CHAPTER 14

The concept of a symbol

The purpose of this chapter is to show that CTM involves a spurious understanding of what symbols are and that, when those misunderstandings are corrected, many of the classical objections to CTM are vindicated.

If a coffee-machine accidentally produced the sounds "snow is white", nothing would have been said; no symbols would have been tokened. Of course, one might think that the coffee-machine was actually producing symbols. But one would believe this *only* in so far as one believed that, in making those sounds, the coffee-machine was actually animated by certain intentions (or that the coffee-machine had been created by somebody with certain intentions and that, consequently, somebody was speaking *through* the machine, much as one can speak *through* a letter or an email). The moment one accepts that the coffee-machine is inanimate and that its production of those sounds is simply an accident, and not the result of its creator's intentions, one immediately accepts that nothing has been said and that nothing linguistic has occurred.[1]

This suggests that the psychological underpinnings of a symbol-token are not just causally, but also *constitutively*, involved in that tokening. Even though it is imperceptible to others, one's mental state is constitutive of the symbol-token one is producing. When the symbol-type "snow is white" really is instantiated, a constituent of that instance is some kind of mental state. When somebody says "snow is white", the symbol-token that has been produced actually comprises at least some aspects of the mind of the person who produced that token.

Sometimes the psychological aspects of a symbol-tokening are spatiotemporally remote from its morphological aspects. Suppose that I am typing on a keyboard that forms letters out of giant rays of light in some remote galaxy. In that case, the morphological aspects of the token of "snow is white" that I produce are estranged by millions of years and miles from the psychological aspects of that same token. But we mustn't let this lead us astray. The psychological underpinnings of those remote galactic events are no less constitutive of that sentence-token than those galactic events themselves.

Symbols not strictly psychological entities

At the same time, psychology cannot single-handedly turn an otherwise dead sound into a living symbol-token. Suppose that, while in the grips of a psychotic delusion, Smith

1. See Searle (1984) for a similar line of thought.

utters complete gibberish, thinking that, in English, it means *snow is white.* In that case, nothing has been *said.* Of course, one might be able to figure out that Smith meant *snow is white.* But this would be a case of one person's having psychoanalytic or criminological insight into another. It would not be a case of two people speaking a common language.

Indeed, it would not be a case of anybody using any language. Let L be any language other than English. L may be a public language, or it may be a "private language" – a code that Smith invented for his own use. Smith's intention was not to say something in L – his intention was to say something in English. So even if the noise he made sounded just like the L-translation of "snow is white", his making that noise was an accident, and was thus no more an affirmation that snow is white than the coffee-machine's production of that same noise.

Here is a similar argument. Suppose that I try to produce the English words "snow is white" but that, solely because of a vocal chord malfunction and not because of a repressed urge to speak French, I end up producing the sounds "la neige est blanche." Witnesses would assume that I had spoken a French sentence. But they would be wrong. It may even be that I *do* speak French. But given that my French knowledge was not implicated in my saying "la neige est blanche", my producing those sounds was not meaningfully related to anyone's knowing French, and thus wasn't a case of French-speaking. Similar considerations apply to Smith's case.

Why any symbol-token necessarily has non-morphological components

In order for a symbol to be tokened, it is necessary that it arise in a certain way from a certain kind of social knowledge. (Later on, we will try to state more precisely the nature of this knowledge.) In this context, the word "necessary" denotes constitutive, not causal, necessity. So the just-described psychological and social conditions are veritable *constituents* of the symbol-token.

This doesn't conflict with our presumption that symbol-tokens, unlike social practices, are discrete entities. If our analysis is right, a symbol-token is a convergence of a number of different factors, and that convergence-point *is* discrete, even though the converging forces are not.

A symbol-token has morphological, psychological, and sociological components. Up to a point, but not entirely, the function of the morphological component of a symbol-token is to indicate the existence of the psychological and sociological components. Thus, there can no more be a symbol-token in the absence of either of the latter two components than in the absence of the first. So far as we think otherwise, it is a reflection of our tendency to telescope metaperceptual into perceptual data. This last point may require elucidation. We see a mere ink-mark. This is perceptual knowledge. We know that various social and psychological factors conspired to produce that ink-mark. This is meta-perceptual knowledge. If it weren't for this meta-perceptual knowledge, the ink-mark wouldn't be a symbol-token, at least not for us. So far as we think

that, in and of itself, that ink-mark constitutes a symbol-token, it is because we believe ourselves to be perceiving what, in fact, we are meta-perceiving. It is because we are conflating meta-perceptual with perceptual content.

Here it is crucial to keep in mind the distinction between saying and expressing. When I produce the sounds "snow is white", that symbol-token doesn't *mean* that such and such socio-psychological factors are now operative. Semantically, that symbol-token ascribes a certain color to snow, and that is all it does. But given that a symbol-token meaning *snow is white* has been produced, it can be inferred that certain psycho-social factors are operative. So that symbol-token *expresses*, but does not *state*, the existence of those factors.

Let us sum up. A symbol-token is what results when a bit of morphology is produced in consequence of a convergence of social and psychological factors. A symbol-token is thus best thought of as a convergence of psychological and sociological forces. (Later we will try to give a more precise definition of the terms "symbol" and "symbol-token.")

Of course, we instinctively gravitate towards the position that symbol-tokens are discrete physical objects – bits of ink and noise. In fact, that position may even seem a truism or a definition. But it is only because we labor under a number of misconceptions that we gravitate towards that view (I tried to identify one of those misconceptions a moment ago), and I think that the following comparison may help loosen the hold that those misconceptions have on us. (The comparison is one that is found in Wittgenstein (1958). Much of what we will say in connection with money resembles what is said there.)

What is money? Here is an answer to which we are instinctively drawn:

($) Dollar bills and nickels and quarters are money. They are also physical objects. They are bits of paper or metal. Money is created (in mints), and it can be destroyed. In the U.S. it is actually illegal to destroy money. If you burn a dollar bill, you are breaking that law. So a dollar bill is just a piece of paper. For exactly similar reasons, a nickel is just a piece of metal.

Though there are elements of truth in it, ($) is false on the whole; and it is worthwhile to say precisely why it is false, since there is such a strict parallelism between ($) and the conception of symbolhood that I am attempting to refute.

Let B be some particular dollar bill – e.g. the dollar bill that is in your pocket. Now imagine the following scenario. Tomorrow, the U.S. is taken over by a foreign power. A new system of government is imposed, and this involves changing the currency that is used. The bits of paper that Americans previously used to make purchases are now as useless as Monopoly money.[2]

2. In his novel *The Stand*, Stephen King describes a situation where the government collapses and dollar bills become (as King himself puts it) "as worthless as Monopoly money." I am borrowing his apt expression.

Even if B hasn't been damaged or otherwise altered, it is now completely worthless and therefore isn't money anymore. This shows that B *per se* never was money. What once had value, and was therefore money, wasn't B, but was rather B's being embedded in a certain way in a certain set of social practices. B was merely a constituent of something that had value, and was therefore only a *constituent* of bona fide money.

Up to a point, B's purpose was only to *indicate* the operativeness of the relevant social conditions.

To be sure, it would be an overstatement to say that B was *merely* an indicator of the presence of the just-described social factors. (Let us now assume that the political takeover earlier described never occurred, and that B remains valid currency.) Suppose that a copy of a legal contract is destroyed. If a duplicate is found, then the legal situation hasn't changed: *ceteris paribus* the judges, lawyers, and so on, are still going to do whatever they were going to do prior to the destruction of the first copy.[3] The physical copy of the contract is there *merely* to indicate that some kind of agreement has been entered into; its legal purpose is *merely* evidential. But if B is destroyed, its possessor has lost certain entitlements: he cannot get on the bus; he cannot enter the theatre or the ballpark. It is irrelevant whether B was destroyed in plain view of a judge or public notary.

But in this context what is relevant is that, since money has causal powers that a mere piece of paper does not, a piece of paper is, at most, a constituent of a unit of money, the other constituent being some political-legal apparatus. And up to a point, though not entirely, the function of the piece of paper is merely to indicate the operativeness of that apparatus.

Analogues of the points just made about money hold of symbols. Noises are partial constituents of symbol-tokens. If somebody tries to speak, but fails to make a sound, he has said nothing. But, by itself, a noise isn't anything linguistic. And when somebody manages to say something by making a certain noise, the function of that noise is, up to a point, to show something about the speaker's attitudes towards certain social rules. (We will develop this obscure point in Chapter 25.) Hopefully, then, our discussion of money illuminates, and also corroborates, our contentions about symbols.

Also, our discussion of money brings to light some principles that will enable us to neutralize an obvious criticism that might reasonably be made of our analysis of symbolism:

> You may be right to say that, as a matter of analytic necessity, there are genuine symbol-tokens only where there are socio-psychological factors of the sort you describe. But you are guilty of a blatant *non sequitur* in inferring from this that those factors are veritable *constituents* of symbol-tokens. As a matter of necessity, I couldn't have had different parents from the ones I actually had. But my parents are not in any sense *constituents* of me. The idea that a symbol is *just* a bit of morphology is thus perfectly compatible with the view (held by you and others) that symbol-tokens necessarily originate in socio-psychological conditions of a certain kind. Those

3. To simplify discussion, let us set aside all of the legal niceties relating to the authentication of duplicates of contracts.

> views are no more incompatible with each other than the obvious fact that people are distinct from their parents is incompatible with the thesis that no one other than Y and Z could have been X's parents if those two are in fact his parents. Your thesis that psychological and social conditions are actual *constituents* of symbols thus seems to be, at best, a proposal that redefine the term "symbol" (or "symbol-token"). Currently, that term refers to ink-marks and noises; and you are proposing that we have it refer to ink-marks (or noises) *plus* their social and psychological concomitants. What you've said in connection with symbols is false if taken as an analysis, and sterile if taken as a proposal to change our nomenclature.

