

Georges REY\*

## Keeping meaning more in mind<sup>1</sup>

### Garder la signification présente à l'esprit

**Résumé :** De nombreux philosophes supposent que les arguments de Quine contre la distinction analytique/synthétique ont condamné les approches de la psychosémantique en termes de "rôles conceptuels". Je distingue deux objections dans ces arguments: une objection "horizontale" touchant à la distinction entre signification et croyance et une objection "verticale" concernant la réductibilité de la première distinction à des termes non-sémantiques. Je soutiens que la gravité de ces deux objections a été exagérée. Les raisons mêmes qui militent en faveur de certaines théories covariationnelles récentes du contenu justifient également qu'on leur ajoute certains traits du rôle conceptuel d'un terme, ce que les avocats de ces théories, trop impressionnés par les arguments de Quine, ont à tort cherché à éviter. Si cela est juste, l'inquiétude suscitée par le second défi est prématurée: il n'y a pas de raison de saborder des approches par ailleurs plausibles d'une science simplement parce que personne n'est encore en mesure d'en donner une réduction en des termes extérieurs à cette science.

A spectre is haunting philosophy and psychology. It is widely supposed that Quine's arguments against the analytic/synthetic distinction have doomed not only the philosophical enterprise of *a priori* knowledge, but also certain approaches to the more empirical task of a psychosemantics, or a theory of the content of mental states. Theorists frightened by this spectre have constrained their theories accordingly — to an extent that strikes me as almost phobic. Or, anyway, that's what I want to try to establish here. In an earlier paper (Rey 1994), I have tried to show how Quine's arguments are effective only against a narrow, positivistic conception of philosophy, and leave traditional conceptions of that

---

\* George Rey, Dept of Philosophy, Skinner Hall, University of Maryland, College Park, MD 20742; (301)-405-5707; rey@umiacs.umd.edu.

<sup>1</sup> This paper is an expansion of Rey (forthcoming), some passages of which it shares. I'm grateful to Michael Devitt, Pierre Jacob, Gabriel Segal and Karen Neander for helpful comments.

enterprise unscathed. In this paper I want to try to dissolve the spectre over psychosemantics as well.

Psychosemantic theories of the sort I will discuss have arisen in the past decade within the context of, broadly speaking, functionalist, but more specifically, "computational/representational" theories of mind. The general idea is that it is the specific causal relations among brain states that provide them their condition of individuation.

This individuation can, however, be understood along two different dimensions. One can be interested in what distinguishes one *attitude* from another, for example, a *belief* from a *desire* or from an *imagining*. Here the hope is that, corresponding to each different attitude, there is a different *computational* relation to what could nonetheless be the same representation; thus, a belief that it's raining is distinguished from a desire that it rains from the different computational role some representation of the proposition, [It rains], plays inside a person's brain.

Alternatively, one could be interested in what distinguishes one *content* of a belief or a desire from another *content*: for example, what distinguishes a belief that it's raining from a belief that it's snowing. It is to marking this kind of distinction in some theoretically satisfying way that a theory of the *content* of mental representations, or a "psychosemantics," is directed. Although most recent authors are functionalists along the first dimension, there is substantial disagreement about whether functionalism will suffice for the second. Roughly speaking, "internalists" about semantics suppose it will, "externalists" that it won't. And the scepticism of the externalists is driven in part by the worries about Quine's attack on the analytic whose seriousness is my present topic.

Quine's arguments against the analytic can be regarded as involving two, somewhat orthogonal challenges to a psychosemantics that are not often distinguished: the usual focus of attention has been upon what might be called the "*vertical*" challenge to "naturalize" intentional and related mental properties, showing how they can be "reduced" to (defined/explained/constituted by) non-intentional, non-mental properties independently available from the other sciences. But another focus is upon what might be called the "*horizontal*" challenge to specify a *principled distinction within psychology* between matters of meaning and matters of mere factual belief, between connections in thought that are *constitutive* of concept possession and those that are not, between "semantics" and "epistemics."

Now, while both these challenges do raise important issues with which any mentalism must *eventually* contend, what I want to argue here to day is

that the seriousness of both challenges has been exaggerated. Current failures to meet them should not be taken, as Quine and others have claimed, to indict the project of a mentalistic psychology as a whole. This seems to me especially true with regard to the vertical, "reduction" challenge. I see no reason to scuttle otherwise plausible approaches to psychosemantics simply because no one as yet can produce a reduction of them to non-intentional terms. I have in mind recent "locking"<sup>2</sup> theories of content that identify the constitutive conditions for an internal representation's I having a certain content C in some fact about the functional (Millikan 1984, Dretske 1988) and/or covariant (Stampe (1977), Fodor (1987, 1990)) relations of tokenings of I with some external condition C. These theories are thought by their adherents to be accountable to a "purity" constraint": *the locking relation must not be specified in intentional terms, lest the account fall afoul of Quine's challenges*. I want to argue here that the very reasons for favoring such theories supply also a reason for abandoning such purism. Indeed, properly understood, such theories already contain within them the seeds of a reply to the horizontal challenge, which can --and I think must-- be met without such purity; and we have no reason to think they mightn't eventually meet a reasonable version of the horizontal challenge: we have no reason to think, that is to say, that they won't ultimately be as "reducible" or "naturalizable" as any other macro-theoretic property of the world (although specifying what that amounts to is probably a lot more difficult than Quine and others have supposed).

