FODOR

3 Rabbit Redux (or, 'Reference Scrutinized')[1]

The second lecture in this series closed with a tentative proposal about the individuation of thoughts: Having a thought is being in a three place relation between a thinker, a (broad) content, and a mode of presentation. Since modes of presentation are linguistic expressions (e.g., sentences of Mentalese) and since linguistic expressions are individuated (inter alia) by their syntax, token thoughts are type distinct if they differ *either* in their contents *or* in their modes of presentation.

This treatment of the individuation of thoughts is, of course, tailor-made to mediate between a semantics that wants to reduce meaning to information, and a psychology that wants to reduce thinking to computing. On the one hand, the informational theory says that content is constituted by symbol-world relations. It is therefore hard put to see how *Jocasta* thoughts could differ in content from thoughts about O's M; or, to vary the example, how thoughts about water could differ in content from thoughts about H_2O.[2] On the other hand, computational psychology requires syntactic differences between *water* thoughts and *H_2O* thoughts *whether or not* they are identical in content. This is because they have different causal powers, and, according to the

Turing picture of psychological processes, the causal powers of mental states supervene on their syntax.

So, then, if thoughts with identical contents may nevertheless be distinguished by their syntax, semantics gets a solution to the Frege problems that's compatible with its externalism, and psychology gets a solution to the Frege problems that's compatible with its computationalism, and both have cause to rejoice. Suppose it turns out that *the very same* syntactic structures that semantics needs mental representations to have in order to accommodate the Frege cases will also serve to define the domains of computational mental processes. That would show beyond any serious doubt that Turing and Dretske between them have solved the mind/body problem. The foundations of cognitive science would then be secure, and the philosophy of mind would have nothing left to worry about. (Except consciousness.)

This all seems rather promising, but of course it isn't free. According to the present treatment, lots of what are intuitively differences between the *contents* of thoughts turn out to be *syntactic* differences between thoughts of the same content. It turns out, for example, that 'H_2O' and 'water' are synonyms and that 'water is H_2O' is analytic, i.e., true in virtue of meaning (though not, of course, knowable a priori). To think that water is wet or that H_2O is wet is thus to think the same propositional content, albeit having the thoughts is being in different mental states. Pretheoretic intuition, not having considered the possibility that thoughts might differ otherwise than in their contents, is, no doubt, affronted. Myself, I don't know how much weight pretheoretic intuition can bear in such cases; content, synonymy, analyticity and the like are, after all, technical notions. If *all* that's wrong with a theory is that it affronts intuitions, perhaps the thing to do is get the intuitions fixed.

Anyhow, here's a soothing thought: if you can't get a *semantic* difference between the concept WATER and the concept H_2O, you can perhaps get the next best thing: despite their synonymy, the conditions for *having* the concepts are different. You can have H_2O only if you have the concepts H[YDROGEN], 2 and O[XYGEN]; but having the concept WATER requires none of this. (You can't have H_2O without H, because H is a syntactic constituent of H_2O; and concepts, since they are linguistic entities, have their constituent structures essentially.)

As I say, this all seems sort of promising; but we're about to have serious trouble. According to the suggested analysis, 'H_2O' and 'water' carry the same information and are therefore synonymous. The familiar examples of failures of substitutivity in contexts like 'has the concept . . . ' are explained by assuming that *content* identity is necessary but not sufficient for *concept* identity. This idea works—if it does—because, although Frege cases show that concepts that carry the same information are not always the same concept, at least Frege cases are compatible with a semantic constraint on concept identity that I'll call condition C:

C: *Concepts that carry the same information are always coextensive.*

J and O'S M are true of the same woman, and WATER and H_2O are true of the same stuff, so C survives both cases.

Suppose, however, that C were to prove unreliable. Then we could no longer pursue the strategy of claiming that informationally equivalent concepts are ipso facto semantically equivalent, and appealing to syntax to explain away apparent counterexamples. Since *semantically equivalent expressions must apply to the same things*, the reliability of C is a necessary condition for the reduction of content to information. If C fails, pure informational semantics fails too.

We are about to consider some examples where C does fail; you get informationally equivalent expressions which *don't* apply to the same things and therefore can't be synonyms. These examples are worrying in a way that mere Frege cases aren't. Frege cases suggest that informational semantics is insufficiently refined to be the whole story about conceptual identity, but they are quite compatible with conceptual identity being a *conservative extension* of informational identity. For all that the Frege cases show, 'carries the same information as' distinguishes fewer things than 'is the same concept as' (it's less 'fine grained'), but at least the latter respects all of the distinctions that the former draws. By contrast with the Frege cases, examples where C fails suggest that taxonomy by informational identity and taxonomy by extensional identity *cross-classify* the concepts. As far as I can tell, they imply that the theory of content can't be either purely informational or purely atomistic. Concessions will have to be made.