Remember what we said a moment ago about money: we strip money of many of its causal powers if we identify it with mere bits of paper. Of course, we do not thereby strip it of *all* of its causal powers: but we do strip it of those causal powers that distinguish it from mere scraps of paper. Like money, symbol-tokens have causal powers. If we identify symbol-tokens with bits of ink or noise, then we strip them of many of their causal powers. Of course, we do not thereby strip symbol-tokens of *all* of their causal powers: but we do strip them of those causal powers that distinguish them from mere ink-marks and noises. If we say that a token of "snow is white" is *merely* a sound or ink-mark, then we strip that token of the causal powers that distinguish it from the barking of a dog or the grumbling of someone's stomach: we strip it of its distinctively *symbolic* causal powers. Because so many aspects of human thought and behavior are to be understood in terms of the manipulation of symbol-tokens, it isn't reasonable to take the position that symbol-tokens don't have any causal powers not also had by mere noises and ink-marks. Such a position would either nullify obviously correct analyses of human conduct or it would demand a tedious and unproductive rewording of those analyses. If we are to accommodate the fact that symbol-tokens have causal powers not had by mere ink-marks and noises, then we must take the position advocated earlier, and denied by the objector, namely: any symbol-token has a morphological *component*; but the function of that component is, at least in part, to *indicate* the presence of socio-psychological factors that are themselves components of that token.

The concept of morphological elasticity

Consider the following three statements (in this context, the term "ink mark" is short for "ink mark or noise or pattern of light..."):

(i) An ink-mark is, at most, one *component* of a symbol-token.

(ii) Whenever an ink-mark is a component of a symbol-token, its function is in large part to *indicate* the existence of certain socio-psychological conditions.

(iii) Two tokens of a single symbol-type may be morphologically very different from each other. More precisely, there is no limit to the extent to which two symbol-tokens may differ from each other in respect of morphology.

We've already seen that (i) and (ii) are true. And it is clear that (iii) is a consequence of (i) and (ii). It is also clear why (i)-(iii) pose a problem for CTM additional to those already discussed. CTM needs there to be a morphological characterization of certain semantic notions; it needs to bring semantic facts (e.g. the fact that one sentence entails another) into alignment with morphological facts. Supposing that our semantic rules permitted little or no morphological divergences between two tokens of a given symbol-type, such an alignment might be possible (though only to an extremely limited extent, given various well-known incompleteness-theorems). But given what symbol-tokens are, the *a priori* probability of there ever being such an alignment is vanishingly close to zero, and the conditions under which such alignments would occur would constitute obfuscations, not expressions, of the basic facts about what symbols are. Given that any symbol-token has non-morphological components, and that the function of the morphological component is (up to a point) to indicate the existence of the other non-morphological components, we cannot reasonably demand of a language that its morphological aspect remain stable. That would be like expecting the pressure of the gas in a sealed vessel to remain constant, while we raise its temperature. We cannot expect dependent variables to behave like constants or like independent variables.

CTM needs the morphological and semantic aspects of language to be in lockstep. For the reasons just given, the circumstances that permit such alignments are in the nature of singularities. There are situations where, within limits, a morphological characterization of certain semantic notions are possible. These are cases where what is in fact the dependent variable (morphology) *appears* as though it is one of independent variables (psychology and sociology), thereby creating the misleading appearance that the socio-psychological components of symbol-tokens are to be understood in terms of their morphologies. Such situations thus doubly obscure the nature of symbolism, as they invert the structure characteristic of symbol-tokens, and then proceed to fuse the ordinarily distinct entities constitutive of that structure. In taking morphology as primary, CTM erroneously takes such singularities as its paradigms, and erroneously regards non-singularities as deviations or irregularities.

Some fiction may help clarify my point. In the year, 2025, human beings lose the ability to speak and to write. They cannot generate *new* sounds or *new* ink-marks. Fortunately, there are still tokens of every letter in the alphabet, and people can form new utterances out of these. Unfortunately, there is only one token of each letter. All of these tokens are housed in a universal letter depository. (Since these tokens are all made of marble, they cannot be bent or warped.) So if I wish to propose to my fiancé, I must go with her to this letter depository and then physically assemble a set of marble letter-tokens in such a way that there is a token of the sentence WILL YOU MARRY ME?

Under *these* circumstances, the morphological elasticity associated with symbols of actual languages would be entirely absent. There would be only *one* way to say KARL PEARSON DESERVES CREDIT FOR MANY OF THE INSIGHTS ATTRIBUTED TO THE LOGICAL POSITIVISTS or MARTIN SCORSESE'S MOVIE "CASINO" IS GROSSLY UNDERRATED.

Under *these* circumstances, thanks to the complete petrifaction (literally) of symbol-morphology, it *would* be possible to give a strictly morphological characterization of at least some semantic notions. (For reasons given earlier, the "automatization" or "mechanization" of thought would be no more possible under these circumstances than under our own. But here we are setting this issue aside, so as to develop a new and independent criticism of CTM.) But notice how much artifice is needed to create the needed alignment of morphology and semantics, and notice how unstable that alignment would be. (It would vanish the moment somebody created new letter-tokens.) This suggests that it is not in the nature of symbol-tokens to be amenable to such an alignment, and this in turn confirms our thesis that symbol-tokens are not mere bits of morphology. We've already discussed why this last fact is problematic for CTM.

Let us now develop a point made earlier. It is obvious that, given any physical object O, and given any meaning (or referent) M, there is some possible language L, such that O (or, in any case, some object morphologically identical with O) means M in L. But there is a stronger point to make. Given a language L, and given an object O that tokens some symbol of L, O's morphology, taken by itself, completely fails to fix the identity of the symbol-type tokened by O.

Suppose that Jones, Brown, and Aaron are all competent speakers of English. From a strictly acoustical standpoint, Smith's utterance of "we're having excellent weather today" may bear much less resemblance to Brown's utterance of that same sentence than it does to Aaron's utterance of "weird squirrels abound in that stack of hay."

Of course, it would be an overstatement to say that morphology imposes *no* constraints on semantics. Obviously it imposes crucial constraints: if you fail to enunciate, or if your calligraphy is poor, I won't know what you are saying. In fact, if your utterances result from an intention to produce the wrong kind of morphology, then your utterance isn't a case of speech at all.

Here is an illustration of this last point. Smith produces the sounds "snow is white." His intention is to say that grass is green. He says "snow is white" because, not having a good command of English, he thinks that it means *grass is green.*

If Smith has affirmed anything, it is either that grass is green or that snow is white. He hasn't affirmed that snow is white: he had the wrong intention. He hasn't affirmed that grass is green: he made the wrong noise. Therefore, Smith hasn't affirmed anything. So it is at least partially in virtue of its strictly morphological properties that a physical event or structure qualifies as a symbol-token.

Nonetheless, as we've seen, morphology only has an extremely partial role in the way of determining whether a physical object betokens this as opposed to that expression-type.

Why there cannot be too tight a connection between morphology and symbolhood

Too much of a connection between semantics and morphology is as threatening to the existence of a usable language as the absence of such a connection. I have probably never heard two utterances of "he's over there" that sound quite the same. The class of circumstances that permit a perfect instantiation of a given morphology is only a tiny subset of the class of circumstances where we wish to communicate the proposition associated with that morphology, and language would therefore be useless if there were too close a connection between morphology and meaning.

Consider the deposit of ink to the right of the upcoming colon: beggars can't be choosers. It is not often that we can produce a physical object that is morphologically just like that ink-deposit. But it is often that we wish to affirm the sentiment meant (in English) by that ink-deposit. Obviously it is easy to express that sentiment: our ability to do so isn't diminished by our inability to produce an object having a hyper-specific kind of morphology. This confirms our point that too much of a connection between morphology and meaning is as threatening to the existence of a viable language as an all-out absence of such a connection.

A corollary is that a certain plasticity is built into any aspect of a creature's functioning that has a genuinely symbolic dimension. To the extent that a creature's behavior is strictly morphology-based, it doesn't bear any significant resemblance to anything that is *uncontroversially* an instance of symbolic activity. Genuinely symbolic activity does, whereas strictly morphology-driven activity does not, allow for an unlimited degree of deviation between a creature's linguistic paradigms and its reproduction of those paradigms. Genuinely symbolic behavior has a plasticity not had by its mechanical analogues.

At first it might appear that, if indeed this last point is true, that is merely a reflection of an evanescent, if not already extinct, fact about our technology. These days one can buy scanners that have "character recognition." So if you scan a typed document, what is uploaded is confined to semantically relevant material, and excludes the semantically irrelevant material (e.g. coffee-stains, crushed insects, smudges) that tends to gather on documents. A scanner with character recognition can recognize the word "snow" in physical objects having very different morphologies. So such a scanner seems to have the plasticity that, according to what was just said, they lack.

Also, there are technologies that transcribe the spoken word. These technologies are currently imperfect. But the imperfections are rapidly being eradicated; and it is not unreasonable to expect that, in a few years, there will no longer be a need to type. Again, it seems that a technological fact counterexamples my analysis.

But this line of thought presupposes the very conception of symbolhood that it is meant to support and that we have seen reason to reject. Let M be some transcribing-device that has maximally good character recognition (in other words, suppose that M's character-recognition capabilities are as good as they can be, given that M's behavior is

strictly morphology-driven), and suppose that M has no trouble "recognizing" the words that Brown has just uttered, even though Brown has a medical condition that makes it impossible for him to enunciate properly. Even though it satisfies these conditions, M will *still* fail to recognize symbol-tokens that any English-speaker would, in virtue of being an English-speaker, easily recognize. The following scenario illustrates this.

X and Y are conversing with each other. They are both reasonable people of average or above average intelligence, and their various sensory modalities are in excellent working order. At some point in the conversation, Y asks X a question, namely: "what did Smith do today?" As X and Y both know, and as each knows the other to know, there are only two answers that, given Smith's personality and general situation in life, have any chance of being correct, namely: "he went to O'Reilly's [the local bar]" and "he spent the whole day in bed." Further, X knows that Smith went to O'Reilly's. Under the circumstances, it isn't necessary for X to produce a crisp, lucid utterance of: "he went to O'Reilly's." It is necessary *only* that X produce a sound that is obviously *not* a token of "he spent the whole day in bed." Such a sound could, and probably would, be little more than a grunt that has a certain intonation and arc. It could be that X is able to answer Y's question with a sound that, in purely acoustical terms, is much closer to a typical utterance of "we are now having high-tea" than it is of a typical utterance of "he went to O'Reilly's."

But S (the previously described scanner) would wrongly transcribe X's utterance. By hypothesis, S's character-recognition capabilities allow it to respond correctly to symbol-tokens that are morphologically extremely degenerate. But from a purely morphological standpoint, X's utterance is *so* degenerate that even S cannot recognize it. Even if S can recognize *that* particular utterance, there will be others that, for analogues of the reasons just given, it will not be able to recognize.