My strategy will be as follows. I shall briefly set out the rationale of locking theories, and mention two classes of counterexamples to purist accounts of them (there will not be time here to consider purist responses). These counterexamples suggest that the way in which locking theories are motivated by their role in psychology is also a reason for supplementing them with features about the inferential role of a state or symbol in a person's mental life. Appreciating this motivation ought, I think, to lead one to be wary of ruling out such a supplementation because of its apparent defiance of the Quinian challenges. By comparison to that motivation, those challenges, I will conclude in §III, are often ill-motivated, and in any case prematurely pressed. As I believe the history of science attests, foundational challenges are seldom met fruitfully in advance of the development of the actual theories they are intended to ground.

---

<sup>2</sup> Loewer and Rey (1992) use this term, having in mind the way in which homing rockets "lock" onto their targets. Locking theories are to be distinguished from "internalist" theories that look not to external phenomena to individuate content, but to features of a representation --e.g. its imagistic properties, or its role in inference-- that are entirely *internal* to the agent.

## I. LOCKING THEORIES

### A. Ideal Co-Variation

Locking theories are externalist theories in which the external relation is some or other kind of co-variation relation between a symbol in an animal's brain and the phenomena in the world that thereby constitutes its meaning. Of course, not just any covariation will do. The ur-idea that seems to me to drive such theories is that the co-variation be one that occurs under ideal epistemic circumstances: at least to a first approximation (we'll look at second ones shortly), what a person means by an expression has to do with what she'd apply it to were her knowledge of the world otherwise complete. You and I have the same concept [alligator] iff were our information about the world otherwise complete, we would agree about what and what aren't alligators. If even under complete information about other matters, we were still to disagree about what are alligators, that would seem constitutive of our talking past each other, expressing different concepts by our words. Locking theories in this way suggest a badly needed principled basis for isolating an interesting *semantic stability* from issues of *epistemic* differences. In particular, a locking theory allows us to capture what in the world an agent is "getting at" in her use of a symbol, isolating that from her relative epistemic success or failure in reaching it. It provides a basis for beginning to predict how an agent will react to further evidence and argument, enabling us to distinguish cases alterable by such processes from those not so alterable, cases in which one needs to get the agent to deploy a different concept and "to think of the phenomenon differently." As the "locking" metaphor is intended to convey, insisting on such distinctions is like insisting on a distinction between guided missiles that end up at a certain location because that's where they were aimed, from those that, aimed elsewhere, ended up there because of an error in navigation.

A particularly striking feature of externalist locking theories is that they draw this distinction between semantics and epistemics in such a way as to allow for *arbitrary divergences in people's beliefs*. It certainly looks as if people could believe *anything*. This is most evident in the case of philosophers: with enough intervening theory, people have convinced themselves that all is water, fire, ideas or texts. But it arises also in comparing people across widely different cultures, different times, and different age groups (children, the senile). By confining semantics to merely an idealized causal covariation under special circumstances, such divergences in peoples' beliefs in *actual* circumstances are easily tolerated as due to intervening factors, failures largely of epistemic position or clear reasoning. In this way, one might hope to meet at least

Quine's *horizontal* challenge, allowing that *no* belief is constitutive of content.

Of course, ideal epistemic covariation by itself probably won't solve the vertical problem: as many have noted, ideal epistemics is probably intentionally characterized epistemics. In an ingenious effort both to recognize the above latitude in interpersonal epistemics *and* to avoid this problem, locking theorists like Dretske and Fodor try to capture the requisite co-variation in a purist way.

## II. PROBLEMS WITH PURISM

### A. Fortuitous Lockings

The first problem with purist proposals is that they run the risk non-psychological instantiations; what I call "fortuitous lockings." Both Dretske's 2nd-order causal account and Fodor's asymmetric dependency accounts are cases in point. It seems perfectly possible for the patterns to arise in many clearly non-mental cases: in the case of Dretske's account, one need only consider mere biological cases in which the fact that some state indicates, say, damage to an organ is responsible for some chemical response in the system (e.g. release of adrenalin); and in the case of Fodor's asymmetric dependencies, consider one law that facilitates another, without being facilitated by it: an automated relay, in which object B gets to a certain destination only if A does, but A's getting there doesn't depend upon B's doing so (the troop train goes from Minsk to Pinsk only if the supply train does; but not vice versa)<sup>3,4</sup>. If such cases are possible, then one of my cognitive states, say, an entokening of a symbol 'S', could turn out to be one of them: 'S' might be locked onto *cows* in a standard cognitive way, but be also, unbeknownst to me or anyone, locked onto the probings of some neurosurgeon. A purist would be thereby committed to 'S' having a semantics that is *ambiguous* between [cow] and [probe on area A] --just as ordinary uses of 'gas' in (American) English mean either a state of matter or liquid petroleum.

