

norms is not necessary for rational belief-formation, it nevertheless remains that establishing that beliefs are indeed rational depends on such norms. And if so, then even Knowles's quietist naturalist will not be able to do without norms, if part of her task involves the establishment of particular (especially controversial) beliefs and theories *as* rational. To forswear this project really does seem to reduce our philosophical ambitions overly much—even the naturalist shouldn't be *that* unambitious.

There are many interesting aspects of Knowles's discussion that I haven't space to address here, but which are well worth sustained examination. Knowles is clear that his main thesis and criticisms of NE depend upon his prior embrace of naturalism. In this respect, his central arguments are not aimed at defeating the traditional epistemologist. Rather, they are aimed at naturalized epistemologists who pursue the traditionalist project of explicating and justifying epistemic norms but from within the constraints imposed by naturalism. Knowles poses an important challenge to naturalized epistemologists, one that traditionalists can safely ignore, as long as they can satisfactorily reject quietism. I have offered some reasons for thinking that they can indeed do that.

Department of Philosophy
University of Miami
PO Box 248054
Coral Gables, FL 33124-4670
USA
hsiegel@miami.edu
doi:10.1093/mind/fzi424

HARVEY SIEGEL

Radical Interpretation and Indeterminacy, by Timothy McCarthy.
Oxford: Oxford University Press, 2002. Pp. 253. H/b £35.00.

In *Radical Interpretation and Indeterminacy*, Timothy McCarthy undertakes an attractive project that attempts to bring together two dominant themes in the philosophy of language of the last half century, which have largely been treated separately from each other. One theme concerns the very foundations of the theory of meaning, the other concerns the choice of a theory of reference and meaning for the various different categories of linguistic expression. The first theme raises the spectre of meaning scepticism. McCarthy's target is a certain scepticism concerning the theory of reference, epitomized in Quine's famous thought experiment of radical translation in *Word and Object* (Cambridge, MA: MIT Press, 1960). The attractive thought is that the explanatory role of meaning and reference gives us a general strategy which, when brought to bear on the semantics of the particular categories of linguistic expression (the second theme), can rule out deviant interpretations that give rise to the threat of meaning scepticism (the first theme). McCarthy develops, in the first two chapters, what he takes to be the general explanatory role of meaning and ref-

erence, and the general strategy this affords for ruling out deviant interpretations of speakers. He then goes on in the third chapter to show how this general strategy applies to particular categories of expression, importantly, natural kind terms. The final chapter is devoted to logical constants, tackling in an interesting and novel way the problem of demarcating the logical constants of a language. Our worries about the project concern whether the explanatory role of meaning and reference really does provide an adequate strategy for ruling out semantic indeterminacy and the threat of meaning scepticism that such indeterminacy seems to pose. That will be the principal focus of this review.

As in the case of external world scepticism, one would expect a successful response to the meaning sceptic to show us either how alternative hypotheses *can* be ruled out after all or how the inference from the possibility of alternatives to the sceptical conclusion can be blocked. One difficulty with McCarthy's discussion is that it is not clear which strategy he wants to adopt. Like David Lewis (in his 'Radical Interpretation', reprinted in his *Philosophical Papers, Vol. 1*, New York: Oxford University Press, 1983; originally published 1975), who adopts a *non-sceptical, realist* approach to the theory of meaning, McCarthy sets out to provide 'a relatively modest set of constitutive principles of interpretation, ... [so] as to resolve the indeterminacies of interpretation that naturally present themselves' (p. 2). And, like Lewis, he suggests that, if a set of constraints fails to yield a unique interpretation, then we ought to search for further, as yet unarticulated, constraints on interpretation (see p. 30). However, though it seems natural to read McCarthy as offering a realist account of interpretation, this reading is hard to mesh with 'instrumentalist' claims he appears committed to as he develops his account.

McCarthy focuses almost exclusively on alternative interpretations that allegedly support the so-called inscrutability of *reference* of terms rather than indeterminacy in sentence interpretation. Now, arguments for inscrutability are typically designed to show that, even assuming a fixed assignment of truth-values to the sentences of a language under study, the assignment of extension to terms of the language can vary, consistently with all the speech dispositions of its speakers. (In addition to Quine, see Davidson's *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984, Essays 15 and 16, and Putnam's *Reason, Truth, and History*, Cambridge: Cambridge University Press, 1981, pp. 32–48, 217–8.) Thus, using his famous 'gavagai' example, Quine suggests that a speaker's verbal and non-verbal behaviour (spontaneous or solicited)—*even if supplemented by causal relations to the speaker's environment*—may be entirely insufficient for determining whether a speaker refers to whole rabbits rather than to, say, undetached rabbit parts. (As an aside, we want to point out that, given McCarthy's focus on referential indeterminacy, it is puzzling that he takes as his point of departure what he calls the 'minimal framework' (see paragraph 2.2), which comprises principles whose leading proponents, Davidson and Lewis, actually *embrace* Quine's inscrutability thesis.)