Also, supposing that *in that particular case* S is sufficiently sensitive to generate the right symbol-token, that same sensitivity will dispose it to assign the *wrong* symbol-token in other cases. There will be cases where S is given precisely the same acoustical data that it was given here, but where it should generate a token of "we are now having high-tea" or "Pete likes Green Sleeves." Because S has the sensitivity needed to generate the right symbol-token in *this* case, it will *ipso facto* be locked into tokening the wrong symbol ("he went to O'Reilly's") in these other cases.

To the extent that context-makes it clear what one is saying, it isn't necessary that morphology do so. Obviously morphology always has *a* role; context can never completely do the job of morphology. But there is no limit to how much context may marginalize the role of morphology.

A corollary is that, even if we focus on a single language and even if we leave aside cases of homonymy, a single morphology can instantiate any number of different expression-types.

A related corollary is that, for any symbol S and any morphology M that tokens S, we can imagine a morphology that is different from M that also tokens S. So supposing that $M_1 \ldots M_n$ are the various morphologies that, thus far, have tokened the sentence-type

"snow is white", there is a kind of morphology that is not included in that set that *could* token "snow is white." Thus, anything whose operations are strictly morphology-driven is *ipso facto* incapable of keeping pace with such context-generated variations. For these reasons, no significant parallel can be established between symbol-driven and morphology-driven activity.

Of course, there are several objections to be made to this line of thought:

> You may be right to say that whole sentences can be tokened with noises that bear little acoustical resemblance to the paradigms through which English is taught, and that symbol-morphology deteriorates as context assumes a greater role in communication. But even after we allow for the morphological deterioration characteristic of everyday speech, you chose a case of symbol-tokening that is *too* degenerate to be regarded as representative and that, consequently, must be regarded as a kind of singularity. So far as a hypothesis is counterexampled only by degenerate or singular cases, it isn't counterexampled at all. So, given that you can make your case only in terms of such a tendentious and artificial example, it follows that you have no case.

A symbol-token has both morphological and socio-psychological components, and the morphological component has done its job provided that it directs the auditor (or reader) towards the right social-psychological component. So there is nothing *linguistically* degenerate about X's utterance, even though it may fail to satisfy the aesthetic strictures of a diction-coach.

Of course, X's utterance is *morphologically* degenerate, as I myself said. But that is just an elliptical way of saying that *in most contexts* an utterance acoustically like X's wouldn't provide the information needed to direct the auditor towards the socio-psychological component of the relevant symbol-token.

There is no sense in which the story just told concerns a singularity or degenerate case. While it is true that, in the exchange between X and Y, context is doing more than it ordinarily would, and morphology is doing less than it ordinarily would, these quantitative differences don't constitute *structural* or *categorical* differences. We are dealing with a garden-variety expression of the forces involved in the composition of any symbol-tokening and, therefore, in any linguistic exchange. The dialogue between X and Y could be described as a "degenerate" or "singular" case only if it were assumed that any utterance that isn't in lock-step with the aesthetic strictures of a diction-coach is *ipso facto* degenerate. But in addition to being question-begging in this context, such an assumption would demand that we categorize virtually every utterance ever produced as degenerate.

Let us consider another objection:

> You are confusing statistical with dynamic facts. Technically speaking, there are no straight lines, rigid bodies, or smooth surfaces. But the physical world must be understood in terms of such idealizations. So even though they are *statistically* absent, they are *dynamically* or *explanatorily* present. *Statistically* there are very few

> utterances of "Smith went to O'Reilly's" that satisfy the stricture of a diction-coach. But, in this context, that is irrelevant. What is relevant is that, so far as they discharge successful speech-acts, the bits of morphology that we produce *approximate* to such paradigms. X's utterance of "Smith went to O'Reilly's" is to be understood in terms of a diction-coach's crisp utterance of that same sentence. Of course, what we just said about X's utterance is true in general. Utterances are understood only in so far as they are associated with the right paradigms. Such paradigms are therefore *dynamically* ubiquitous, even though they are *statistically* few and far between. You may be right to say that CTM insists on a much tighter connection between morphology and semantics than is found in everyday speech. But contrary to what you say, that doesn't show that CTM is wedded to an artificial, or otherwise erroneous, understanding of what it is for a physical object to be a symbol. On the contrary, it shows that CTM embodies a sensitivity to the relevant *dynamic* facts and, unlike you, doesn't naively derive semantics from statistics.

I agree with the objector on many points. Just as he says, literal meaning cannot be understood in statistical terms. (It should be pointed out that Kent Bach 1984 makes exactly this point.) "I would like a vodka tonic" literally means *I would enjoy having a vodka-tonic right now*, even though it almost always communicates the imperative: *bring me a vodka tonic*. It seems, indeed, that were *complex* expressions are concerned, there is practically no limit to how much literal and understood meaning may pull apart. (But where simple expressions are concerned, literal meaning must ultimately collapse into understood meaning.) That is why, as semanticists often point out, there are well-formed sentences that are *systematically* misunderstood, for example: "The man the boy the girl hit kissed moved."[4]

But the objector is misapplying this point, as we may see by contrasting the dialogue between X and Y with a situation that actually exemplifies the principle that the objector is discussing. Consider the sentence:

(a) "some number is bigger than every number."

(a) is false. But it is conceivable that English speakers might *take* those words to mean

(b) given any number x, there is some number y such that y is greater than x,

which is true, instead of

(c) there is some number x, such that given any number y, x is bigger than y,

which, being equivalent with (a), is false.

What somebody who wishes to affirm (b) should utter is:

(d) "there is no greatest number."

4. This example is borrowed from Fodor and Pylyshin (1988: 35).

There is indeed *a* sense in which utterances of "Smith went to O'Reilly's" should bear a greater resemblance to a diction-coach's pronunciation of that sentence than Smith's does. But it is not comparable to the sense in which somebody who wishes to affirm (b) *should* utter (d) as opposed to (a). If you utter the sentence "Smith went to O'Reilly's" with the intention of affirming the proposition *Smith went to O'Reilly's*, people are generally more likely to *know* that your speech-act is in keeping with the relevant linguistic rules *if* your utterance is more like a diction-coach's than it is like Smith's. It is easier to know what a diction-coach is saying than it is to know what is being said by somebody who is mumbling or has laryngitis. But this doesn't mean that the expressions produced by a diction-coach are more *correct*, relative to the rules constitutive of the language in question, than what is said by somebody who has laryngitis. It only means that *ceteris paribus* it is easier in the one case than the other to identify the speech-acts that have *in fact* been performed. So the sense in which one should speak like a diction-coach is comparable to the sense which one *should* speak directly into a microphone. The "should" here is of a generally instrumental nature, and has no specifically linguistic significance.

But supposing that I wish to affirm (b), the sense in which I *should* utter (d), as opposed to (a), *is* of a distinctively linguistic nature. In uttering (a), I am guilty of linguistic incompetence. But I am guilty only of garden-variety instrumental incompetence if I mumble (d) with the intention of affirming (b).

There is another objection to our analysis that we should consider:

> You start with a correct point and end with a false one. There are indeed many ways that a given symbol-type of a public language can be instantiated, and natural languages are thus characterized by a certain "morphological elasticity." As you point out, this makes it harder than it would otherwise be to bring morphology into alignment with semantics. But this is completely irrelevant to CTM. According to that doctrine, thinking consists in manipulating symbols of some *internal* language, not of some public and external language like English or Spanish. To immunize CTM from the criticisms that you've just put forth, we need only assume that this internal language ("Mentalese") is petrified, i.e. that where that language is concerned, there are strict limits as to how much two instances of a given symbol-type can differ from each other in respect of morphology.

This line of thought is a non-starter, the reason being that it is inherent in the nature of symbolism that morphologically different entities be capable of betokening the same symbol, and also that morphologically identical entities be capable of tokening different symbols.

Whether a given sound (or ink-mark or pattern of light....) betokens this as opposed to that symbol is a function, not of that sound's morphology, but of the relation between its morphology and the socio-psychological context of tokening. That is why two morphologically identical sounds (or ink-marks or patterns of light...) can betoken different words, and that is why two morphologically very different sounds (or ink-marks...) can betoken the same word.

Let T_1 be an occurrence of the word "bank" (financial institution), and let T_2 be a morphologically identical occurrence of the word "bank" (river's edge). Why do T_1 and T_2 betoken different words? In the one case, the context of utterance involved Smith putting a fishing pole into the back of his pick-up truck. In the other case, the context of utterance involved Jones angrily snapping his briefcase shut after unsuccessfully attempting to borrow money from his brother. Even though the morphology is the same, the relationship between morphology and context of utterance has changed; and that is why *two* words have been tokened, instead of one. Obvious extensions of this reasoning show why two different tokens of a *single* expression-type may be very different in respect of morphology.

This line of thought accommodates the fact that morphology has *some* constitutive involvement in the tokening of a symbol. For a given symbol to be tokened, it is necessary that there exist a physical object (e.g. a noise or ink-mark) whose *morphology* stands in a certain relation to the context. This means that there is no symbol (or symbol-token) where there isn't a bit of morphology. It also means that, all other things being equal, two tokens of a given expression-type will be morphologically similar.

But this same line of thought accommodates the fact that two tokens of a given expression-type can be morphologically very different. Given a sufficiently large change in the context of utterance, a commensurately large change in morphology will be needed to ensure that the *relation* between morphology and context is preserved. (If you gain two hundred pounds, the clothes that you wear must also change *if* they are to fit, i.e. if their relationship to your new physique is to coincide with the relationship had by your old clothes to your old physique.)

A symbol-type is instantiated in a given context iff there occurs an event whose morphology has a certain relation to that context. A symbol-token is thus an instance of a certain kind of *relation*: a relation between morphology and context. A symbol-type *is* such a relation. For reasons already discussed, the word "context", as it occurs here, is to be understood, at least up to a point, in social and psychological terms.

Because symbols are *relations* between morphology and context, it is not an option to suppose that there are, or even could be, morphologically petrified languages. Here is a stark illustration of the principle in question. A ratio is, by definition, a relation between two integers. So the ratio 1:2 is identical with the ratio 3:6, since the relation between one and two is the same as the relation between three and six. Suppose that somebody decided to speak of "numerator-frozen" ratios. Given two integers m and n, m:n would be a numerator-frozen ratio if there were no integers m* and n*, distinct from m and n, such that m:n was identical with m*:n*. The concept of a numerator-frozen ratio is obviously an absurd one. Given that symbols are *relations* between morphology and context, one's thinking involves a similar absurdity so far as it presupposes the existence, or even the possibility, of morphologically frozen languages.