It's important to note that the issue here isn't merely one of intuitions, but of good psychological explanation. In the case of a genuinely ambiguous term like 'gas' it is surely important that *a speaker takes* two different phenomena to be the *right* kind of causes of her uses of 'gas'. By contrast, neither Penfield nor his patients have the slightest reason to think that the representations caused by his pokes are the least bit ambiguous. The subjects' memories and reports, say, "I'm recalling the cows we kept" didn't have as their truth condition *either* [I'm remembering the cows we kept] *or* [My area A is being electrically probed]. Indeed, the subjects needn't have had the slightest conception of either electric probes or the relevant areas of their brains. I doubt that anyone in *my* community would

---

<sup>3</sup> E.g. a system of specific keys and skeleton ones, where the skeleton works only if the specific does, but not vice versa. This might be arranged electronically, so that the department secretary has access to individual offices only when their occupants do, but not vice versa.

<sup>4</sup> HIV may provide a case: As things stand it incapacitates the immune system. But the search for a vaccine is a search for a benign form that, were the system to respond to it, it could then respond to the deadly form. Break the response to the benign form, and you break the response to the deadly form --indeed, the system is destroyed. But break the response to the deadly (introduce one more mutation), and you still could have the response to the benign.

be at all inclined to think they've been "getting at" just *any* phenomenon that happens to be entirely non-cognitively locked onto a state of their brain. Such causes do not give rise to any meaning intuitions, they would not enter into cognitive deliberations, and, most importantly, they would not seem to figure in any *cognitive psychological* laws:<sup>5</sup> cognitive manipulation by *evidence or reasoning* related only to the probes would not *per se* affect the entokenings of 'S'.<sup>6</sup> Worse, unlike the earlier innocuous cases, here the semantic assignments actually *interfere* with the psychology.

### B. Neglecting Distinctions

The second difficulty I want to discuss is that a pure externalist account is too crude. Sometimes we are interested not only in what worldly phenomena, if any, an agent is getting at, but *what way of thinking of that phenomenon* she is trying to get at as well; and we have no reason to think that the world is sufficiently rich in phenomena independently of our minds to distinguish among them. Indeed, it's notorious that "extensional" accounts run into standard difficulties of distinguishing among *co-instantiated concepts*, i.e. concepts that are instantiated in all the same possible worlds and/or counterfactual situations, such as 'triangle' and 'trilateral'<sup>7</sup>, 'eucalytus' and 'gum', 'rabbit', vs. 'undetached rabbit parts', vs. 'temporal stage of a rabbit'<sup>8</sup>.

Fodor does suggest a reply to cases like these. Both in his (1990) and in his recent Nicod lectures (1995), he hopes to confine them to cases in which the actual representations in the brain have different logical forms. What Fodor is of course hoping is that all the work asked of a theory of concepts as *senses* --in particular, the work of distinguishing among co-instantiated concepts-- can be performed by concepts as *internal representations*.

---

<sup>5</sup> For example, laws specifying the decision-theoretic links between beliefs, preferences and choices of action; or about patterns in (rational and/or fallacious) reasoning; or about the relation of specific intentional states (e.g. fear of immediate danger) to non-cognitive effects (release of adrenalin).

<sup>6</sup> Cases of the essential reliance on experts would, of course, involve a borderline case of these considerations, and would require subtler discussion than is possible here --but in any case than is available on Fodor's theory.

<sup>7</sup> If you find 'triangle' vs. 'trilateral' a little too fine for your taste, try 'equiangular Euclidean triangle' vs. 'equilateral Euclidean trilateral' --the terms that in high school you had to *prove* were indeed necessarily co-extensive.

<sup>8</sup> Quine (1960:chap 2) deserves credit for discovering this latter, distinctive category and the interesting puzzles to which it gives rise, even if it's not at all clear that he's entitled to derive from it his sweeping "thesis of the indeterminacy of translation."

However, there are a number of difficulties in relying on internal representations in this way. In the first place, one might wonder why what is logically complex in English need be logically complex in the language of thought; and, in a related vein, whether what is logically complex for one person is so for another.

Consider in this connection the phenomenon of "subception," whereby many animals are able to recognize groups of things of certain (usually modest) cardinality (Gallistel 1990:ch 10). Take a pigeon that has been trained to peck at "three-membered patterns" Along the lines of a pure locking theory, we can suppose that one of these animals actually does have an internal symbol that locks onto, say, the property of triplicity: for any sortal property, F-ness, that it can otherwise discriminate, it can also discriminate 3F-ness. Now, does this animal plausibly have the same concept [three] that I have? There's this reason to think not: the concept I (and most of us) have is a concept controlled by something like Peano's axioms whereby we can be led by reasonings into understanding a potential infinity of complex arithmetic truths. As Gallistel remarks:

To discriminate on the basis of numerosity is not, however, to have a concept of number, if by "having a concept of number" we mean that an animal is capable of manipulating [numerical representations] in accord with the relational and combinatorial operations of arithmetic. (1990 : p. 348).