I'll argue, however, that the concessions that have to be made are harmless. Although informational semantics isn't strictly true, what's wrong with it doesn't threaten either Realism or Naturalism about meaning. And it doesn't invite Meaning Holism either.

Quine's Puzzle

How do we know that 'rabbit' refers to rabbits and not to *undetached proper parts* of rabbits (hereinafter urps)? Conversely, how do we know that 'undetached proper rabbit part' (hereinafter 'urp') refers to urps and not to rabbits? Call this question Q (for Quine and for convenience). I propose to answer Q presently, but some preliminary comments are required. These follow in no particular order.

1. For present purposes it's convenient not to distinguish the question whether 'rabbit' is *referentially* indeterminate between rabbits and urps from the question whether it is indeterminate between *meaning rabbit* and meaning *urp*. In effect, wherever it doesn't matter, I shall speak as though meaning determines reference; in particular, as though synonymous expressions are ipso facto referentially identical. I would be surprised if the argument proved to depend on this.

2. Q is about *content* individuation, not (just) *concept* individuation: 'rabbit' and 'urp' can't be synonyms because they aren't even coextensive; in fact, anything that either applies to thereby fails to satisfy the other. So, (mere) syntax won't answer Q; we can't, for example, exploit the fact that 'part' occurs in 'urp' in the way that we were able to exploit the fact that 'H' occurs in 'H_2O'. 'Water'/'H_2O' is arguably (just) a grain problem; 'rabbit'/'urp' is a cross-classification problem.

From this perspective, the examples of putative referential inscrutability that one finds in the philosophical literature are a mixed lot. Though no rabbit is an urp, I suppose that every rabbit is and must be an instantiation of rabbithood, and that nothing else can be. That is, 'rabbit' and 'instantiation of rabbithood' are necessarily coextensive. Because they are, the question why 'rabbit' doesn't mean *instantiation of rabbithood* is not crucial for an informational semantics in the way that, according to the present analysis, the question why 'rabbit' doesn't mean *urp* most certainly is. It would, for example, be open to an informational semantics to hold that 'rabbit' *is* synonymous with 'instantiation of rabbithood', the difference between them being not in what they mean but in the concepts they express.

3. Q looks to be an epistemological question, and the philosophical literature often takes it that way. But I don't. The question I propose to answer is metaphysical and *not* epistemological. It's something like: *What, if anything, makes it the case* that 'rabbit' refers to rabbits and 'urp' refers to urps? On what, if anything, does this difference in reference supervene? Epistemological considerations have no status in metaphysical inquiries according to my religious principles. I stress this because, according to the answer that I'll give, that 'rabbit' means *rabbit* rather than *urp* in Smith's mouth depends, inter alia, on what inferences Smith accepts. And one might wonder *how one would tell* what inferences Smith accepts, given, as it might be, facts about the (e.g., verbal) behaviors that Smith emits. Or how a 'radical translator' or a 'radical interpreter' could tell, consonant with the constraints that define their epistemic positions. Wonder what you will, of course, but for present purposes I have no interest in these questions. I assume that there are facts about what Smith (and others) are prepared to infer from what. I propose to appeal to such facts freely in what follows.

4. 'Rabbit' and 'urp', though not coextensive, are nevertheless invariably *coinstantiated*; every rabbit has and must have undetached rabbit parts, and every urp must be undetached from some or other rabbit. It is therefore true in this and every other possible world, that a situation is one in which rabbithood is instantiated iff it's one in which urphood is. A fortiori, any event that contains the information that either is instantiated contains the information that the other is instantiated too. I conclude that no purely informational semantics can distinguish the meaning of 'rabbit' from the meaning of 'urp'.

I can't prove this, of course; it depends on what one's notion of information is, and who knows what notions of

information may still await discovery? But I do think we'd better assume it. In fact, I think we'd better assume that no purely *externalist* semantics can prefer "rabbit' means *rabbit*' to "rabbit' means *urp*'. Here's why: Externalist semantics has only two ways to distinguish between expressions for properties that are locally coinstantiated. When they are *not* coextensive it does so by appealing to counterfactuals; in effect, by finding a possible world in which only one of the expressions is satisfied. If all and only the rabbits in our world have rabbit flies, and if 'rabbit' nevertheless means *rabbit* and not *rabbit fly*, then there must be some *other* world where the rabbits come without the flies or the flies come without the rabbits. By contrast, if symbols that are coinstantiated in point of conceptual or metaphysical necessity are also necessarily coextensive ('triangular' v. 'trilateral'; 'water' v. 'H_2O'; 'rabbit' v. 'instantiation of rabbithood'), externalist semantics bites the bullet, assumes that they are synonymous and distinguishes them by their syntax, as previously explained. But though 'rabbit' and 'urp' are *not* coextensive (and hence, a fortiori, are not synonymous), they are nevertheless invariably coinstantiated; there *aren't* any worlds in which one but not the other is satisfied. Pure externalism has, therefore, no resources left to cope with them.