We've seen some reason to think that the concept of a symbol-token is psychologically pregnant. Much of the debate between advocates and opponents of CTM concerns whether this is actually true and, if so, whether it has any bearing on the thesis

that thinking consists in symbolic-manipulations. But the criticism that we just put forth is independent of this debate, and doesn't presuppose the truth of any particular resolution of it. Given that a symbol-token is an instance of a *relation* between morphology and context, it immediately follows that there is no purely morphological characterization of *any* symbol-type. It immediately follows that there is no strictly morphological characterization of *any* of the symbol-types constituting any language, whether actual or possible. Of course, we've seen reason to believe that the contexts in question have a psychological (and also a sociological) dimension, and advocates of CTM are likely to deny this. But the argument just presented has nothing to do with the specific nature of the contexts in question, and it goes through regardless of whether we were right to see such contexts as being psychologically pregnant. Once it is granted that a symbol-token is an instance of a relation between morphology and context, then *regardless of whether that context is understood in psychological terms* it immediately follows that there is no one morphology that two tokens of *any* symbol-type of that language must have in common. More exactly, given any symbol-type S of any language L, if two physical events are morphologically similar, it cannot be *solely* in virtue of their both being instances of S. So supposing that DOG is a symbol-type of Mentalese, if two brain-events are morphologically similar, that is not *solely* in virtue of their both being instances of that symbol-type; and given only that two brain-events are both tokens of that symbol-type, there is no limit to how much they can differ in respect of morphology.

Thus, symbolic manipulations never track semantics *in virtue of being* morphology-driven. This must be understood aright. Obviously morphology-driven operations sometimes track semantics. A monkey can manipulate letter-blocks in such a way that they form a meaningful sentence-token. But given any language, and given any series of inscriptions (or noises...), the morphologies of those inscriptions leave it entirely open what, if any, expressions of that language have been tokened. Because symbols are *relations* between morphology and context, we cannot even coherently conceive of a language of which that isn't true.

It could well be that, circumstances being what they are, there is *in fact* only one way to token the symbol-types of a particular language. We described such a scenario earlier. But it can never be inherent in any language that its symbols are thus frozen.

Conceivably, an advocate of CTM might respond by saying that, even though Mentalese expression-types can *in principle* be tokened by different morphologies, the structure of the brain prohibits different morphologies from instantiating a given symbol-type – much as, in our story about the letter-depository, contingent circumstances prohibited different morphologies from instantiating any given English word.

But, in light of what we've said, it is clear why this move won't work. Because it is never strictly in virtue of its morphology that a physical object is a token of some expression-type, it is no longer an option to say that it is ever strictly in virtue of the fact that shapes are being manipulated that a symbolic operation, or therefore a computation, is being performed. Suppose that, because he just ate a rotten fish, Harry becomes

violently ill. In that case, the cause of Harry's becoming ill coincides with his eating a certain fish. But it is in virtue of his eating something rotten, and not in virtue of his eating a fish, that Harry becomes ill. To put it another way, if X is the fish in question, it is not X's being a fish eaten by Harry, but rather X's being a rotten thing eaten by Harry, that leads to Harry's illness. (We must remember that it is states of affairs, not objects, that have causal powers.) Similarly, even though interactions among bits of morphology may coincide with interactions among symbol-tokens, nothing that is symbolic occurs in virtue of anything that is strictly morphology-driven: because it is never solely in virtue of its morphology that anything is a symbol-token, no strictly morphology-driven interaction of symbol-tokens is itself symbolic. Thus, given two symbol-tokens A and B, unless it is (at least in part) A's having certain *non*-morphological properties that brings about B, the juxtaposition of those two symbol-tokens does not itself constitute anything symbolic. So, supposing that our thoughts are mediated by symbol-tokens, it must be the *non*-morphological properties of such tokens that determine in what sequences they occur. Otherwise those sequences, though *consisting* of symbol-tokens, will not themselves be symbolic. And in that case, those sequences will consist of thoughts that do not add up to larger thoughts, the result being a complete absence of thought-*processes* of any kind, whether computational or not.

Why CTM strips the mental of any causally significant properties

A consequence of our analysis is that, in direct opposition to what Fodor believes, CTM is inconsistent with the fact that mental states have causal powers. In fact, CTM makes it harder than it would otherwise be to accommodate the obvious fact that it is often in virtue of their representational properties that mental states have causal efficacy. I would now like to say, as explicitly as possible, why this is so. (Doing this will involve repeating some points made earlier.)

Here is Fodor's view:

> To think is to compute. To compute is to manipulate symbol-tokens. Symbol-tokens are shapes. Shape is causally efficacious. (In virtue of being round, as opposed to square, an object has causal properties that it wouldn't otherwise have. That is why square wheels would be so useless.) So if we identify thinking with computing, we have no trouble accommodating the fact that our thoughts have causal powers – that thoughts generate other thoughts and also generate actions.

An expression-type is not a shape, and an expression-token is not an instance of a shape. An expression-type is a relation between morphology and context, and an expression-token is an instance of such a relation. So to identify thinking with computing is to identify thinking with the manipulation of instances of relations between morphology and contexts. (More exactly, it is to identify thinking with the manipula-

tion of instances of relations of instances of morphology to instances of certain kinds of context.) It is *not* to identify thinking with the manipulation of bits of morphology.

Suppose that S is a token of the Mentalese word SOCRATES. If our analysis is correct, S is an instance of a relation between morphology and context. Given CTM, the morphology in question must be had by some brain-state B. But what is the relevant context?

There are different possible answers to the question, and none of them validates CTM. Let us start with the answer that Fodor himself gives. According to Fodor, so far as B represents Socrates, it is because some state of affairs of which Socrates was a part is causally responsible for B's existence. So if Fodor is right, the relevant context is some long-gone state of affairs of which Socrates was a part. In that case, any manipulation of S would involve a manipulation of some vast stretch of human history. But it obviously makes no sense to say that vast stretches of history are manipulated. In any case, thinking obviously doesn't consist of such manipulations. It should also be pointed out that in holding that thinking consists in manipulations of symbol-tokens, CTM is committed to the most aggressive form of content-externalism imaginable. This is because the symbol-tokens in question would involve remote stretches of space-time, and CTM therefore becomes no more capable than any other form of content-externalism of explaining how our thoughts can have any causal properties at all.

Fodor's view is that B – the intra-cranial bit of neuro-morphology – is the relevant symbol-token and that, consequently, there is no trouble explaining how it is that B can be manipulated. But, given what the symbols are, CTM cannot coherently say that B itself is the relevant symbol-token.

We've seen that Fodor's answer to the question "what is the relevant context?" doesn't enable CTM to accommodate the fact that thoughts are not causally inert (i.e. that, in virtue of being a thought, a brain-event has causal properties that it wouldn't otherwise have) and, further, that in virtue of being a thought with a *specific* content C, a brain-event has causal properties that it wouldn't have if it instead had some other content. But is there some other possible answer that would accommodate this fact?

Obviously such a symbol-token would have to be entirely intra-cranial. Otherwise, as we just saw, we are forced to say, absurdly, that thinking consists in manipulating vast stretches of extra-cranial space-time. It immediately follows that any version of CTM that can accommodate the fact that thoughts have any causal powers is completely incompatible with content-externalism. This is a problem for Fodor, since one of his principle reasons for advocating CTM is that he accepts content-externalism and believes that an acceptance of CTM is needed to reconcile content-externalism with the fact that thoughts are causally efficacious. But does it follow that CTM itself is in jeopardy?

Yes. As we've just seen, one must accept content-*internalism* if those symbols (or symbol-tokens) are intra-cranial. One must take the view that, if a brain realizes mental states with certain contents, that is to no degree a function of the causal origins of that brain's current condition, and is thus entirely a function of those properties of that brain that can be understood in abstraction of those origins. But if one takes that view, then there is no need to deny that it is in virtue of their *representational* properties that

brain-states have causal powers, and (if like Fodor, you are an advocate of SCT) there is thus no need to deny that it is in virtue of their *semantic* properties that brain-states have causal powers. We can continue to say what we have a strong pre-theoretical wish to say, namely: beliefs and desires have causal powers (or, more precisely, that if x is a thought with a certain content, then in virtue of that fact x has causal powers that it wouldn't otherwise have).

In identifying cognitive operations with non-semantic operations, CTM demands an abandonment of the our presumption that my desire to drink water is at least part of what causes me to walk over to the water fountain and drink from it. In fact, an abandonment of that presumption is really the essence of CTM, since that doctrine is given by the thesis that semantics does nothing and morphology does everything. We've just seen that if symbol-tokens are intra-cranial, there is no need to deny this presumption, and thus no need to accept CTM. We've also seen that CTM fails if Mentalese symbols are, even in part, extra-cranial. So there isn't any delineation of the term "symbol" that validates CTM.

Let us consider a point that might be made on behalf of CTM:

> Expressions of English are tokened by bits of noise or ink, whereas expressions of Mentalese are tokened by patterns of neural stimulation. Tokens of English expressions are typically audible (or visible), whereas tokens of Mentalese expressions are not. One has to *learn* English, but one doesn't have to learn Mentalese. These are but a few of the many differences that we may expect to find between Mentalese and any public language. No advocate of CTM has ever denied that such differences abound, and none of these differences warrants the rejection of CTM. Consider the fact that, if they exist, tokens of Mentalese expressions aren't made of ink or noise. It would be absurd to conclude on this basis alone that thinking does not consist in operations on symbol-tokens belonging to an internal language. It seems to me that all of your criticisms of CTM are comparably absurd. All you've done is to uncover more evidence in favor of the trivial point that Mentalese differs from English (and other public languages) in many ways. Leaving aside your view that CTM is wrong, you haven't said anything that any advocate of that doctrine would necessarily disagree with. At most, you've helped CTM hone its conception of the nature of the internal symbols that mediate thought, by eliminating wrong views about those symbols.

Consider the English word "dog." Here is what we have said about it. If a noise tokens that word, it is not *solely* in virtue of its morphology. Rather, it is in virtue of its morphology *plus* facts about the context of utterance. More exactly, it is in virtue of the relationship between the morphology of that noise and the context in which it occurred. There are instances of expression-types when, and only when, there are instances of relations of the sort just described. This suggests (as we previously observed) that, where public languages are concerned, expression-types just *are* such relations and expression-tokens are instances of such relations.

Supposing that this analysis is correct, anything that *isn't* an instance of such a relation doesn't have any significant similarity to the things that are ordinarily referred to as "expressions." We've discussed why CTM cannot coherently regard tokens of Mentalese expressions as instances of relations of this kind. CTM thus cannot coherently regard cognition as an operation on anything that can appropriately be referred to as an "expression" or "symbol."