Humans who can actually reason arithmetically have what might be called the concept [Peano-three]. [Peano-three] and [subcept-three] certainly pick out the same worldly phenomenon; for a symbol to lock onto the one is for it to lock onto the other. But there is good reason for distinguishing them from a psychological point of view: the cognitive manipulations for the one are not the same as for the other --no amount of argument or reasoning would get a pigeon to realize that  $3^3=27$ : the bird just doesn't have a symbol controlled by the relevant rules. Indeed, it may well be that normal human beings have both concepts, and that it's informative to learn that subcept-three = Peano-three.

A particularly crucial set of cases of necessarily co-instantiated concepts are the *necessarily uninstantiated* ones: e.g. [largest prime], [round-square]. These may not be limited to logically complex cases either. Arguably, nothing could possibly satisfy [miracle], [magic], [unicorn], [monster], [free will] or [soul].<sup>9</sup> Perhaps some of these cases are

---

<sup>9</sup> Slote (1975) has argued, nothing could possibly satisfy our concept, [monster]: when "Nessie" of the Loch Ness is ultimately captured, she'll quickly be dissected, analyzed and classified by the techniques of natural science, after which she'll no longer be a "monster", but just another animal (for example, a wayward dinosaur).



spurious. But all that is needed are two. What is a pure locking theory to say of *them*? That they are one?

Fodor (1990) briefly addresses this problem, and speculates that

*no primitive symbol can express a property that is necessarily uninstantiated.*  
(1990:101).

Now, possibly it will turn out that there is metaphysics enough for all our primitive thoughts, a property for every primitive predicate. But it would seem pretty rash to bet one's theory of content upon it.

Perhaps with enough maneuvering a purist theory like Fodor's could provide an account of the content of all purely descriptive (if you like, "scientific") concepts, like [water], [gold], [alligator], [unicorn]. There is still the problem to be faced that not all our concepts are descriptive or scientific. As many philosophers have stressed, a great range of our concepts involve varying degrees of *normativity*, very often a quite obvious *moral* normativity. *Pace* Dennett, I am *not* thinking of ordinary mentalistic concepts like [belief]. Rather I am thinking of what have come to be called "response-dependent" concepts like [shameful], [tragic], [funny], [bizarre]: although it's not clear to me that possession of these concepts actually requires any specific behavioral response in a person, it's hard to see how they don't require some appreciation of certain conceptual connections, e.g. between being funny and being an occasion for laughter, being tragic and an occasion for tears. But quite apart from the intuitive connections that would seem to be required in such cases, there is the crucial point that *the same normative concept can cause its possessors to lock onto different phenomena*. After all, it is a commonplace that different people find different things funny, shameful, tragic, bizarre; and this is not in all cases plausibly due to differences in their epistemic position. You and I may disagree about what's funny, well, just simpliciter, perhaps as a result of simply very brute differences between our nervous systems: you may simply have a different threshold for, say, gallows or toilet humor. Consequently, there is no reason to expect convergence even under ideal circumstances; nor consequently any reason to think that a purely externalist theory like Fodor's could account for what is the nonetheless patent fact that you and I could perfectly well share the concept [funny] (or, *mutatis mutandis*, [shameful], [tragic], [bizarre]).

These three sorts of counterexamples to a Purist theory like Fodor's -- gratuitous lockings, necessary co-extensives and the normative-- suggest that any co-variational theory will need to be supplemented by some facts about a term's inferential role: `cow' plays a certain role in a person's scientific reasoning; `triangle' bears a direct relation to `angle' that `trilateral' lacks; `undetached-proper-rabbit-part', and `rabbit' play different

roles in mereological inferences;<sup>10</sup> [free will], but not [unicorn], is tied to impossible claims about spontaneity and moral responsibility; [funny] is tied to laughter. Moreover, it is by requiring that any locking relations be brought about as the result of the deployment of a defining inferential role that a locking theory can avoid the fortuitous lockings that we saw a pure theory invites. Such defining inferential roles are arguably "senses" --they *quite literally* determine reference; and it is they, for the reasons I have given, that are as indispensable to psychology as psychology is to them.

### III. QUINEPHOBIA

Now, Quine would, of course, have none of this. Leave aside whether he would buy into ideal co-variation theories at all (although note that they are only a generalization of his (1960) account of "stimulus-meaning"). Supplementing a theory of content with an appeal to inferential role would seem to flout both the vertical and horizontal challenges mentioned at the beginning that he raises against any semantic theory. In view, however, of the problems we've encountered with a purist account that attempts to meet those challenges, it is worth considering how serious these flouts --or the challenges themselves-- might be. I shall consider the horizontal challenge first.