5. I think that Q has an answer, but I'm leaving it open whether every similar question does. That is, I'm leaving it open that there may be *some* referential indeterminacy left over when all the metaphysical facts are in. That it is sometimes indeterminate whether 'x' refers to *x*s wouldn't entail that there aren't *any* facts about what refers to what; it wouldn't entail that reference isn't real.

It is, for example, perfectly OK for someone who is agnostic about whether it's determinate whether number words denote sets to hold nonetheless that *of course* 'rabbit' denotes rabbits. Correspondingly, that it tolerates indeterminacy is not *per se* an objection to informational (or, indeed, any other) semantics; a theory that is true tolerates whatever there is. Quine's question is embarrassing because it suggests that an informational semantics tolerates indeterminacy *where it seems intuitively obvious that there isn't any.* It seems intuitively obvious that 'rabbit' means *rabbit* and not *urp,* and this seems, prima facie, to be an intuition that informational semantics can't capture. If the intuitions were that the reference of 'rabbit' is indeterminate, then its failure to answer Q would argue *for* an informational theory.

6. I take it that Q is a question about reference *rather than truth.* I take its implication to be that the predicate 'is a rabbit' is indeterminate between meaning *is a rabbit* (hence being satisfied by rabbits) and meaning *is an urp* (hence being satisfied by urps), and that this is so *even if the truth values of all the sentences that contain that predicate are fixed.* Specifically, if Q is unanswerable, then any English sentence of the form 'a is a rabbit' is equally legitimately analyzed as being true iff some individual designated by 'a' is a rabbit, or as being true iff some individual designated by 'a' is an urp. On this reading, Q *grants* the syntactic notions *term of L, sentence of L* and *predicate of L* and suggests that the extensions of its predicates and the reference of its terms are underdetermined by the truth values of L's sentences. Q expresses the intuition that there is something wrong with the semantics of terms and predicates—i.e., with the referential semantics of the syntactic constituents of sentences—even if there is nothing wrong with the idea that sentences *have* syntactic constituents or with the idea that they have determinate truth values.

I will therefore take the syntactic notions *sentence, predicate* and *term* and the semantic notion *truth value* for granted in what follows.

7. I propose to consider Q in the following, austere form. Imagine a game involving two linguists (Ling1 and Ling2) and an informant (Inf) who speaks a language L. For convenience, let L be English, though nothing turns on this. Ling1 says that 'rabbit' means *rabbit,* Ling2 says that it means *urp.* The game consists of their attempts to defend these theses in face of the data that Inf provides. In particular, the linguists are allowed to specify any pair they like of a sentence of L and a possible situation, and Inf will tell them whether he takes the sentence to be true in that situation. Inf is, in effect, the embodiment of a (partial) function from situations and sentences of L to truth values. So, for example, given the data Inf provides, the linguists know that Inf holds 'there's a rabbit' true in a situation iff he holds 'there's an urp' true in that same situation.

I further assume that the linguists are given the semantics of the sentential connectives of L for free. (Remember: the intuition behind Q is that the semantics of *sub*sentential expressions is indeterminate even if the semantics of the sentential expressions is fixed. There is supposed to be a metaphysical problem about reference *over and above* whatever metaphysical problems there may be about truth.)

Finally, in order to make it absolutely clear that the issues about to be discussed aren't epistemic, I assume that the linguists' access to their data is unlimited and that the informant is always *right* about which sentences in his language express truths. In effect, I read Quine as betting that even a linguist who knows which sentences *God* holds true couldn't distinguish an ontology according to which 'is a rabbit' applies to rabbits from an ontology according to which it applies to urps. I'm betting that Quine is wrong to hold this.

OK; here's how the game is played. When Inf judges sentence S to be true in situation N, Ling1 must show that S's being true in N is compatible with 'rabbit' meaning *rabbit* in L. For example, if 'rabbit' is a term in S, then there must be a rabbit in N for it to refer to; if '(is a) rabbit' is the predicate of S, there must be a rabbit in N for it to apply to. And so on. Similarly, mutatis mutandis, Ling2 must show that S's being true in N is compatible with 'rabbit' meaning *urp*.