We must make a distinction. It is one thing to say that there are no interpretations of terms like "symbol" and "linguistic expression" that validate CTM, and it is quite another to say that we don't think in symbols or linguistic expressions. What we've seen so far is not that we don't think in symbols. What we've seen is that there are no interpretations of expressions like "symbol" and "sentence-token" that validate CTM's thesis that thinking consists in *formal* symbolic operations. We have *not* yet seen that thinking doesn't consist in the tokening of expressions of *some* kind or other.

CTM and SCT are distinct doctrines. CTM presupposes the truth of SCT, but SCT doesn't presuppose the truth of CTM. I will argue that SCT is indeed false. But it is not false *because* CTM is false; it is false for its own distinctive reasons.

Mental causality revisited: Fodor's syntactical approach

For reasons that we discussed in Part I, content-externalism seems to be incompatible with the causal potency of the mental. We discussed a number of attempts to refute this, but found them to be fallacious. But there is one important approach to this problem that we did not discuss – that of Jerry Fodor.

Fodor's analysis makes heavy use of the concept of syntax. We were not in a position to evaluate Fodor's analysis in Part I, since we hadn't yet analyzed that concept. Now that we have done so, we can assess the merits of Fodor's view.

As we've discussed, Fodor is a hard-line content-externalist. In his view, given only that, leaving aside facts about the causal origins of their conditions, Max and Twin-Max are qualitatively identical, it doesn't follow that they have anything in common in terms of the representational contents of their mental states. At the same time, Fodor rightly takes it for granted that *psychologically* they are indistinguishable, i.e. that what an omniscient psychoanalyst would (in his capacity as psychoanalyst) have to say about the one would be identical with what he had to say about the other. Fodor's position is thus consistent with the causal efficacy of the mental and, therefore, with the presumption that psychology has integrity as a discipline.

Fodor reconciles these two views by saying that, whatever causal properties a thought has, it has them in virtue of its *syntax*, not its semantics (representational content). According to Fodor, given that Max and Twin-Max are atom-for-atom duplicates of each other, the syntactic structures of Max's thoughts must be identical with those of Twin-Max's thoughts. What those thoughts represent is a function of spatiotemporally remote, and therefore causally inert, facts about the environments in which those

individuals are embedded. So the *semantics* of a person's brain-states are to be understood in terms of what is remote and therefore causally inert. But *syntax* is an entirely internal affair. Given two subjects that are atom-for-atom duplicates, their thoughts cannot differ in respect of syntax, it being irrelevant what environmental facts led to those thoughts.

Is this position tenable? No. First of all, thoughts have syntactic structure only if they are sentences. Obviously this is not an innocuous view (later we will find it to be false). But for argument's sake, let us suppose that it is correct and that mental states do indeed have syntactic structure. Given this, let P be the meaning of

(a) "Mary loves Tom",

and let P* be the meaning of

(b) "Larry punched Bob."

Further, let

(A) **Mary loves Tom,**

and

(B) **Larry punched Bob**

be tokens of the Mentalese translations of (a) and (b).

There is no denying that (A) and (B), being brain-states, have causal properties. They have mass, shape, temperature, and various other causally efficacious properties. The question is not whether brain-states have causal powers or whether Fodor's view strips them of such powers. The question is whether, supposing that Fodor is right to strip semantics of causal power, he can coherently say that *syntax* has causal power. In other words, can syntax be causally efficacious if semantics is not?

No. As we saw earlier, a sentence-token's syntax lies in *how* it means what it means. Syntax, as we saw, is meaning-how (whereas semantics is meaning-what), and meaning-how no more supervenes on morphology than does meaning-what. Given only its morphology, a token of (a) doesn't have to mean *Mary loves Tom*, and the same is obviously true of (A).

An advocate of CTM could respond by saying that, where Mentalese symbol-tokens are concerned, syntax *does* supervene on morphology. But given what we said in the last chapter, this would mean that Mentalese symbol-tokens (so-called) were so different from their English and Spanish counterparts that we couldn't refer to the former as "symbols" without rendering that term ambiguous.

The Regress-argument revived

We saw in Chapter 13 that syntax is semantic decomposition. A sentence's syntactic structure lies not in what it means, but in *how* it is assigned that meaning by the semantic rules of the language to which it belongs.

Supposing that our thoughts have syntactic structure, the relevant semantic rules are either mentally represented or they are not. If they are mentally represented, those representations cannot themselves coincide with Mentalese sentences (or sentence-tokens), since (for well known reasons[5]) no language can "contain its own truth-predicate", as Tarski put it. (In other words, no consistent language can express all of its own semantic rules.) Those rules must therefore be represented either in some non-linguistic form or in Meta-Mentalese (a language distinct from Mentalese that expresses the rules that assign meaning to Mentalese expressions). If we say that those rules are represented in a non-linguistic form, then we are giving up on the thesis that we think in sentences. If we say that they are represented in sentences (or sentence-tokens) of Meta-Mentalese, then we embark in a vicious regress, since everything that we said about Mentalese is true of Meta-Mentalese.

So let us suppose that the semantic rules of Mentalese are not mentally or neurally or otherwise internally represented.[6] In that case, the operativeness of those rules does not supervene on any causally effective fact about one's person. But then the syntactic structures of Mentalese sentence-tokens are stripped of any causal powers. It straightforwardly follows that Fodor is wrong to hold that it is ever in virtue of their syntactic properties that brain-states do any causal work.

In light of these points, we can see why one of the classic arguments against SCT is in fact cogent. If we think in a language, then presumably we understand that language – we know how to interpret its sentences. But interpreting such a sentence either involves translating it into another language that is already understood or it involves generating a *non*-linguistic representation of its meaning. In the first case, there is a vicious regress. In the second case, there is an acknowledgement that thinking ultimately does *not* consist in the tokening of sentences. Either way, SCT fails.

This argument is well-known to advocates of SCT, and here is how they deal with it:

> (NSK) To know a *public* language, like English, you must mentally represent its rules. It would indeed be viciously regressive to suppose that one must mentally represent the rules of Mentalese. But Mentalese is not a public language; and to circumvent the argument that you just put forth, advocates of the view that we think in Mentalese symbols need only say that you *don't* have to know the semantic rules of Mentalese to know Mentalese. All that is necessary is that, as Lycan

5. See Kuczynski (2002, 2005) for a detailed discussion of these reasons, and Russell (1950: 371) for a pithy and lucid discussion of them.

6. This is the position taken by Fodor (1975) and Lycan (1984: 237) in response to the regress-argument just given.

> (1984: 247) puts it, you be "built to conform" to those rules. You use English-expressions in the right way because you *know* how to use them. But even though you use Mentalese expressions in the right way, it is not because you *know*, or otherwise mentally represent, the corresponding semantic rules. It is because your nervous-system is so structured that you are physically compelled to token those expressions in the appropriate ways. The compulsion here doesn't involve the mediation of semantic, and therefore regress-generating, activity.[7]

To quote Fodor himself:

> [Y]ou don't have to represent the rules of a language that you are able to use: all that's required is that you are so constructed as to conform to the rules. (Cf. the way computers are 'built to use' the machine language.) Of course, it's plausible that you DO internally represent the rules of (say) English; but that's not because English is a language; it's because it's a language that you LEARN; and it's quite plausible that to learn a language is to come to internally represent its syntax and semantics (presumably in 'Mentalese').[8]

If (NSK) is right, Mentalese is so different from any paradigm-case of a language that we must once again question whether it is appropriately described as a language. The answer to that question is "no", as we will now see.

Given any one English-speaker, the English-language could exist if that person didn't exist or did exist but didn't speak English. But suppose that *nobody* spoke English. In that case, the English language wouldn't exist. One might respond by saying that languages are purely function-theoretic pairings of physical objects with meanings. In a moment, we will consider this position, finding it to be false. But even if it is true, it remains an uncontroversial fact that, if nobody spoke English, English would be defunct and thus as good as non-existent.

Knowing English consists in knowing the relevant semantic rules. English is dead if nobody knows those rules. It is obvious that, in general, whether a public language is alive or dead is constitutively dependent on whether people know the relevant semantic rules. (This doesn't mean that any one person has to know all of those rules, but somebody must know some of them.) It is an essential fact about any public language that its very existence (or, in any case, its not being dead) is constitutively dependent on its rules' being known to somebody.

But for a language to be alive, it is not sufficient (though it is necessary) that somebody *know* its semantic rules. A knowledge of those rules must be causally responsible for the judgments that people make as to what noises and ink-marks mean, as the following story shows. You privately learn how to read and write some long-dead language L, and a number of other people simultaneously do the same. But nobody knows that anyone else knows L, and nobody uses their knowledge of L to communicate any-

7. See Lycan (1984: 247), Kuczynski (2002, 2004d).

8. Fodor, private correspondence. The capitalizations are Fodor's own.

thing to anyone else, or even to themselves. Under these circumstances, L is obviously a dead language (though it could be readily revived). So supposing that $E_1 \ldots E_n$ is a complete list of the expression-tokens that people produce, a language L is alive only if, for some i, a knowledge of L it is causally responsible for somebody's producing E_i. (Obviously this is not a *sufficient*, but only a necessary, condition for L's being alive.)

If (NSK) is right, then nobody knows the semantic rules for Mentalese. *A fortiori* a knowledge of those rules is inert. (Even if Mentalese does exist, and a few cognitive scientists manage to figure out what those rules are, that knowledge obviously isn't to any degree what sustains the continued existence of Mentalese. The fact that a few people know Sanskrit isn't enough to make Sanskrit be a living language.) Mentalese is therefore crucially different from anything that is uncontroversially a language.

A related point is that, as a matter of analytic necessity, where a knowledge of semantic rules is inert, there *are* no semantic rules, at least no living ones. If (NSK) is right, the semantic rules of Mentalese exist, and remain active, even though nobody knows any of them. For this reason, the so-called semantic rules associated with Mentalese are so different from those associated with English that it is misleading to use the term "semantic rule", without qualification, in connection with both representational systems.

How are the (so-called) semantic rules of Mentalese relevantly similar to those of English? Like their English counterparts, the (so-called) semantic rules of Mentalese assign representational content to bits of matter (or events). But if anything that assigns content to a bit of matter is *ipso facto* a *semantic* rule, it becomes trivial to say that our thoughts are linguistic, and the thesis that we think in symbols is reduced to the platitude that we have thoughts that have content or, more simply, that we have thoughts.