#### A. Fear of the Horizontal

Quine's horizontal challenge to distinguish matters of meaning from mere matters of belief is, of course, of a piece with his argument against the analytic/synthetic distinction. This argument involves four strategies: an attack on truth by convention, an appeal to revisability, claims of confirmation holism,<sup>11</sup> and charges of explanatory vacuity. Elsewhere (Rey 1994), I argue that none of these strategies actually succeed: they depend largely on either a behavioristic theory of linguistic processing, or, more importantly, on what I call "superficialist" assumptions about meaning: that what someone means is a fact readily available in her introspections or her behavioral dispositions. Neither of these views are ones that a cognitive scientist, interested in deeper facts about the mind, need take seriously for a second.

---

<sup>10</sup> This is not intended to be a full reply to Quine's (1960) interesting challenge; only an indication of where on my view a reply, if any, is to be found.

<sup>11</sup> "Confirmation holism" is the view (suggested by writings of Duhem and defended by Quine) that a claim is confirmed not in isolation but only in conjunction with the whole of the rest of an agent's beliefs. It is distinct from "semantic holism," which is the view that the *meaning* of a claim can't be specified in isolation, but only in relation to the rest of an agent's beliefs.

Indeed, as I argued in that paper, I see no reason why analyses might not express rules that control the deployment of predicates, but which are (in Stich's (1978) phrase) "sub-doxastic" and not readily available to either introspection or behavioral dispositions. Stich rightly distinguished the different roles that grammatical rules can play for a linguist like Chomsky: they could be rules that are consciously known to a language user, in the way that, say, the rules of chess ordinarily are to any competent chess player. Here it is important that the rules are available for almost any sort of reasoning in which the agent may choose to engage; in another of Stich's phrases, they are "informationally promiscuous." But it's pretty clear that the rules that Chomsky is after aren't of this sort: even most linguists (not to mention non-linguist adults and children) find the task of articulating the rules of their native tongue immensely difficult; and what rules they may postulate are fairly obviously not available for most ordinary reasonings. The theories that linguists like Chomsky are trying to articulate are not likely to be the least conscious or informationally promiscuous. They are "sub-doxastic," or below the ordinary level of belief.

Such sub-doxastic rules in semantics are, then, compatible with the familiar facts about the unlimited revisability of beliefs so stressed by Quine, and even with reasonable versions of confirmation holism. Indeed, they fully allow the extreme divergences of belief that we noticed were a main motive for locking theories: given enough interference from other quarters, *doxastic* states may, after all, diverge arbitrarily from sub-doxastic ones --this is, of course, precisely how we might distinguish "perverse" philosophers like Berkeley or Hegel from the genuinely conceptually odd, whose *sub-doxastic* semantic rules do differ in these ways from our own.

It is important to notice that this proposal does deviate from *certain* traditional conceptions of the "analytic" in one particular respect that has plagued the discussion at least since Kant: it is *not* a requirement of the above suggestion that analyses be actually accessed every time a thought involving the analysandum is processed: someone might well have thoughts involving the concept [bachelor] without actually accessing the analysis [unmarried, eligible male]. Indeed, the evidence suggests that this is often the case: the vast experimental literature of "prototypes" (e.g. Rosch (1978), Smith and Medin (1981)), seems to confirm what a moment's introspection strongly suggests: that we often reason with stereotypes and (representations of) remembered exemplars, not with anything like the serious analyses that we may nonetheless know are

entirely available.<sup>12</sup> Analyses need play only an indirect role, as backup, controlling the deployment of the predicate in hard cases like those that tend to arise only in e.g. philosophy and law.

Nevertheless, it might be thought that Quinian scepticism about such rules has received further support from Putnam's and Kripke's familiar observations on the semantics of natural kind terms. These are terms, like 'water', 'gold', 'multiple sclerosis', which seem to receive their proper definitions only in *a posteriori*, empirical theories of, e.g. chemistry or medicine (it is such observations that lead Fodor (1990) to claim that the meaning of 'water' may be the same as the meaning of 'H<sub>2</sub>O'). If such observations are correct, then how could the content of a non-expert's usage of these terms (or their mentalese equivalents) be determined by any sort of internal rule, sub-doxastic or otherwise?

A natural suggestion is that the rule in the head involves --along the lines of Grice (1965) and Putnam (1975)-- some sort of "blank space to be filled in by the specialist," a kind of indexical element that permits a full semantic content to be determined by the context with which the agent interacts, much as the semantics of indexical terms like 'I', 'now', 'this' and 'that' do.<sup>13</sup>

Moreover, there are limits to externalism.<sup>14</sup> It certainly appears that *some* contents are in fact *fixed*, and *not* marked for empirical completion. Indeed, as George Bealer (1982) has emphasized, it wouldn't be suprising if many of the traditional *philosophically interesting* cases of e.g. [knowledge], [justice], [piety], [free will] did have such fixed contents, while the more *scientific* examples emphasized by Kripke and Putnam (as by Locke and Kant before them) were just the ones that were marked for empirical completion. After all, it was probably only Jerry Katz who insisted it was *analytic* that cats are animals, and so generated a debate with Putnam about robot cats that left many of us wondering who cared. [Cats], as a natural kind term, is *just* the sort of example that might involve the above indexical elements, and so be highly manipulable by context in the way that that debate showed it to be.