Here's how the game is scored: To find a datum that Ling1 can cope with and Ling2 can't would be to answer Q, so if there is such a datum, I win. If both linguists can cope with all the data, reference is inscrutable and Quine wins. If Ling1 fails and Ling2 doesn't, then 'rabbit' determinately means *urp*, nobody wins, and it's the end of the world.

I'm about to propose what I take to be a winning strategy for Ling1. I'm going to do this, however, in two trips; first, I'll suggest an answer to Q which, though it looks promising, turns out on inspection not to work. I think the way that it fails is illuminating and justifies the indirection. I'll then say what I take the right answer to Q to be and what morals it has for informational semantics.

First Fling at Q

Here's a gambit Ling1 might try. Suppose, for reductio, that 'is a triangle' means *is an undetached proper part of a triangle* and 'is a square' means *is an undetached proper part of a square*. And now, consider the situation illustrated in figure 3.1, where a square overlaps a triangle at a point A. A is a proper part of a triangle, so Ling2 predicts that Inf accepts 'A is a triangle'; A is a proper part of a square, so Ling2 also predicts that Inf accepts 'A is a square'. But, presumably, here and elsewhere, Inf accepts 'A is a triangle' only if he *rejects* 'A is a square' and he accepts 'A is a square' only if he rejects 'A is a triangle'. This seems to show that either 'is a triangle' doesn't mean *is a part of a triangle* or 'is a square' doesn't mean *is a part of a square*. Or, of course, both. Parallel arguments would show that Inf doesn't mean *urp* by 'rabbit'.[3]

This is, first blush anyhow, an attractive line of argument. After all, *being a square* and *being part of a square* are different properties; if they weren't, we wouldn't be having our present difficulties. Since they are different properties, it's not implausible that there should be some other property P such that

A thing's instantiating P is incompatible with its being a square, but compatible with its being a part of a square.

Being a triangle will do in the present case since, though nothing is both a square and a triangle, some triangles are parts of squares. If Inf's behavior signals that a thing has P,

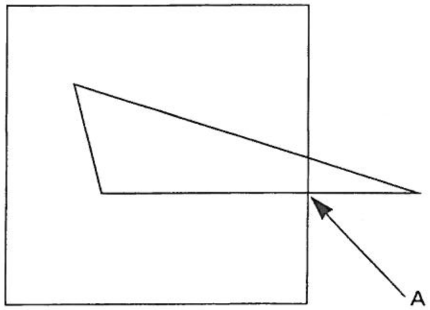

Figure 3.1
A triangle intersects a square at point A. See text.

it *thereby* signals that the thing may be a square part but can't be a square. The rest is duck soup.[4] It seems, at a minimum, that actually *showing* that Inf's ontology is indeterminate between *so and sos* and *such and suches* requires showing that there isn't a property that is situated with respect to *so and soness* and *such and suchness* in the way that being a triangle is situated with respect to being a square and being a part of a square. For all I know, however, there always are such properties; in which case there is no referential indeterminacy.

As I say, this seems, first blush, an attractive line of argument. But, on second thought, it begs the question against Ling2 and Quine. The tactic was to construct a situation in which Ling2 is forced to make a prediction that is contrary to fact—viz., that Inf will accept both 'A is a triangle' and 'A is a square'. That this prediction *is* contrary to fact is supposed to follow from two assumptions: first, that Inf is an informant about English, and second, that Inf accepts only sentences that are true in English. But it does *not* follow from these assumptions. You need also that 'A' unambiguously names A in both sentences; or, if you like, that the individual named by 'A' in 'A is a triangle' is the very same individual that is named by 'A' in 'A is a square'.

Let us pause to recapitulate.

Recapitulation

We could rule it out that 'triangle' means *triangle part* (mutatis mutandis, that 'rabbit' means *urp*) if we could establish that Inf rejects at least one of 'A is a triangle' and 'A is a square' in the situation diagramed by figure 3.1. To show that, we need:

i. Inf is a truth teller

ii. 'A is a square' is true only if 'A is a triangle' is false.

To get ii, however, we need iii *and iv*:

iii. 'is a square' means *is a square* and 'is a triangle' means *is a triangle*.

iv. 'A' unambiguously names A.

We get i–iii by assumption. But where does iv come from?

The moral seems to be something like this: Its being a fact that 'A is a triangle' means *A is a triangle* in Inf's mouth crucially depends on its being a fact that 'A' in Inf's mouth unambiguously names A. But for 'A' to be unambiguously the name of A is, at a minimum, for every token of 'A'*to name the same individual* as every other. So, the most that the first-fling argument could show is that we can rule out Ling2's ontology *if we can determine when two of L's expressions name the same thing*. Perhaps that's progress, but it's surely a good way short of answering Q. An inquiry into the ontology of reference must not presuppose the notion of *co*reference.

