wedgwood's arguments, I want to raise some general qualms about his project as he construes it, and suggest a better way that he might try to capture what he's after.

---

16. Indeed, in view of the indefiniteness of the conditions that could interfere, one might wonder whether it's even necessary to lasoo together all the defeating conditions. As Pietroski and I argued in our (1995), *ceteris paribus* clauses are best regarded as checks written on the banks of independent theories, their acceptability turning on the explanatory merits of those theories, not on whether the clause can be replaced by some specification of the conditions under which the unqualified law would hold. So Wedgwood's *proviso* needn't involve *any* characterization, normative or otherwise.

4.1 Intuitions as Evidence of Concepts, not Properties

As we noted in §3.1, Wedgwood claims to be interested not in claims about our *concepts* of intentional states, but in their real essences.[17] However, as Putnam (1965, 1975) so admirably stressed, claims about the real essences of phenomena that exist independently of our thoughts require empirical evidence about how the phenomena fit into the independently existing world.

Wedgwood nowhere appeals to any serious empirical psychology, which, on the face of it, doesn't seem remotely committed to Normativism. To the contrary, as I already mentioned, Kahneman suggests that the fast, "intuitive" sub-systems may be hopelessly disposed to error, and, quite independently of his work, theorists of vision and language comprehension argue that these tasks are performed by "informationally encapsulated modules" that produce illusions that they are constitutionally incapable of correcting (see Fodor 1983). Indeed, discussing vision, Pylyshyn (2003) argues that the system don't engage in "inference" at all:

> Although it might be possible to characterize the operation of the visual system in terms of "rules," these differ significantly from rules of inference since they only apply to representations arising directly from vision and not to those with a different provenance.
>
> –(Pylyshyn 2003:39fn8)

But note that this doesn't lead Fodor, Pylyshyn or other vision theorists to doubt that the systems traffics in intentional representations of, e.g. of edges, objects, surfaces, or spatial points.[18] This fact alone should lead one to wonder about supposed rational constraints on intentional ascription.

4.2 Diagnosis of Intuitions

In the absence of any serious psychological theory about the real essences of the attitudes, it seems to me that Wedgwood's arguments for Normativism must rest in the end on his conceptual intuitions. They seem to me perfectly *good* intuitions –I certainly feel the pull of his claims about [yellow] and [if]. The question that divides us is not the intuitions themselves, but their explanation. Wedgwood thinks they reveal the role of rational norms. I'm inclined to a simpler, more traditional account: they are simply the intuitions philosophers and many others have about "the meanings" of the relevant words.

However, although I share such intuitions, I also share the embarrassment many philosophers have come to feel about them since reading Quine (1953, 1956) and wondering what distinguishes them from merely deeply ingrained beliefs. It was because Quine despaired of drawing that distinction that he and

---

17. He does allow that an *expressivist* about normativity might have to restrict the view to concepts (NE:2), a restriction to which he himself curiously retreats in discussing the concept of a rational attitude (Ch7:7).

18. This seem sometimes to be denied, as in Chomsky (2000, 2003); in my (2003) I argue it shouldn't be.

Davidson turned to charity and Normativism to stabilize ascriptions of content.[19] But, if Normativism is problematic in the ways I have suggested, perhaps we ought to reconsider Quine's despair.

Such reconsideration is not without independent motivation. Many of us have been sceptical not only of Normativism, but in general of Quinean efforts to explain away analytic intuitions. Rather than being particularly "central," which many of them patently are not,[20] they seem on a par with the sorts of intuitions about syntax that regularly inform fairly rich and well-supported theories of grammar (see, e.g., Katz 1972, Chomsky 2000). This is not to say that intuitions are infallible (cf. fn 5). Nor is it to say that the theory of lexical semantics is anywhere near as developed and convincing as the theory of syntax. It's just that the epistemic situation seems so much the same: there are clear convergences in people's judgments, not only about what's rational and revisable, but about, e.g., intelligibility, possibility, synonymy, redundancy, antinomy, and "logical" entailment. To answer Quine, of course, any explanation will need to ground the distinction between these and other judgments, showing, for example, how it is part of a general distinction between a system for language and one for general thought; and what seems to be needed for that is a well supported theory that goes beyond merely the superficial(ist) evidence of mere conceptual intuitions or dispositions to overt behavior. As the case of grammar has illustrated for the last fifty years, the theory needs to be informed by empirical theories about the structure of the mind.