---

<sup>12</sup> See Fodor, Garrett, Walker and Parkes (1980) and Armstrong, Gleitman and Gleitman (1983). The needlessly strong supposition to the contrary is encouraged by Kant's remarks when he introduces the analytic, and by the intense interest in psychology in performance *processing* models of reasoning, rather than in a characterization of conceptual *competence* and *possession conditions*.

<sup>13</sup> Of course, on such a view, there might be *wide* analytic truths, for whatever they may be worth (being straightforwardly empirical, they probably would not do any of the work they were originally being asked to do).

<sup>14</sup> Here I've been influenced by Bealer (1982) and Keil (1989).

In any case, the hypothesis that there are sub-doxastic analytic rules seems worth exploring as an empirical hypothesis separate from the assumptions with which Quine burdened traditional semantics. It is a hypothesis about the mechanism of meaning that avoids the problems about e.g. unrevisability, divergence, decomposition, unavailability and externalism that worry Fodor and Quine. At the same time it provides a basis for ruling out the fortuitous lockings that a pure locking theory permits, and for capturing the fine distinctions that such a theory neglects.

### **B. Fear of the Vertical**

It is at this point that Quine might press his more vertical complaints. These have evolved over the years from verificationist ones about providing an empirical test for synonymy to his more metaphysical thesis of the indeterminacy of translation.

The verificationist challenge is most conspicuous in "Two Dogmas" and "Carnap and Logical Truth." There, he argues that there seems to be no way out of an "intentional circle" of notions of analyticity, synonymy and modality:

a dichotomy of...truths...into analytic and synthetic, truths by meaning postulate and truths by force of nature, has been given no tolerably clear meaning (1956:131-2).

What is it to give a distinction a "tolerably clear meaning"? Well, Quine calls himself an "empiricist," and, for any meaningful statement, "the factual component must, if we are empiricists, boil down to a range of confirmatory experiences" (1953:41). Indeed, as he later makes quite explicit:

to learn a language is to learn the meaning of its sentences, and hence to learn what observation sentences to count as evidence for and against them. -- (1973:p38)

So Quine's verificationist complaint is that the analytic/synthetic distinction can't be reduced to a specific range of confirmatory experiences.

The trouble with this complaint is that it, itself, falls afoul of his own argument in "Two Dogmas" and elsewhere on behalf of "confirmation holism." Claims --and distinctions-- are justified "not individually, but only as a corporate body" (1953:41). Why mightn't there be *indirect*, as yet not altogether clear ways of drawing the analytic/synthetic distinction, along the lines of the sub-doxastic rules mentioned above, or in terms of an empirical hypothesis about a specific computational organization of our minds or brains? After all, if "for predicting future experience in the light

of past experience, physical objects are conceptually imported...not by definition in terms of experience, but simply as irreducible posits" (1953:44), why shouldn't the same hold for rules of meaning or analysis?<sup>15</sup>

But, of course, Quine could and has put the above complaint in other ways that can't be dismissed quite so easily. Leave reduction to experience aside: the "indeterminacy" thesis is that, apart from certain observational cases, facts about meaning cannot be captured but by any *non-semantic* facts about agents or their relations to the world. Quine tries to establish this through the kinds of examples of necessarily co-divided predicates ('rabbitt', 'undetached rabbit parts') that we already considered against Fodor's purist theory. He does recognize that appeals to inferential roles (captured by what he calls "analytical hypotheses") would help sort things out, but argues that there are incompatible such appeals that would each be consistent with a person's observational speech dispositions (which are, he claims, the only evidence we could possibly have for translation). In an interesting gloss on his own conclusion, he compares it to the

thesis of Brentano's ... that there is no breaking out of the intentional vocabulary by explaining its members in other terms. ... Using the intentional words 'believe' and 'ascribe', one could say that a speaker's term is to be construed as 'rabbit' if and only if the speaker is disposed to ascribe it to all and only the objects he believes to be rabbits. Evidently, then, the relativity to non-unique systems of analytical hypotheses invests ... intentional notions generally. --(1960:220-1)

Now, although the examples of co-divided predicates are challenging ones, neither Quine's nor Brentano's grand hypotheses follow from them alone.<sup>16</sup> After all, our assessment of these examples rests at a crucial juncture on our ability to imagine a way of specifying physical facts that would nail a particular translation down. Who's to say that we've considered all the possibilities, or all the possible evidence? Do we really understand either psychology or our ordinary mental idiom so well to agree to his behavioristic restrictions on possible data?