We are now well situated to see why it is so natural to despair of answering Q. Let us, therefore, reformulate.

Reformulation

It's untendentious that the data Inf supplies constrain the linguists in the following way:

v. If L takes 'is F' to mean *is F*, then, if there is in situation N an individual which is F according to L's ontology, L must predict that Inf takes 'is F' to be satisfied in N.

Thus, Ling1, who takes 'triangle' to mean *triangle*, must predict that Inf takes 'is a triangle' to be satisfied in figure 3.1; and so too, mutatis mutandis, must Ling2, who takes triangle to mean *triangle part*. But, of course, both 'is a triangle' and 'is a triangle part' *are* satisfied in figure 3.1; so v doesn't suffice to rule out Ling2's deviant ontology.

What *would* do the trick, however, is vi.

vi. If L takes 'is F' to mean *is F* and 'is G' to mean *is G*, then, if there is, in N, an individual which is both F and G according to L's ontology, L must predict that Inf takes 'is F' and 'is G' to be satisfied *by the same individual* in N.

That vi is stronger than v is evident on the face of it; and figure 3.1 shows that vi can't be met on the assumption that 'triangle' means *triangle part*. But we can't enforce vi unless we know *not just which expressions Inf takes to be satisfied, but which expressions he takes to be satisfied by which individuals*. That is, we can't enforce vi unless we already know a lot about how Inf is ontologically committed. So it seems that v is too weak, and vi begs the question. Dilemma.

To break this dilemma, we need some premise which, if granted, would license the inference from 'Inf takes F to be satisfied in N' and 'Inf takes G to be satisfied in N' to 'Inf takes F and G to be satisfied *by the same thing* in N'. The thesis that reference is inscrutable can now be seen as the claim that only question-begging premises could warrant such inferences.

So much for reformulation. We are now in a position to answer Q.

Reference Scrutinized

Suppose we translate 'A is a square' and 'A is a triangle' according to the deviant scheme, *leaving it open* whether the 'A's are coreferential. In effect, we know that what he accepts commits Inf to *some x satisfies 'is a square'* and to *some y satisfies 'is a triangle*', but we don't know whether he is committed to $x = y$. Is there any further fact about what sentences Inf accepts that could tell us that he is so committed?

Well, yes. If 'A' is an ambiguous name, then 'A is a square' and 'A is a triangle' can both be true in the same situation. *But 'A is a square and a triangle' can't be.* To put it the way linguists do, you can't conjunction reduce across a referential (or other) ambiguity. So, then, if we know that Inf accepts the inference from 'A is F' and 'A is G' to 'A is F and G', we *thereby* know that 'A' is referentially unambiguous in the premises. This isn't a special property of conjunction. Parallel remarks apply if, for example, Inf accepts the inference from 'Neither A is F nor A is G' to 'A is neither F nor G'. Etc.

In English and every other language I've heard about, the semantic function of predicate structures like 'is A (connective) B' is to insure that, when the predicate is evaluated, 'A'and 'B' are applied *to the same individual*. So, for example, if the connective is a conjunction, the predicate is satisfied iff the same individual satisfies all of the conjuncts. Knowing whether Inf accepts an inference from sentential to predicate conjunction thus gives one the same sort of information that one gets from knowing which individuals Inf takes to be identical and/or which names he takes to be unambiguous. So, if we know that Inf accepts the inference from:

(1) 'A is a triangle'

and

(2) 'A is a square'

to

(3) 'A is a square and a triangle'

we don't need further premises to decide whether 'square' means *square part* and 'triangle' means *triangle part*. Inf means *square part* by 'square' and *triangle part* by 'triangle' only if he takes it that 'is a square and a triangle' is satisfied in figure 3.1. But, by assumption, Inf *never* holds 'is a square and a triangle' to be satisfied, either in figure 3.1 or elsewhere.[5] So Inf doesn't mean by 'square' and 'triangle' what Ling2 says he does. So far so good.

We now have is what is called in Quinese an 'imminent' (as opposed to a 'transcendent') solution to the problem raised by Q. In effect, we can reject the deviant ontology for a language *in which we can identify such constructions as, e.g., predicate conjunction.* Since we know that 'is ——— and ———' is the construction that expresses predicate conjunction in English, we can reject the hypothesis that 'triangle' means *triangle part* in English. That the deviant ontology of L is excluded by determining which sentences with predicate connectives Inf holds true is, I think, not without interest; you might have supposed that it can only be excluded by determining which sentences that express identities Inf holds true. Quine says things about the inscrutability of reference that suggest that he does think this.