But even this overstates the similarity between Mentalese semantic rules and their English counterparts. English semantic rules have normative content. They are to the effect that, if one wishes to affirm such and such, one *should* utter thus and such (or, at least, that one can do so). It is a semantic rule that "pickles like to listen to harpsichord-musirc" means pickles like to listen to harpsichord-music. But that rule is not to the effect that anyone *in fact* uses that sentence to convey that proposition. That rule is to the effect, if one wishes to convey that proposition, one *ought* to use that sentence (or, at least, that one can do so). But given the Fodorean assumption that Mentalese semantic rules are unknown, those rules are without normative content: they are not to the effect that, if one wants to convey such and such, one ought (or can) do so by doing thus and such. Those rules are to the effect that such and such bits of matter *in fact* have thus and such content: unlike English semantic rules, they are prescriptively and permissively empty. This confirms our point that there are no meaningful similarities between the things that we *know* to be semantic rules, on the one hand, and the semantic rules of Mentalese, on the other.

(NSK) thus amounts to nothing less than a *rejection* of the SCT. Supposing that (NSK) is right, the only significant property that Mentalese expressions have in common with their English (or Korean or French...) counterparts is that, in both cases, the expressions (so-called) in question have content. Any other putative similarity ends up being skewered on an extension of the line of thought just put forth.

For example, suppose that one were to suggest, as Fodor does, that Mentalese expressions, like their English counterparts, have syntactic structure. So far as English expressions have syntactic structure,that it is because somebody knows at least some of the rules of English semantics. We've already seen that a knowledge of syntax is indistinguishable from a knowledge of the kind of semantic information that is given by contextual definitions. One knows English syntax because one knows (inter alia) how "loves" combines with other expressions into meaningful units – because, in other words, one knows the relevant recursive semantic rules. Syntactic knowledge is combinatorial semantic knowledge. We've seen that there would be no English semantic rules if nobody knew any such rules. The same is therefore true of English syntactic rules, given that syntactic rules are (combinatorial) semantic rules. These points *mutatis mutandis* hold with respect to the syntactic rules of anything that is clearly a language. So if we say that there could be Mentalese syntactic rules *even if nobody knew any of them*, we are saying that Mentalese is so different from anything that clearly deserves to be described as a language that it does not itself deserve to be so described.

There is another problem with Fodor's syntactical approach. Let us look at our paradigms again (as before, suppose that (A) and (B) are the Mentalese translations of (a) and (b)):

(a) "Mary loves Tom",

(b) "Larry punched Bob."

(A) **Mary loves Tom,**

(B) **Larry punched Bob**

Since (a) and (b) are syntactically identical, the same is presumably true of (A) and (B). Supposing that Fodor is right to say that there are mental-causal differences only where there are syntactic differences, it follows that *ceteris paribus* one's belief that Mary loves Tom has the same causal properties as one's belief that Larry punched Bob. Since this consequence is false, so is Fodor's analysis.

Because he identifies syntax with morphology, and not with semantic decomposition, Fodor would not be moved by these arguments.[9] But syntax is not morphology, as we have seen, and the argument just given therefore stands.

One might object that, for special reasons, the Mentalese translations of (a) and (b) *don't* have the same syntactic structures. But what we just said about (A) and (B) will hold of any two Mentalese sentence-tokens that are semantically different but syntactically identical.

In response, an advocate of SCT might say that, where Mentalese is concerned, it is impossible for two sentences to have the same syntax but different semantics. (This is not far from the position that Fodor himself sometimes appears to take, since he sometimes says that there are syntactic differences whenever there are morphological

9. See, for example, Fodor (1987: 18-20). See also Cain (2002: 135).

differences.) But this suggestion is a non-starter. Syntax is recursive semantics. If Mentalese satisfied the condition just described, the same recursions could never be used twice. Since anything comparable to English or Korean – anything that clearly deserves to be described as a "language" – just *is* a set of recursions defined over a certain lexicon, that is tantamount to saying that there is no such language as Mentalese.

I would like to end this chapter with a general point about symbolism. We saw earlier that symbol-types are relations between morphological and psychosocial factors, and that symbol-tokens are instances of such relations. (By a "psychosocial factor" I mean a case of somebody's having a belief about, or at least some kind of propositional attitude towards, some social institution, e.g. a convention of some kind.) But not just *any* such relation qualifies as a symbol, for there are obviously non-symbolic relations holding between morphological and psychosocial factors. Right now, while sitting at my desk, I am contemplating the fact that there is little crime in Sweden. Surely this confluence of circumstances isn't an instance of some symbol's being tokened, even though it is an instance of a relation between morphology and psychosociology. So in saying that symbols are relations of the kind just described we are giving, at most, a partial analysis of the concept of symbolism. What is the right *complete* analysis of that concept? We will outline an answer to this question in Chapter 25. (We will not be able to do so before then.)

CHAPTER 15

Event-causation and the root-problem with CTM

There is an objection to our criticisms of CTM that is more important than any thus far considered:

(FRM) Suppose that Smith (a human being) performs some computation. It is obvious that Smith's behavior is not strictly morphology-driven. It is obvious that, in general, paradigm-cases of computation differ in significant ways from the so-called computational processes that, according to CTM, realize our cognitive activity. As far as I can tell, you've only confirmed this triviality, and you have therefore failed to make a case that CTM is false.

Let's suppose that there is some operation –call it "computation*" – to be defined thus. A computation* is a purely morphology-driven operation on bits of matter. It is left open what, if any, semantic or otherwise representational properties those bits of matter have. If we define CTM as the doctrine that thinking consists in computations*, all of your criticisms of CTM crumble. Those criticisms all presuppose that the "computations" posited by CTM are like the computations performed by engineers and mathematicians. And that presupposition is false, given our new definition of CTM.

There are two points to make here. The advocate of (FRM) might be right to say that my criticisms of CTM are null and void *if* one takes the computations posited by CTM to be nothing more than bits of morphology colliding with other bits of morphology. But given the points we've made in connection with terms "symbol", "algorithm", "syntax", "form", and the like, it isn't clear what intuitive motivation there would be for (FRM). The idea behind CTM is that, in at least some cases, a *cognitive achievement* results from strictly morphology-driven interactions among bits of matter. The question is: why believe that semantically antiseptic interactions among dead bits of morphology can constitute cognitive achievements? As it is traditionally defined, CTM answers by saying that such interactions are relevantly similar to the cognitively pregnant symbolic-operations performed by logicians and mathematicians. But (FRM) rejects this answer and doesn't provide an alternative to it.

It could well be that, when somebody has a thought, that *is* in virtue of the fact that some neural state of affairs has certain morphological properties. In fact, there is probably an element of truth in that view, given that our cognitive activity is almost certainly not entirely independent of the morphological properties of our brain-states. So

(FRM) may be right. But (FRM) doesn't have any meaningful similarity to (CTM), as it is traditionally understood, since (FRM) isn't similar to the view that thinking is computing (at least not if, by "computing", is meant anything even remotely similar to what is conventionally meant by that word).

There is another difference between (CTM), as traditionally understood, and (FRM). Supposing for the sake of argument that there is a one-one correspondence between neural morphology and mental content, (CTM) aspires to *explain* that correspondence. (CTM) aspires to *explain* how thought arises out of unthinking matter. If we grant (CTM)'s supposition that strictly morphology-driven operations among bits of matter are the essence of cognitive achievements like proof- and theorem-generation, it does indeed become clear how it is that bits of dead matter can realize thought, and that is obviously at least part of the reason that (CTM) is so widely accepted. But (FRM) explicitly refuses to endorse that very supposition, and it thus strips (CTM) of its *raison d'être*. So even if it turns out that (FRM) is correct, that hurts (CTM) more than it helps it.

The root-problem with CTM

But there is a much more important point to make in response to (FRM). It can be shown that, so far as interactions among brain-states (or physical objects any kind), constitute instances of *thought*, those interactions must be driven by the representational (semantic) properties of those states. This doesn't mean that strictly morphology-driven interactions don't generate thought. But it means that, so far as they do, it is because representational properties are identical with, or realized by, morphological properties and, consequently, morphology-driven interactions are *ipso facto* semantics-driven.

Let us start by recalling our point that it is not objects, but states of affairs, that have causal properties and that figure into adequate causal explanations. It is not the rock that breaks the window. What does so is the rock's colliding with the window with a certain amount of force at a certain instant. Language sometimes obscures this fundamental fact about causality. We say "Hitler was the cause of World War II." But this is obviously elliptical for some statement to the effect that such and such actions on Hitler's caused (or were partial causes of) World War II. Also, it is states of affairs, not objects, that *are* caused. What is caused is not *the statue*, but is rather the existence in a certain place and time of a piece of marble with certain properties. (In Book 2, Section II, of the *Physics*, Aristotle writes that "Polyclitus [a famous sculptor] is the cause of the statue", and he also writes that "one can speak of the cause of a piece of bronze." But it is such and such acts on Polyclitus' part, not Polyclitus *per se*, that bring about the existence of the statue; and it is not the bronze *per se*, but rather the event of its coming into existence, that is caused. Many of Aristotle's views on causality, and indeed much post-Aristotelian thought on that topic, involve a failure to see that causality is a relation between states of affairs, not objects.)

I should point out that, in this context, the term "state of affairs" is meant to cover both events as well as static conditions. Some artist is causally responsible for the fact that there is a certain statue in a certain place and time. Here what was caused was a (relatively) static condition. Of course, that condition was caused *by way of* events: the artist had to chip away at a hunk of marble and then arrange for it to be moved to a certain place. So static conditions are created by way of changes. In any case, both will be referred to as "states of affairs." (Incidentally, our usage of that expression is not stipulative or neologistic, and is perfectly consistent with its existing meaning. As we will see later, what we refer to as "static" states of affairs – e.g. a statue's remaining in a certain place – are uniform changes, and what we refer to as "events" – e.g. the statue's being moved from one place to another – are changes in the manner of change. So both events and so-called static conditions are both changes, and our usage of the term "state of affairs" in no way renders it ambiguous and thus doesn't involve any equivocation.)[1]

A given object has many different causally effective properties. (Here the term "causally effective" refers to any kind of causal potency, and is not being used in the narrow technical sense that it had in Chapter 12.) Let R be a given rock. R has a certain mass, a certain weight, a certain shape, a certain structure, a certain temperature, a certain color, and so on. In virtue of having any given one of these properties, the rock has certain causal powers that it wouldn't otherwise have. Thus, these properties are not causally inert. Of course, the rock does have some causally inert properties. For example, it has the property of a thing x such that x is either a square or is not a square. But the properties listed a moment ago are not in this category, and in this context we will focus exclusively on the rock's causally effective properties.