Of course, we want some reply to Quine's examples --eventually. Just as we want some account of how psychology relates to more general theories

---

<sup>15</sup> Quine argues that even if the "phenomenalist" attempt to define (or "reduce") material object talk to claims about merely sense experience fails, still material object talk as a whole can be defended as the best explanation of that experience. Material objects are "posited" in the same way as the theoretical entities of any other science.

<sup>16</sup> It is interesting that Quine seldom provides a *general* argument for the thesis. The only place I know of in which he tried is in a short piece, "On the Reasons for the Indeterminacy of Translation," in which not only is the argument specious, but the conclusion sufficiently vague to be compatible with any determinacy any mentalist has ever cared to claim. It is notable, however, that the article has never been reprinted (his confidante, Burton Dreben, has in lectures in Cambridge and Paris claimed that he has renounced it).

--eventually. But I see no reason why we shouldn't hold out for something a bit more sophisticated than behaviorism or the above proposal of Quine's merely linking translation directly to belief. Certainly the appeal I proposed above to a term's inferential role is not *viciously* circular in the way that Quine's particular proposal would seem to be. Such appeals would have to be judged by the overall simplicity, coherence etc, they bring to our explanations on the whole. To return again to the comparison with material objects, why shouldn't intentional states be "irreducible posits" needed for science, in this case a science of intelligent phenomena? Indeed, why shouldn't they be like the posits of other macro-sciences, where it is also improbable that terms could be "defined" or "reduced" one by one to terms in other sciences, but rather only be "justified as corporate bodies." Is "belief that p" really any worse off in this respect than 'species', 'survival', 'heart', or 'kidney' in biology? Or 'capital', 'inflation', 'marginal utility' in economics? Is anyone really prepared to offer, much less insist upon, a reduction of 'is a capitalist' to non-economic terms?

However, suppose that, contrary to fact, we really did have an adequate psychology, but still couldn't think of how to explain its determinacies by more general theories. Where would that leave us? Should we conclude forthwith that psychology was indeterminate? Well, one does often hear it said<sup>17</sup> that the reason one ought to reject claims about telepathy for example, is that we can't imagine any physical mechanism that could mediate it. But while it may be true that the lack of an imaginable mechanism ought certainly to give us pause, surely what really argues against telepathy is the simple lack of any substantial evidence for it *at its own, horizontal level*, independent of vertical reduction. Indeed, suppose that certain people did turn out to be *terrifically good* at passing telepathy tests: they could guess the card the experimenter had in mind *in every case*, and even know when their loved ones (and even their plants!) 1000 miles away were in some sort of danger. If there really were such robust data, then reductionism and maybe even present physics be hanged: we would just have to revise our view about reducibility and/or about what's physically possible (cf. the radical revisions being considered by the results of the "Bell Inequality" experiments in quantum mechanics).

Now, unlike the case of telepathy, the horizontal evidence for intentional explanation is overwhelming. Consider just the "standardized tests" like the SAT and GRE that millions of students suffer through every year; try to explain what are (in even these days of declining scores) the staggering correlations between questions and answers, and in answers

---

<sup>17</sup> Barry Loewer has argued this to me a number of times.

across students, without assuming that the students have minds.<sup>18</sup> Quine, of course, thinks he does have a better hypothesis in his Skinnerian behaviorism, but this is now known to be empirically bankrupt. Mentalism may not be the most well-worked of our theories --I submit we're only just beginning to understand it-- but it's by far the best suggestion we've got, and the lack of a satisfactory reduction of it to the other sciences is as yet quite insufficient to tell against it. One sometimes gets the impression from Quine (and from other proponents of mental indeterminacy, like Dennett 1991) that he thinks with respect to psychology that we're in the position of Einstein, reflecting in 1905 on issues in deep and well-worked out theories of motion and electro-magnetism. But surely our position is more like that of Zeno reflecting on the possibility of motion without the resources of modern physics and mathematics.

So why the impatience in meeting the vertical challenge? Why demand clear explanatory foundations for psychology when both the target theory and the general demands on reduction are so unclear? Not to get too psychoanalytic, what I suspect underlies Quinephobia is a fear of dualism: unless we meet Quine's vertical challenge *before* working out a serious psychology, the worry is that we'll be stuck with a "ghost in the machine" afterwards. Witness Quine's worries about an "autonomous science of the mind." But who said the only alternative to reduction was *complete autonomy*?