We're not, however, out of the woods. Here's the problem. In effect, we have it that if 'is ——— and ———' means predicate conjunction in English, then 'is a triangle' doesn't mean *is a triangle part* in English. And, of course, 'is ——— and ———' *does* mean predicate conjunction in English, so the argument goes through. Where, however, does the minor premise come from? On what does the fact that 'is ——— and ———' means predicate conjunction itself supervene? What we want is that it should supervene *on facts that are fully specified when one says which sentences Inf holds true.* Otherwise it's left open that the intuition that 'is ——— and ———' means predicate conjunction is itself a product of such ontological intuitions as that 'rabbit' refers

to rabbits and not urps. In which case, appealing to the first intuition in support of the second merely begs the question.

This is, course, a characteristically Quineian style of polemic: You could fix the ontology of English given the intuition that 'same' means *same* since the satisfaction conditions for 'same rabbit' are different from the satisfaction condition for 'same rabbit part'. But whether 'same' means *same* is itself up for grabs unless the ontology of L is determinate. In particular, if Quine is right, it isn't determined by facts about what sentences Inf holds true; viz., by facts of the kind that, by the rules of the game, constitute the linguist's data. And the rules of the game aren't gratuitous. If we were to assume that the linguists know not only what expressions Inf takes to be satisfied but also what he takes them to mean or which individuals he takes to satisfy them, we would be taking for granted precisely the semantic facts whose status we are attempting to determine.

Short form: Q is answered if we can identify predicate conjunction (and the like) in L just on the basis of which sentences in L Inf holds true. Well, can we?

No. But what we can do is just as good for the purposes at hand. We can identify predicate connectives if we know which sentences Inf holds true *and what inferences he is prepared to draw from them.* The details would likely enough be quite complicated, but the basic idea is clear enough: A predicate connective '*' is predicate conjunction if(f?):

Inf always takes sentences of the form 'A is F * G' to imply the corresponding sentence conjunction 'A is F and A is G'; *and*
whenever Inf is prepared to accept 'A is F * G', he is prepared to infer 'A is F * G' from 'A is F and A is G'.[6]

Notice that it is left open that there may be cases where Inf accepts 'A is F and A is G' but is *not* prepared to infer 'A is F * G'. Intuitively, these are the cases where 'A' is ambiguous in the premises (or where 'F' or 'G' is ambiguous between the premises and the conclusion).

The moral, to repeat, is that if you want to know which structures in L are predicate conjunctions, you need to know not just which sentences L-speakers hold true, but also (some things about) which inferences L-speakers are prepared to draw. This should seem unsurprising on reflection since it is entirely plausible that if you want to know which structures in L are *sentence* conjunctions, you also need to know (some things about) which inferences L-speakers are prepared to draw.[7]

A word about inferring, since it has now begun to loom large: The metaphysical context of this entire discussion has been a certain naturalistic framework within which semantic notions are reconstructed in, roughly, causal/nomological terms. In that framework, it's reasonable to assume that inferring is fundamentally a matter of *causal relations among sentence tokens*. In particular, *there is some (probably quite complicated) causal/nomological relation (call it CN) such that Inf infers 'S' from 'P' if(f?) Inf bears CN to (ordered) pairs of tokens of 'S' and 'P'*. No doubt CN will involve not only Inf's actual causal history but also his dispositions with respect to merely possible tokens of the types 'S' and 'P'. So be it.

Many philosophers will find this sort of treatment of inferring tendentious, not to say unbearable. Ah, well. My point is that it's natural for a naturalist to assume it, and that *he does not beg question Q by doing so*. The reason it's un-question-begging is that we have Turing's assurance that inferring is a computational process, hence that CN is a relation that sentence tokens enter into just in virtue of their syntax. We don't have to determine that Inf means *a is F* by 'a is F'—indeed, we don't have to know *anything* about what Inf means by 'a is F'—to determine that he infers 'a is F' from 'a is G'.

So, then, assuming that inferring is naturalized by invoking CN, we get the following account of predicate conjunction. '*' means predicate conjunction in Inf's mouth if(f?) R:

R:

(i) Inf bears CN to <<'a is F * G'>, <'a is F and a is G'>> whenever he accepts 'a is F * G'. (In effect, Inf is prepared to infer a sentence conjunction whenever he accepts the corresponding predicate conjunction;) and

(ii) If Inf accepts 'a is F * G', then he bears CN to <'a is F and a is G', 'a is F * G'>. (In effect, Inf is prepared to infer any predicate conjunction he accepts from a corresponding sentence conjunction.)[8]

My claim is that R characterizes predicate conjunction in L in terms of sentence conjunction in L, and that it does so *without* assuming that the terms/predicates of L are unambiguous. In particular, R determines predicate conjunction in L relative to a specification of the syntax of its sentences, the truth values that Inf assigns to them, and the inferential cum causal relations among their (actual and possible) tokens. I also claim that this determination is transcendent (it works for any language that has both predicate conjunction and sentence conjunction) and that it does not, in and of itself, beg the question against referential inscrutability.