Not every property of the rock is involved in every causal transaction in which the rock is involved. The fact that the rock has a temperature of 72° as opposed to 82° is irrelevant to the fact that the window broke. This must be understood aright. Any two non-simultaneous states have *a* causal connection with each other. So the rock's having

1. Incidentally, this line of thought suggests a certain approach to an old metaphysical question. It is often asked whether objects – rocks, trees, people, and so forth – have essential properties. Is there such a thing as the essence of the rock or the essence of your friend Smith? The question is ill-formed, given that it is not objects, but states of affairs, that are causally and explanatorily important. When we talk about "essences" and "essential properties", we are presumably referring to things that have some kind of causal or explanatory significance. As we've seen, *Smith* is a non-entity, as far as explanation and causation are concerned. What is *not* a non-entity is Smith's having a million dollars or Smith's resenting you for your good looks. Obviously it is essential to Smith's having those properties that he have certain other properties. One does not *just* have a million dollars. One's having a million dollars supervenes on one's having various other properties – for example, one is in possession of bits of green paper that play a certain role in the society of which one is a part. Since it is essential to Smith's having a million dollars that there exist some other, lower-level state of affairs of the kind just described, it seems to follow that Smith's having a million dollars has an essence. But it doesn't follow that *Smith* has an essence. And to the extent that essences are supposed to be explanatorily significant entities, it is absurd to say that he, as opposed to his having this or that property, could have an essence.

a temperature of exactly 72° has *a* causal connection with the shattering of the window. But it doesn't have any special or relevant connection: the window would have shattered even if the rock had had a temperature of 82°. So even though it is causally effective in general, the rock's having that temperature is inert in this particular context.

Of course, everything that we just said is true of brain-states. Let B be a particular state (or structure or pattern of neural stimulation...) characteristic of Smith's brain, and let us suppose that B realizes Smith's belief that two is a prime number. B has a number of causally effective properties. It has thermal, electrical, and structural properties, and each of these is causally potent. In virtue of having any given one of these properties, B has causal powers that it wouldn't otherwise have. My own view is that B's *representational* properties belong on this list: in virtue of being a belief that 1+1=2, B has causal properties that it wouldn't otherwise have. But that is just what CTM denies, and it is just what I am trying to establish; so it cannot be assumed in this context.

Not every case where B is involved in a certain causal transaction is a case where B's having a specific temperature is causally potent. In some contexts, B's having that property is potent, and in others it is not. For exactly similar reasons, supposing that B's being a belief that two is prime is causally potent, not every case where B is involved in a causal transaction is a case where B's having that particular property has any special causal role.

Some fiction will help us move onto the next phase of our argument. At time t Smith suddenly thinks *Socrates was wise*. B is the brain-event realizing this mental event. Later, at time t*, Smith suddenly thinks *somebody was wise*. B* is the brain-event realizing *this* mental event. Further, B is what *causes* B* to come into existence. Of course, given what we just said about causality, what this really means is: some state of affairs involving B is causally responsible (at least in part) for some state of affairs involving B*.

Given only what we've said, does it follow that Smith's thinking *Socrates was wise* had anything to do with his later thinking *somebody was wise*? No. Given only that B is what caused B* – more exactly, given only that some state of affairs involving B is what led to there being some state of affairs involving B* – it doesn't follow that Smith's thinking *Socrates was wise* had any special connection with his subsequently thinking *somebody was wise*. That would follow only if *B's being a belief that Socrates was wise* is what led to the state of affairs consisting in *B*'s being a belief that somebody was wise*. If it was B's having a certain temperature that was responsible, and not its being a belief that Socrates was wise, then Smith's thinking *Socrates was wise* at t didn't have anything to do with Smith's thinking *somebody was wise* at t*. We thus cannot conclude that Smith *inferred* that somebody was wise on the basis of his belief that Socrates was wise. So we cannot conclude that Smith thought: *Socrates was wise, therefore somebody was wise.*

Obviously Smith's thinking *Socrates was wise* constitutes a cognitive achievement, and his thinking *somebody was wise* constitutes another cognitive achievement. But it doesn't follow that his thinking *both* of those thoughts constitutes some third cognitive achievement. For this to follow, it would be necessary (though not sufficient, as we will see) that his thinking *Socrates was wise* was responsible for his later thinking *somebody was wise*. If two thoughts are causally unrelated to each other, then surely they don't

jointly constitute some third thought. We've seen that Smith's thinking *Socrates was wise* doesn't necessarily have any (relevant) causal connection to his later thinking *somebody was wise*. So we cannot conclude that his thinking *both* of those constitutes some third cognitive achievement. For example, we cannot conclude that it involves his making an inference. It is irrelevant that B is what caused B*, and it is irrelevant that the proposition encoded in B logically entails the proposition encoded in B*.

We can reinforce this point by extending our story. Later that day, after reading an article about some historical figure (other than Socrates) who committed an unwise act, Smith suddenly thinks *somebody was not wise*. This happens at time t**. It would be absurd to say that, in virtue of thinking *Socrates was wise* at t, and then thinking *somebody was not wise* at t**, Smith was guilty of a logical non-sequitur. In fact, it would be absurd to say that, in virtue of having those two thoughts, Smith deserved credit for having *any* thought-process that embodied *any* view at all as to how, if at all, those two propositions are related. The reason this would be absurd is obviously that his thinking *Socrates was wise* didn't have anything to do with his thinking *somebody was not wise*. (In any case, those two events didn't have any special or relevant causal connection, even though, like all non-simultaneous events, they have *a* causal connection.) For similar reasons, it would be absurd to say that, in virtue of thinking *Socrates was wise* at t and *somebody was wise* at t*, Smith has had a thought that embodies any view at all as to how those propositions are related to each other. It is therefore out of the question to say that Smith deserves credit for having made a logical inference.

The points just made can be generalized in a few different directions. Let us start with the most obvious direction of generalization. Let P be and P* be any two propositions such that P logically entails P*. Further, let b_p and b_{p*} be two brain-states, both had by a single person X, such that b_p realizes a belief that P and such that b_p* realizes a belief that P*. Supposing that b_p causes b_p*, it doesn't follow that b_p's being a belief that P has anything to do with b_p*'s coming into existence. It could be that it was b_p's having some non-semantic property that was thus responsible. It therefore doesn't follow that X's thinking P has anything to do with his later thinking P*. There is no reason to believe that X inferred P* from P or, indeed, that he had any thought-process that embodied any views, whether meritorious or not, as to how those two propositions might be related.

Let us further generalize our analysis. Let T and T* be any two non-simultaneous events or objects that are representational. For the sequence consisting of those events to form a *single* representation, it is not enough that *T* be what causes T* to come into existence: it is necessary that T's *being a representation* be what does so. A bit of fiction may illustrate my meaning. At 3:00 p.m. I am feeling very sad. I attempt to write "I am suffering from great sadness." I manage to write "I am suffering from great." But right after I write the "t" in "great", I am struck by the beauty of my own penmanship, and this causes me to feel great joy. Because of this, I forget about my earlier intention to write "I am suffering from great sadness", and I form the intention to write "sadness

never lasts and, ultimately, life is a great joy." But right after I write the second "s" in "sadness", I am interrupted by the doorbell. I never write in my diary after that.

Under these circumstances, have I produced a token of "I am suffering from great sadness"? No. I have only produced an ink-mark that looks just like a token of that sentence. To be sure, the token that I produced of "I am suffering from great" is causally responsible, at least in part, for the adjacent token of "sadness." Nonetheless, I never managed to carry out my original intention of tokening the sentence "I am suffering from great sadness." Therefore, that sentence was not tokened in this context, even though the ink-mark I have produced is morphologically just like a token of that sentence.

Let T be my inscription of "I am suffering from great", and let T* be my inscription of "sadness." It was T's having certain aesthetic properties, and *not* its having a certain meaning, that led to my writing T*. So *T's being a symbol* (or, in any case, T's being a symbol that has a certain meaning) had nothing to do with T*'s coming into existence. So far as they have semantic properties, the inscriptions "I am suffering from great" and "sadness" are causally unrelated (except in the trivial sense in which any two non-simultaneous entities are causally related), and thus have nothing to do with each other, that being why they don't jointly form a sentence-token.

Supposing that X and Y are symbol-tokens, the juxtaposition of X and Y doesn't constitute a single symbol-token, unless X's being a symbol (or representation) is what generates Y (or *vice versa*). If this condition isn't met, then X has no semantically significant relation to Y, in which case it immediately follows that they don't jointly form any semantic unit. The juxtaposition of these symbols may look (or sound) like a meaningful unit, and the causal mechanism that mediates between X and Y may be completely reliable. But if it isn't X's *being a representation* that generates Y (or *vice versa*), then X and Y aren't parts of a single representation. Where semantics is causally inert, representations cannot combine to form more complex representations, and nothing has any representationally significant relation to anything else.

Some thoughts consist of other thoughts. If I think *grass is green and snow is white*, that thought involves the thought *grass is green* and also the thought *snow is white*. CTM says that semantics is causally inert. Given what we just saw, CTM thus cannot accommodate the fact that there are thoughts that consist of other thoughts. CTM is therefore inconsistent with the fact that there are complex thoughts.

Necessary versus sufficient conditions for the formation of complex thoughts

Suppose that I think *Socrates was wise* and then later think *somebody was wise*. Let b and b* be the brain-states mediating those thoughts. As we just saw, if b and b* are to constitute a single thought, or are to be part of the same thought-process, it is *necessary* that b's being a thought that Socrates was wise be at least part of what generates b*. But it is not *sufficient*, as the following story shows.

A hypnotist programs me to think *something was wise* whenever my heart rate goes above 90 beats per minute. One day it suddenly occurs to me that Socrates was wise. (I finally understand some obscure passage of *The Republic*, and I see that, contrary to what I used to think, Socrates really was wise.) The resulting euphoria causes my heart-rate to sky-rocket. Because of my hypnotic programming, this causes me to think *something was wise*. Here we have a case where my thinking *Socrates was wise* causes me to think *something was wise*. So if b and b* are the brain-states associated with those two thoughts, we have a case where b's being a thought that Socrates was wise is responsible for b*'s being a thought that something was wise. But we don't have a case where I have thought *Socrates was wise, therefore something was wise* and, more generally, we don't have a case where a single thought or thought-process comprises both of those thoughts. An exact analogue of what we just said about b and b* could be constructed in connection with ink-marks or noises or any other symbol-tokens.