In a somewhat different vein, Kripke famously invites us to reflect on what could determine which arithmetical function a speaker associates with the sign '+', or the expression 'plus', for example, given that any speaker's speech dispositions are finite. McCarthy seems to argue that we *could* rule against the 'quus' scheme by determining whether the speaker instantiates a Turing machine for *plus* rather than *quus*. A Turing machine instantiated by a speaker who means *quus* rather than *plus* by 'plus' could be expected to 'execute a subroutine that counts the numbers represented' so as to decide whether to add, or rather quadd, them (see McCarthy, p. 96). If we understand this to mean that a *plus*-user will have different behavioural and mental dispositions from a *quus*-user, then the point is sensible, but hardly new. In general, one could argue that 'deviant' schemes of interpretation will incur certain commitments as to the speaker's mental organization and operations, regarding which we could have independent positive or negative evidence. But in response, the sceptic could recast his challenge by claiming that a speaker's mental organization (understood in intentional terms) is itself open to incompatible interpretations. If we understand McCarthy to be proposing instead that a speaker who instantiates the *plus* algorithm may not thereby instantiate the *quus* algorithm, the point is again well taken, but has no force against the sceptic who is questioning what grounds the claim that the speaker is instantiating one algorithm rather than another in the first place.

In chapter two (pp. 43f.), McCarthy considers the *permutation argument*. A permutation argument proceeds from the assumption that we somehow have been able to devise a scheme of reference, *R*, for a speaker's language, compatibly with all our evidence. We are then offered a recipe for constructing an alternative scheme *R'* which permutes the original one by mapping each original referent to a new one, and which (it is argued) we cannot rule out non-arbitrarily. McCarthy argues that stock examples of permutation can be blocked by imposing a *Rigidity Condition* requiring that a given permutation should accord not only with the truth-values of the speaker's elementary first-order sentences, but also with the truth-values of his *modal* pronouncements. The condition would rule out a permutation that maps each physical object onto the mereological sum of its time-slices, or momentary stages (see p. 53). And it could perhaps block other kinds of permutation, too, such as that afforded by a 'proxy function' (which would map each individual in our domain of discourse onto her shadow; and, to compensate, would map each predicate to the set of shadows of the relevant individuals (as in Davidson, 1984, Essay 15)), or a permutation that mapped each object in our domain to an object fifty miles to the north (as in McCarthy pp. 79–80).

McCarthy, however, is not interested merely in constraints that would rule out alternative interpretations. He is after constraints that are independently motivated by considerations about the *point* of radical interpretation. This leads him to formulate and defend the *Principle of Conformality*—the centre-piece of his discussion. An interpretation of an agent should aim to provide

what McCarthy calls a *conformal explanation* of the agent's behaviour (see 2.4.1–2). To use McCarthy's example, consider a suitably programmed robot that has the ability to exit a building in an efficient way. A purely causal explanation that describes the robot's internal workings and the causal effect of its internal states on its trajectory could explain how the robot got from *A* to *B*. Yet such an explanation would miss that the robot has a *plan* for exiting the building—a representation of the layout of the building—and would thereby fail to explain why the robot takes an *efficient* route out of the building. A better explanation of the robot's behaviour would make reference to the *relationship* between the dispositions or internal states of the robot and the layout of the building. If we think of the robot as instantiating a *particular* function from a given location in the building and a movement (go straight, turn left, turn right, and so on) to a resulting new location, we can explain the robot's success in navigating the building.