And, given an un-question-begging characterization of predicate conjunction, we have an answer to Q. For, as we've seen, the facts about which conjoined predicates Inf accepts rule out the ontology according to which 'triangle' means *triangle part* and, mutatis mutandis, they rule out the ontology according to which 'rabbit' means *urp*.

So much for Q. For all Q shows, reference is scrutable after all.

The Cost of Scrutability

Ernie LePore and I used to go around asking philosophers the following sort of question: 'Imagine a language that doesn't have an expression that translates our word 'animal'. Could it have an expression that translates our word 'rabbit'?' (If this question doesn't grab you, try: 'Imagine a mind that hasn't got the concept ANIMAL; could it have the concept RABBIT?') We were interested in such questions because it seemed to us that if the answer is 'no,' then there might well be a slippery slope to the conclusion that no language could translate *any* English expression unless it could translate *every* English expression. For a variety of reasons that we set out in our book *Holism* (1992), this is a conclusion one might well wish to avoid.

To put it in slightly other terms, it seemed to us likely that either translation is an *atomistic* relation, so that what translates an expression of L is independent of what, if any, other expressions L contains; or translation is a *holistic* relation, so that what translates an expression of L depends on *all* the other expressions L contains. We saw no stable middle ground short of wholesale appeals to the analytic/synthetic distinction, which, following Quine, we took to be a Very Frail Reed.

That all this should be so is, after all, exactly what informational semantics predicts. What information 'rabbit' carries depends on whether and in what way 'rabbit'-tokens covary with instantiations of rabbithood. There being such covariation is presumably metaphysically independent of any other symbol-world relations, so information, as such, is

purely atomistic. Accordingly, if meaning is information and translation preserves what an expression means, translation should be atomistic too. Pure informational semantics thus entails that whether a language contains an expression that translates the English word 'rabbit' is independent of any *other* facts about its expressive power.

It seems, however, that pure informational semantics is wrong to say this. 'Rabbit' determinately means *rabbit* and determinately doesn't mean *urp*, and the previous discussion suggests that this depends, in fairly intricate ways, on the logico-syntactic apparatus that English makes available to its speakers. I say it *suggests* this. The most it could actually *show* is that the distinction is supported when the apparatus is intact. It's left open that there could be some *other* way for 'rabbit' to determinately mean *rabbit* rather than *urp*; some way that's compatible with semantics being strictly atomistic. I admit, however, to not having a clue what this other way might be, and I am therefore prepared to concede that the cost of 'rabbit''s referential scrutability is that semantics isn't strictly atomistic and hence that it isn't strictly informational.

Never mind. Even if a language that can translate 'rabbit' has to have predicate connectives, it doesn't follow that it has to have an expression that can translate 'animal'. So, even if the metaphysics of referential determinacy shows that semantic atomism is strictly false, it's still wide open that you can say *rabbit* (and/or think RABBIT) even if you can't say (and/or think) *animal*. The reason is that, so far at least, the inferential apparatus that makes 'rabbit' referentially scrutable is exhaustively "logico-syntactic"; we've found no reason to suppose that it infects the *non*-logical vocabulary.

Here's how you run a slippery slope argument for semantic holism. First you get a guy to admit that nothing in a language that can't say *animal* could mean what 'rabbit' does in English. Then you ask about *carrot*; if, after all, meaning *rabbit* depends on accepting the inference from *is a rabbit* to *is an animal*, why doesn't it also depend on accepting the inference from *is a rabbit* to *likes to eat carrots*? What principled difference could make one of these inferences meaning-constitutive and the other not? 'Gee, I don't know,' your interlocutor replies, having read Quine. Holism follows.

My present point is that you can't run this line of argument if there *is* a principled difference between the meaning-constitutive inferences and the rest; and, for all that the metaphysics of scrutability shows, there perfectly well may be. It looks like various pieces of logical syntax have to be in place for Inf to mean *rabbit* rather than *urp*. And, no doubt, terms in the logico-syntactic apparatus are largely *inter*defined. Presumably a language that didn't have 'not' couldn't have 'if', and maybe a language that didn't have sentence conjunction couldn't have predicate conjunction. But there is no reason at all to suppose that the logico-syntactic vocabulary is itself interdefined with the *non*-logical vocabulary. So, even on the assumption that having *rabbit* requires having predicate-*and*, there is no reason to suppose that having *rabbit* or having predicate-*and* requires having *animal*. For all that the metaphysics of scrutability shows, you *can* have *rabbit* without having *animal*; structuralists in linguistics and conceptual role semanticists in philosophy to the contrary notwithstanding.