Note that Wedgwood's own particular version of Normativism itself requires analytic claims. Spelling it out, he writes:

> the nature of each concept is given *both* by the principle that defines when beliefs involving that concept are correct, and *also* by certain basic principles of rationality that apply to the use of that concept –that is, basic principles that specify certain ways of using that concept as either rational or irrational. –(Ch7:14)

However, deciding which principles for a given concept are rational or irrational will involve precisely the defense of conceptual intuitions that would establish that these "principles" are analytic. But if they are analytic, then that might *by itself* plausibly explain why it would be irrational to deny them (it mightn't be irrational to apply "yellow" to blue things if "yellow" meant [blue]!). And, as we noted earlier, it might also accommodate Evans' (1982) claim on which Wedgwood relies, about how appreciation of truth can't depend upon acceptance of a falsehood. Any further Normativist principles of rationality would seem unnecessary and, given the other problems discussed earlier, unwelcome.

****

19. As we saw above (fn 11), this seemed to be Davidson's tack. It is also Quine's (1954/76:109, 1970:81) in his treatment of the logical particles. Note that there can be another response to Quine's despair, viz., Fodor's (1987, 1990) "meaning atomism," by which he hopes to ground meaning independently of the analytic; but see my (2005) for scepticism about this approach.

20. To take the well-worn examples, "bachelors are male" and "Pediatricians are doctors," are hardly *central* to anyone's belief system. The only reason people are tenacious about them –and why it seems "irrational" to give them up– is that people regard them as analytic! If someone wants to re-define "bachelor" to include women, fine. But that's just a re-definition, not an "empirical discovery" –much less a sudden fit of irrationality (see my (2005) for further discussion).

References

Bach, K. (1981), "An Analysis of Self-Deception," *Philosophy and Phenomenonlogical Research*, 41 (March): 351-70

Cartwright, N. (1983), *How the Laws of Physics Lie*, Oxford University Press

Cherniak, C. (1986), *Minimal Rationality*, Cambridge (MA): MIT Press

Chomsky, N. (2000), *New Horizons in the Study of Language*, Cambridge University Press

Chomsky, N. (2003), "Reply to Rey," in Antony, L. and Hornstein, N., *Chomsky and His Critics*, Oxford: Blackwell, pp274-87

Davidson, D. (1970/80), "Mental Events," in *Essays on Actions and Events*, Oxford University Press, pp207-25

— (1974/80), "Psychology as Philosophy," *Essays on Actions and Events*, Oxford University Press, pp229-38

— (1973/84), "Radical Interpretation," in *Essays on Truth and Interpretation*, Oxford University Press, pp125-39

— (1978/82), "Paradoxes of Irrationality," in R. Wollheim and J. Hopkins (eds.), *Philosophical essays on Freud*, New York : Cambridge University Press

Dennett, D. (1978), *Brainstorms*, Cambridge (MA): MIT Press

Dennett, D. (1987), *The Intentional Stance*, Cambridge: MIT Press

— (1991), *Consciousness Explained*, Boston: Little Brown

— (1995), "Superficialism vs. Hysterical Realism," *Philosophical Topics*, 22 (1-2):530-6

Devitt, M. (1996), *Coming to Our Senses*: *a Naturalistic Program for Semantic Localism*, Cambridge University Press

Dilthey, W. (1894/1977), 'Ideas Concerning a Descriptive and Analytic Psychology', in *Descriptive Psychology and Historical Understanding*, trans. R. Zaner and K. Heiges, The Hague: Martinus Nijhoff

Elster, J. (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge University Press

Elster, J. (1983), *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge University Press

Fodor, J., (1983), *The Modularity of Mind*, Cambridge (MA): MIT Press

Fodor, J. (1987), *Psychosemantics*, Cambridge (MA): MIT Press

Fodor, J. (1990), *A Theory of Content and Other Essays*, Cambridge (MA): MIT Press

Fodor, J. and LePore, E. (1992), *Holism: a Shopper's Guide*, Oxford: Blackwell's

Frisch, M. (2005), *Inconsistency, Asymmetry, and Non-Locality: Philosophical Issues in Classical Electrodynamics*, Oxford University Press.

Harman, G. (1986), *Change in View: Principles of Reasoning*, Cambridge (MA): MIT

Harman, G. (1995/99), "Rationality," *Reasoning, Meaning and Mind*, Oxford University Press

Hornsby, J. (1993), *Simple Mindedness: In Defense of Naive Naturalism*, Cambridge (MA): Harvard University Press

Horwich, P. (1998), *Meaning*. Oxford University Press

Hurtshouse, R. (1991), "Arational Action," *Journal of Philosophy*, 83:291-5

Kahneman, D. (2002/03), "Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice," Nobel Prize Lecture, 8 Dec 2002, available at website: *http://nobelprize.org/economics/laureates/2002/kahneman-lecture.html*. Ashorter version appears as "A Psychological Perspective on Economics," *American Economic Review* (Proceedings) 93(2), Dec 2003, pp162-8.