Actually, I'm not at all sure what a serious doctrine of an "autonomous science of the mind" really is supposed to come to. Surely no one in the debate is *seriously* worried about ghosts. Perhaps some people are worried about Platonistic claims that we just "intuit" truths about abstract entities and relations. But the problem here is not just that we can't imagine a *physical* mechanism: we can't imagine any *explanation* at all! Is the claim then that, as Wittgenstein put it, "explanation comes to an end somewhere" --and, as with both Wittgenstein's shopkeeper, or Godel's claims about seeing numbers, these explanations just end with brute psychological facts? Quite apart from issues of physicalism or dualism, why in the world should we accept such a claim? Even if there *were* ghosts, or non-physical minds, or just abstract objects, that would be no excuse for bringing explanation to a screeching halt. So perhaps that is the problem: explanatory screeching halts are not familiar from any other science (imagine someone claiming there could be no non-biological explanation of life, or no non-economic explanation of economic phenomena). What

---

<sup>18</sup> I cite these tests, since the correlations they provide can be described "objectively," in terms of printed inputs and graphite outputs, without presupposing the intentional idiom for which they are therefore unbiased evidence.



would it be like to be told that *any* property really is, in itself, and not merely in relation to what we happen to now at a particular time, hopelessly inexplicable? Indeed, why shouldn't explanation continue on indefinitely, our finding even deeper properties and principles to explain what we had hitherto regarded as fundamental? At any rate, no non-Quinian mentalist who is as yet unclear about the demands of reduction need accept such halts to explanation.

That is, it's worth still persisting in the effort to naturalize the mind. *My point is only that, in the face of the overwhelming evidence for intentionality at the horizontal level, our inability to think of a satisfactory vertical "reduction" is not a very interesting argument against it, nor a reason to constrain psychosemantic theories in the way that purists insist.* Fear of ghosts is just as silly as believing in them; and fear of doctrines based on fear of ghosts is no better. Such phobias have no place in working out adequate theories of psychosemantics, which should therefore not be daunted by the prospect of an inferential role semantics whose relation to underlying theories may be complex.

Georges REY  
Dept of Philosophy, Skinner Hall,  
Univ of Maryland, College Park, MD 20742

### ***Bibliography***

- Armstrong, S., Gleitman, L, and Gleitman, H., (1983), "What Some Concepts Might Not be," *Cognition* 13(3): 263-308.
- Bealer, G. (1982), "The Philosophical Limits of Scientific Essentialism," in *Philosophical Perspectives I, Metaphysics, 1987*, ed. by J. Tomberlin, Atascadero: Ridgeview.
- Dennett, D. (1991), *Consciousness Explained*, Boston: Little Brown and Co.
- Dretske, F. (1988), *Explaining Behavior: Reasons in a World of Causes*, Cambridge: MIT (Bradford).
- Fodor, J.A. (1987), *Psychosemantics*, Cambridge: MIT (Bradford).
- Fodor, J.A. (1990), *A Theory of Content*, Cambridge: MIT (Bradford).
- Fodor, J.A. (1995), *The Elm and the Expert*, Cambridge: MIT (Bradford).
- Fodor, J.A., and Garrett, M., Walker, E. and Parkes, C. (1980), "Against definitions" *Cognition*, 8, 263-367.
- Gallistel, C. (1990), *The Organization of Learning*, Cambridge: MIT (Bradford).
- Grice, H.P. (1965), "The Causal Theory of Perception," in R.J, Swartz, ed., *Perceiving, Sensing and Knowing*, New York: Doubleday.

- Loewer, B. and Rey, G. (1992), *Meaning in Mind: Fodor and His Critics*, Oxford: Blackwell.
- Millikan, R. (1984), *Language, Thought, and Other Biological Categories*, Cambridge: MIT (Bradford).
- Putnam, H. (1975), "The Meaning of 'Meaning'" in *Mind, Language, and Reality* (Philosophical Papers, vol. II), Cambridge (U.K.): Cambridge Univ. Press, pp. 215-271.
- Quine, W. (1953), "Two Dogmas of Empiricism," in *From a Logical Point of View and Other Essays*, New York: Harper & Row, pp. 3-15.
- Quine, W. (1956b), "Carnap and Logical Truth," in *Ways of Paradox and other Essays*, (2nd ed: 1976) Cambridge: Harvard University Press, pp. 107-132.
- Quine, W. (1960), *Word and Object*, Cambridge: MIT Press.
- Quine, W. (1973), *The Roots of Reference*; LaSalle: Open Court.
- Rey, G. (1993), "The Unavailability of What We Mean: a Reply to Quine, Fodor and LePore" in *Grazer Philosophica*, special edition ed. by J. Fodor and E. LePore, pp. 61-101.
- Rosch, E. (1978), "Principles of Categorization," in Rosch, E. and Lloyd, B. (eds.), *Cognition and Categorization*, Hillsdale, NJ: Lawrence Erlbaum and Associates, pp. 27-48.
- Slote, M. (1975), "Inapplicable Concepts," *Philosophical Studies* 28:265-71
- Smith, E. and Medin, D. (1981), *Categories and Concepts*, Cambridge, MA: Harvard University Press.
- Stampe, D. (1977), "Towards a Causal Theory of Linguistic Representation," *Midwest Studies in Philosophy*, Minneapolis: Univ. of Minnesota Press, pp. 42-63.
- Stich, S. (1978), "Belief and Sub-Doxastic States," *Philosophy of Science*, 45 (4): 499-518.