Semantic atomism is the idea that the meaning of your words—mutatis mutandis, the contents of your thoughts—is metaphysically independent of the inferences you are prepared to draw. In this sense, it's the idea that semantics isn't part of psychology. I think that the puzzles about scrutability show that semantic atomism is probably false in this strong form. The ontology of a language supervenes not on mind/world connections alone, but on mind/world connections *plus logical syntax*. Having said this, however, it's important to add that, for most philosophical purposes, it doesn't matter a damn. It doesn't imply holism about meaning, it doesn't imply that the conditions for a term's meaning what it does are other than well-defined, and what it tells us about naturalism in semantics is only something that we already knew; viz., that the program fails unless there is a naturalistic account of inferring. Since inferences are surely part of the causal structure of the world, this is true *whether or not they are constitutive of meaning*.

Yes, but What Does 'Gavagai' Mean?

Suppose that the distinction between an expression of L meaning *urp* and its meaning *rabbit* depends, metaphysically, on the logical syntax of L. And suppose that the linguistically interesting facts about a certain informant are exhausted by this: He accepts 'Gavagai' when and only when he is visually stimulated by rabbits. (More precisely, when and only when he bears to rabbits, and hence to urps, whatever relation your semantics says is constitutive of carrying information about rabbits, and hence about urps.) What does 'Gavagai' mean in this informant's mouth? Thinking about this question is a way of finding out how far we have, and how far we haven't, departed from a strictly informational semantic theory.

The question has, I think, a perfectly good answer. But I can't tell it to you, and I'm afraid you wouldn't like it if I could. The reason I can't tell it to you is that, to put it very approximately, 'Gavagai' means *gavagai*, and that it does is not something that you can say in English.

Notice, to begin with, that there is no problem about saying what information 'Gavagai' carries in the informant's mouth. Let P be the property that something has iff it instantiates rabbithood[9] or any property that is necessarily coinstantiated with rabbithood. Then Inf's utterances of 'Gavagai' carry the information that P is instantiated; and it's precisely things that instantiate P that the expression applies to in Inf's dialect.

But, of course, 'Gavagai' doesn't *mean* P, assuming that meaning is what translations are supposed to preserve. The trouble is that translation works like indirect quotation and de dicto belief ascription; all three require the preservation not just of content (i.e., information) but also of appropriate relations among modes of presentation. Just which such relations *are* appropriate depends, I think, on the purposes at hand in a given case. For that reason, I doubt that rigorous conditions for translation, indirect quotation or de dicto belief ascription can be formulated (see Fodor 1992).

The problem about translating 'Gavagai' into English is that the only modes of presentation of the property P that English affords are long and disjunctive; and, of course, there is no reason at all to suppose that Inf has anything long and disjunctive in mind when he says 'Gavagai' in the situation we have been imagining. Presumably, what he has in mind is just GAVAGAI.

If, as I've been suggesting, distinguishing rabbits from their undetached parts depends on having access to constructions like predicate conjunction, then speakers of Gavagese can't refer either to rabbits or to their undetached parts; the best they can do is refer to things that instantiate P. So, we can't translate them, but we can refer to things that they can't. As a matter of fact, I don't suppose there are any languages, or minds, that can express P but don't have predicate connectives and the like. So I don't suppose that there are, as a matter of fact, any languages or minds that can't share our ontological commitments with respect to rabbits and rabbit parts, or whose modes of presentation we can't, at least roughly, approximate in translations. But this is at best a *contingent* truth according to the present account. Nothing about the metaphysics of meaning or of reference guarantees it. In ways that pure informational semantics does not, the mildly mixed view at which we have now arrived tolerates the possibility of minds and languages whose ontological commitments are inscrutable to us and to which our ontology is equally obscure.

Qua speaker of Gavagese, Inf is so situated that he can't mean or refer to what we use 'rabbit' to mean and refer to; and it's true that no purely informational semantics—indeed, no purely atomistic semantics—can account for this. But nothing metaphysically important follows; in particular naturalism about semantics is unimpugned. For all that has been shown so far, the meaning of 'rabbit' is fully determinate, and the conditions for referring to rabbits can be exhaustively and precisely specified in nonintentional and nonsemantic vocabulary. *That we can't translate Inf and Inf can't refer to rabbits does not, therefore, make intentional psychology a philosophically interesting science.*

In which case, who cares whether atomistic semantics is literally true? Not me, I assure you.

The 1993 Jean Nicod
Lectures

The Elm and the Expert

Mentalese and Its Semantics

Jerry A. Fodor

1994

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England