Kahneman, D. and Tversky, A. (eds.), (2000), *Choices, Values and Frames*, Cambridge University Press

Katz, J. (1972), *Semantic Theory*, New York: Harper and Row

Kim, J. (1985/93), "Psychophysical Laws," in his *Supervenience and Mind: Selected Philosophical Essays*, Cambridge University Press

Kripke, S. (1972/80), *Naming and Necessity*, Cambridge (MA): Harvard University Press

McGinn, C. (1991), *The Problem of Consciousness*, Oxford: Blackwell's

Mehler, J. and Depoux, E. (1994), *What Infants Know*, Oxford: Blackwell

Moran, R. (2001), *Authority and Estrangement*, Princeton University Press

Nozick, R. (1993), *The Nature of Rationality*, Princeton University Press

Pollock, J. (1991), "OSCAR: A General Theory of Rationality," in R. Cummins and J. Pollock (eds.), *Philosophy and AI: Essays at the Interface*, Cambridge (MA): MIT Press, pp189-213.

Putnam, H. (1962/75), "Dreaming and `Depth Grammar,'" in his *Philosophical Papers*, vol. 2, pp304-24

Putnam, H. (1975), "The Meaning of `Meaning'," in his *Philosophical Papers*, vol. 2, pp215-71

Pylyshyn, Z. (2003), *Seeing and Visualizing: It's Not What You Think*, Cambridge (MA): MIT Press

Quine, W. (1953), "Two Dogmas of Empiricism," in *From a Logical Point of View and Other Essays*

— (1954/76), "Carnap and Logical Truth," in his *The Ways of Paradox and Other Essays*, revised and enlarged edition, pp107-32

— (1960), *Word and Object*, Cambridge (MA): MIT Press

— (1969), *Ontological Relativity and Other Essays*, New York: Columbia University Press

— (1970), *Philosophy of Logic*, Englewood Cliffs (NJ): Prentice-Hall

Reber, A.S. (1992), "The Cognitive Unconscious: An Evolutionary Perspective," *Consciousness and*

Cognition, 1:93-133.

Rey, G. (1988), "Towards a Computational Account of Akrasia and Self-Deception," in *Perspectives on Self-Deception*, ed. by B. McLaughlin and A. Rorty, Berkeley: University of California Press, pp264-96

— (1994), "Dennett's Unrealistic Psychology," *Philosophical Topics*: vol 22 (# 1-2), pp259-289

— (1997), *Contemporary Philosophy of Mind: a Contentiously Classical Approach*, Oxford: Blackwell

— (2002), "Physicalism and Psychology: a Plea for Substantive Philosophy of Mind," in *Physicalism and Its Discontents*, ed. by Carl Gillet and Barry Loewer, Cambridge University Press

Rey, G. (2003a), "Chomsky, Intentionality and a CRTT" in *Chomsky and His Critics*, ed. by L. Antony and N. Hornstein, Oxford: Blackwel3, pp105-39

Rey, G. (2003b), "Representational Content and a Chomskyan Linguistics," for *Epistemology of Language*, ed. by Alex Barber, Oxford University Press (2003), pp140-86

Rey, G. (2005), "Philosophical Analysis as Cognitive Psychology: The Case of Empty Concepts," in Cohen, H. and Lefebvre, C., *Handbook of Categorization in Cognitive Science*, Amsterdam: Elsevier, pp71-89

Rey, G. and Pietroski, P. (1995), "When Other Things Aren't Equal: Saving Ceteris Paribus," *British Journal for the Philosophy of Science*, Vol 46:81-110

Rosen, G., "Brandom on Modality, Normativity and Intentionality," *Philosophy and Phenomenological Research*, 63:611-623 (2001)

Slote, M. (1975), "Inapplicable Concepts," *Philosophical Studies* 28:265-71

— (1989), *Beyond Optimizing: A Study of Rational Choice*, Cambridge (MA): Harvard University Press

Stein, E. (1996), *Without Good Reason*, Oxford University Press

Stich, S. (1983), *From Folk Psychology to Cognitive Science: the Case Against Belief*, Cambridge (MA): MIT Press

Strawson, G. (1987), *Freedom and Belief*, Oxford University Press

Strawson, P. (1962/68), "Freedom and Resentment," in Strawson, P. (ed.), *Studies in the Philosophy of Thought and Action*, Oxford University Press, pp71-96

Weber, M. (1913/81), "Some Categories of Interpretive Sociology," trans. by E. Graber, *The Sociological Quarterly*, 22:151-80

Wedgwood, R. (forthcoming-a), "The Normativity of the Intentional", in Brian McLaughlin and Ansgar Beckermann, eds., *The Oxford Handbook of the Philosophy of Mind* (Oxford: Clarendon Press).
Available also at http://users.ox.ac.uk/~mert1230/normativityofintentional.htm

Wedgwood, R. (forthcoming-b), *The Nature of Normativity*, Oxford University Press

Wilson, T. (2002), *Strangers to Ourselves: Discovering the Adaptive Unconscious*, Cambridge (MA): Harvard University Press

Wittgenstein, L. (1953), *Philosophical Investigations*, trans by G.E.M. Anscombe, New York: Macmillan