McCarthy suggests that we think of an interpretation of an agent—an ascription of a set of attitudes (beliefs, desires, and so on) and a semantics for the agent's language—as affording a similar explanation of an agent's ability to successfully navigate the world. As McCarthy puts it:

The Principle of Conformality tells us roughly that part of the point of interpretation is to explicate the subject's activity in conformal terms, an aim that favors an interpretation that as far as possible represents the subject as believing stories which correctly describe causal-explanatory connections between her prospective basic actions and outcomes she desires. These stories are the basis of conditional explanations of the relevant outcomes that the agent can frame herself. (p. 12)

McCarthy contrasts his own account with that of Putnam. For Putnam 'a theory of reference for Karl's language can help to explain the *success* of Karl's behavior' (pp. 66f.), where success consists in desire satisfaction. Following Schiffer, McCarthy rejects Putnam's account on the grounds that desire satisfaction is itself a semantically loaded notion. To ground the semantic notions employed in the *explanans* in their ability to explain a semantically loaded *explanandum* would be circular. For McCarthy, the agent's desires, and hence the standard for success, are part of the interpretation—part of the *explanans* rather than the *explanandum*. The *explanandum* is the agent's behaviour described in entirely non-semantic terms, for example, the agent's taking an efficient egress out of a building. Desire satisfaction still constitutes success on McCarthy's view, but it is not part of what conformal explanations seek to explain; it is part of what does the explaining.

In addition to desires, an interpretation seeks to explain an agent's behaviour by ascribing to her conditional beliefs about how to realize her goals through performing a sequence of what McCarthy calls 'basic actions' (ones that do not require a contribution from the world for their success, such as shooting at someone, as opposed to killing them). Importantly, it is because these conditional beliefs are framed by the agent *in her language* that the

semantics of her language is implicated. This is supposed to privilege a particular semantic mapping of the agent's language onto the world, because, supposedly, only a certain mapping will explain the agent's ability to navigate the world. The Principle of Conformality then enjoins us to prefer interpretation schemes which provide the best such explanation. McCarthy takes this principle to rule out the sort of putative indeterminacies generated by proxy-functions (such as Davidson's 'shadow' interpretation). An interpretation that construes an agent's attitudes as being about Karl can explain how the agent is able to track Karl, whereas an interpretation that construes her attitudes as being about Karl's *shadow* cannot.

This idea that there is an important connection between content and the explanation of behaviour is reminiscent of Dretske in *Explaining Behavior: Reasons in a World of Causes* (Cambridge, MA: MIT Press, 1988). For Dretske, what makes Mary's belief-state *about* Karl is, roughly, that it is carrying information about *him* that produced the satisfaction of desires and in turn reinforces the behaviour caused by that state. But the resemblance is only superficial. For, whereas Dretske is after an account of what *determines the content* of intentional states, McCarthy offers Conformality rather as a principle that *constrains the attribution* of such states in the situation of radical interpretation. For Dretske, the content of a belief-state is determined by what *does* explain the agent's behaviour. For McCarthy, the appropriateness of an interpretation (an ascription of content) is constrained by what *would* explain the agent's behaviour. In other words, McCarthy's account is ahistorical in a way that Dretske's is not: there are typically many different interpretations that *would* explain an agent's behaviour. This seems to give rise to precisely the kind of indeterminacy that McCarthy aims to rule out. McCarthy's twist on Putnam's view, for example, becomes problematic. Since desires are part of the *explanans*, rather than the *explanandum*, the standards for success (desire satisfaction) become internal to the interpretation. If we ascribe different desires to an agent, then we get different conditions for success. If there are no interpretation-independent standards for success, then Conformality provides no basis for deciding between such rival interpretations.

There is another way in which Conformality fails as a constraint on interpretation. Consider McCarthy's version of Twin Earth, which (unlike Putnam's original version) is an *exact* duplicate of Earth in every respect. We can permute an interpretation of an agent on Earth (who has in fact never had any contact with Twin Earth) so that every term in the agent's language refers to its Twin Earth counterpart. To preserve truth, we can reinterpret the agent's indexical expressions so they refer to the relevant Twin counterparts. Given that Twin Earth is exactly like Earth in every respect, such an interpretation certainly *would* explain the agent's ability to navigate her way around Earth, in much the way that it would explain your ability to navigate a particular building if you knew the layout of a different building with an identical layout. In effect, we explain the agent's ability to navigate Earth by ascribing to her repre-

sentations of Twin Earth. So, at least, such a deviant interpretation *would* provide an explanation of what McCarthy thinks interpretations ought to explain, namely the ability to navigate the world. It would also respect McCarthy's other constraints on interpretation, notably the Principle of Charity, since it would have the agent believing, not false things about Earth, but true things about Twin Earth—for example, that Twin Washington DC is the capital of Twin USA.

Though this would be a *bad* explanation of an Earthling's behaviour, McCarthy cannot rule it out by appeal to its most obvious defect: it explains the behaviour of Earthlings by crediting them with beliefs they do not have (beliefs about Twin Earth), and an explanation with a false *explanans* is a bad explanation. Given the project of radical interpretation, to reject a putative interpretation on these grounds would be question-begging. At any rate, if we are permitted to appeal to the truth or falsity of interpretations, then we need only one constraint on admissible interpretations (call it the Truth Principle): the only acceptable interpretation of an agent is the one that gets the agent's attitudes and the semantics of her language *right*. We can dispense with the Principle of Conformality and all the rest.

McCarthy could appeal to considerations of explanatory economy to rule out the Twin Earth interpretation of a wholly Earth-bound agent. The Earthly interpretation may seem to afford a more *economical* explanation, in that we needn't supplement it with the fact about the similarity between Earth and Twin Earth. But such a principle of explanatory economy would itself be problematic. It might well be that the right explanation of how I know my way around building A is that I have been in building B and the two buildings are identical in layout. Nevertheless my knowledge is about (represents) the layout of building B, even if an interpretation of me as possessing knowledge of A would provide a simpler explanation of my ability to navigate A. There is no guarantee that the simpler explanation will be the *right* explanation.

Even if the Principle of Conformality does the work McCarthy thinks it does, it is not in the end clear how adequate a response McCarthy has provided to the problem of indeterminacy. As McCarthy himself admits, even if we require that an adequate interpretation scheme should 'generate, for the same range of sentences of the agent's language, conformal explanations of the agent's bringing about the realizations of the truth-conditions they associate with those sentences' (p. 13), we can still expect that competing semantic descriptions of an agent's utterances and intentional states will yield equally good conformal explanations of her behaviour and actions. (See especially paragraphs 1.5–1.6.) Are we to adhere to a Lewis-style robust semantic realism, and maintain that this just means we haven't found all the constraints that would serve to ground the determinacy of interpretation? Are we to suppose that all remaining indeterminacies will be in some sense too trivial to worry about? Or are we rather to adopt the *instrumentalist* position that any of the equally best (even if conflicting) alternative interpretations will be correct,

since success in affording the best conformal explanation is *all there is* to correct interpretation? This last position is one famously defended by Daniel Dennett, who, like McCarthy, thinks semantic notions can play an indispensable role in explaining the behaviour of more complicated systems. (See, for example, his *Brainstorms: Philosophical Essays on Mind and Psychology* (Montgomery, VT: Bradford Books, 1978); we note that, given this affinity, we found it very surprising that McCarthy makes no reference to Dennett's familiar view.) But Dennett explicitly endorses Quine's indeterminacy thesis, which sets him apart from 'intentional realists' such as Fodor. Given the possibility of Dennett's instrumentalist position, it seems incumbent on a realist opponent of Quinean meaning scepticism to assure us that (at least non-trivial) indeterminacies will *not* arise.

In the end, we found McCarthy's book rather disappointing. This is a difficult book—in our view, more difficult than it needed to be. McCarthy uses the tools of formal logic even in places where this hinders rather than helps. His use of formalisms too often compromises readability, and at times obscures important underlying philosophical issues. As for philosophical rigour, we would have preferred more anticipation of objections to his view, along with responses to counterexamples familiar from the literature on interpretation and representation. The book does contain some promising insights, however. In particular, we found attractive the idea that an agent is not just the object of explanation but also has the capacity to frame explanations of her own behaviour and that this might constrain interpretation. Whether this idea can really help rebut Quinean scepticism about meaning determinacy remains to be seen.

Department of Philosophy
UNC-Chapel Hill
Chapel Hill, NC 27577-3125
 USA
dbar@email.unc.edu
deanpettit@earthlink.net
 doi:10.1093/mind/fzi429

DORIT BAR-ON AND DEAN PETTIT

Powers: A Study in Metaphysics, by George Molnar. Oxford: Clarendon Press, 2003. Pp. xiv + 238. H/b £32.00.

Dispositions, like the malleability of copper wires or the water-solubility of salt, have long been regarded by philosophers as metaphysically suspect. In part this is because dispositions, by their nature, seem to concern not the present, actual behaviour of the objects that possess them, but rather the future and in some cases merely possible behaviour of those objects. The inflammability of my house, for example, has thankfully never become manifest, and hopefully it never will: my house's inflammability concerns a merely pos-