In general, for the juxtaposition of two representational entities (be they thoughts or symbol-tokens) to constitute a *single*, complex representation, it is *necessary but not sufficient* that the one's being a representation of a certain kind be causally responsible for the generation of the other.

Of course, it is an interesting question what *other* conditions must be fulfilled. But in this context we don't need to know the answer to that question. CTM demands that semantics be *totally* inert, and it thus renders impossible the fulfillment of one of the conditions necessary for the combining of representations into other representations. This by itself is a major strike against CTM, given how frequently such combinations occur.

Semantic versus conceptual complexity

To develop this line of thought, and to see more fully why CTM must be rejected in light of it, we must first distinguish between two different kinds of complexity. Consider an inscription of the word "Socrates." That inscription is physically complex, since it consists of a multiplicity of marks. But it is not *semantically* complex.[2]

Any brain-state that realizes a concept of Socrates will inevitably have a great deal of physical complexity. But it doesn't follow that any such state has representational complexity. I myself believe that anything that realizes a concept of Socrates *will* have representational complexity. (We argued for this in Chapter 1.) But even if anything that realizes a concept *does* have representational complexity, we must distinguish between the kind of complexity that such a thing has *in virtue of* being such a concept and the kind of complexity that it has for some other reason.

2. Strictly speaking, it isn't quite right to say without qualification of that inscription that it is physically complex. Whether it is complex, and if so how, is a function of the explanatory context. But we needn't pursue this.

With this distinction in place, let us tell another story. At time t, Smith thinks *Socrates*. b is the brain state realizing this thought. At some later time t*, Smith thinks *snored*. b* is the brain-state realizing *this* thought. For exact analogues of the reasons just given, unless b's being a representation of Socrates is responsible for b*'s subsequent occurrence, Smith has *not* thought *Socrates snored*, at least not in virtue of the fact that b and b* occurred. Semantics cannot be inert if anyone is to think anything true or false. No true or false thought is *representationally* simple. So nobody can think anything true or false except in so far as it is one's brain-states' having *representational* properties that are governing one's cognitive activity. It follows that CTM strips anyone of the ability to think anything true or false, given that, according to CTM, it is never in virtue of its semantic properties that a brain-state has causal powers.

I myself believe that, contrary to what the argument just given seems to presuppose, one cannot *just* think *Socrates* or *snored*. Supposing that this is right, it might seem that my argument crumbles. But it does not crumble, and this is because, as we discussed earlier, CTM *needs* it to be possible to think just *Socrates* or *snored* or *blue*. CTM demands that we have atomic concepts, and that our thoughts consist of concatenations of these concepts. So our argument shows that CTM fails relative to its own assumptions.

CTM requires the truth of some kind of conceptual atomism. So conceptual atomism would indeed be *de rigueur* if CTM were correct. But we've seen reason to think that CTM is not correct. It involves misunderstandings of the nature of symbolism, formal truth, and causation; and it doesn't appear that any doctrine significantly similar to CTM would *not* involve these misunderstandings. So there is *no* reason to accept conceptual atomism, given only that CTM demands its acceptance. In the next three chapters, we will see why some other reasons to accept conceptual atomism are in fact spurious.

CHAPTER 16

Fodor's first argument for conceptual atomism

Fodor accepts CTM. He also sees that CTM demands the truth of conceptual atomism (henceforth "atomism"), and this is at least part of the reason that he accepts atomism. But Fodor realizes that few are willing to accept atomism, and he thus provides three independent arguments for it (Fodor 1998). In this chapter, we will examine the first of these three arguments. We will examine the other two in Chapters 17 and 18.

Fodor's argument

> There is no doubt that *some* expressions are definable. "Triangle" can be defined as "closed, straight-edged, planar figure such that any two, but not all three, of its sides intersect." "Two" can be defined as "the successor of one." In general, expressions belonging to "bona fide axiomatic systems" can be defined. Further, "patently phrasal" expressions can be defined, at least in a relative sense: "smart brown cow" can be defined as "anything x such that x is smart, x is brown, and x is a cow."
>
> But leaving aside these two classes of expressions, virtually no expression can be defined. No one has produced an adequate definition of "knowledge." (It used to be thought that "knowledge" could be defined as "justified true belief." But Gettier 1963 famously proved this false, and attempts to produce an adequate definition have failed. See Shope 1983.) What we just said about "knowledge" is true of practically any expression denoting a philosophically, or otherwise theoretically, important concept. In fact – what is probably more important – even expressions denoting pedestrian concepts turn out to be incapable of definition. Consider the verb (not the noun) "paint." Contrary to what might initially be thought, this cannot be defined as "to cover in paint" or even "to intentionally cover in paint." When you dip your paintbrush into a bucket of paint, you are not painting it, even though you are intentionally covering it in paint. There are other *prima facie* reasonable definitions of this verb, but they all fail. Given that the word "paint" is indefinable, it follows that the corresponding concept is simple. After all, if that concept were complex – if it consisted of other concepts (e.g. *cover*, *intentional*, and so on) – then there would be some true statement saying *how* that concept were composed of those other concepts (just as there is a true statement saying *how* water molecules are composed of hydrogen and oxygen molecules). There isn't; so it doesn't. "Paint" is indefinable and the concept that it expresses is therefore simple. Given virtually any other expression, what we just said about "paint" is true of that other expression, showing that *all* concepts are simple. (In any case,

> it shows that all, or nearly all, the concepts denoted by semantically non-complex expressions of natural language, leaving aside those concepts that are expressed by artifacts like "integral", are simple.)[1]

Our response

In this context, the word "concept" obviously doesn't denote a psychological entity. So if cogent, the argument just presented shows that concepts *in the non-psychological sense of the word* are simples. That argument does not, at least not directly, say anything about concepts in the psychological sense. To simplify further discussion, let us refer to concepts in the *non*-psychological sense as "concepts$_o$", and let us refer to concepts in the psychological sense as "concepts." So a concept is one's grasp of something, and a concept$_o$ is a platonic entity. (The subscript stands for "objective.")

Fodor is trying to show that concepts (no subscript) are atoms (simples). He is trying to make a case for conceptual atomism. So Fodor must be assuming that if concepts$_o$ are simples, the same is true of our psychological representations of them. In any case, if this assumption is not granted, then Fodor's argument doesn't show that concepts are atoms. So for the sake of argument, let us grant Fodor's tacit assumption that conceptual atomism follows from conceptual$_o$ atomism. (Later we will find this assumption to be false.)

Granting this assumption, there are two major problems with Fodor's argument. Nobody denies that x's being a belief is *necessary* for x's being knowledge, and nobody denies that the necessity in question is analytic, and not empirical, in nature. Nobody denies that putting paint on something is *necessary* for painting it, and nobody denies that the necessity in question is analytic. Given any concept$_o$ C, nothing is easier than to identify conditions that are analytically *necessary* for something's falling under it. (Actually, Quine and a few others *do* deny that *if x is knowledge, x is a case of belief* is analytic. At the end of the present chapter, we will examine their reasons for holding this. For now we will operate on the reasonable – and, we will see, demonstrably correct – view that many propositions are analytic.)

But surely analytic necessities correspond to conceptual$_o$ structure. *x is an instance of knowledge* does not entail *x is an instance of putting paint on something*. But the latter is entailed by *x is a case of painting*. Surely this tells us that the architecture of the one concept$_o$ is different from the architecture of the other.

In fact, this last point would seem to be a truism: given that to fall under the one concept$_o$ but not the other, x must be an instance of belief, those two concepts$_o$ are *ipso facto* structurally different. When we talk about the "structures" (or "compositions" or "architectures") of concepts$_o$, we obviously aren't referring to their physical structures. We are referring to relations of logical dependency, not of spatiotemporal juxtaposition. With a minor qualification to be stated shortly, for concepts$_o$ C and C* to differ in respect of their structures just *is* for *x is a C* to entail something not entailed by *x is a*

1. Fodor (1998: Chapter 4).

C*. A corollary is that, as long as we can give *necessary* (and non-trivial) conditions for x's falling under C, it follows that C has conceptual$_0$ structure. (A trivial condition would be given by a statement like: x falls under C iff x falls under C.)

Taken in conjunction with this line of thought, Fodor's own argument would seem to show that any given concept$_0$ has an *infinite* amount of structure. Getteriologists may not have provided necessary and sufficient conditions for something's being knowledge. But they have succeeded in providing a number of necessary conditions (Shope 1983). Given what we said a moment ago, each of these new necessary conditions corresponds to a newly discovered fact about the architecture of the concept$_0$ of knowledge. If there are *infinitely* many such conditions, then that concept$_0$ has an infinite amount of structure. Supposing that there are only finitely many such conditions, that concept$_0$ has a finite, but non-null, amount of structure *and*, more importantly, that concept$_0$ *is* definable. Either way, Fodor's argument collapses.[2]

---

2. Let us address a possible objection to our argument:

"Not *all* analytic dependencies correspond to conceptual$_0$ structure. *x is knowledge* entails *x is not a square circle*. But this analytic dependency doesn't show us anything specifically about the concept$_0$ of knowledge. After all, *x is a case of painting* entails *x is not a square circle*, as do *x is a house* and *x is a barber*. The fact that *x is knowledge* entails *x is not a square circle* evidently tells us *nothing* about the structure of the concept$_0$ of knowledge. So it would appear that you were wrong to see analytic dependencies as indicating conceptual$_0$ structure, and that your argument against Fodor was misguided."

This objection actually reinforces our criticism of Fodor's argument. Consider the entailment given by the statement:

(*) "For any x, if x is knowledge, then x is belief."

If the occurrence of "knowledge" is replaced with an occurrence of "a house" or "case of painting", what results is a falsehood. So (*) *does* give us differential information about the structure of the concept$_0$ of knowledge, thus confirming our point. It is an interesting question what exactly the difference is between (*), on the one hand and:

(**) for any x, if x is knowledge, then x is not a square circle,

on the other. More generally, it is an interesting question what precisely the difference is between those analyticities that reveal conceptual$_0$ structure and those that do not. But it is not a question that has to be answered here. For our purposes, it is enough to know that analyticities typically (though not universally) reveal conceptual$_0$ structure.

But for what it's worth, it seems to me that the difference between (*) and (**) is that (**) is *semantically* true, whereas (*) is not. As we saw in Chapter 13, if a sentence is semantically true, that is because of structural properties of its derivation tree, a consequence being that its truth has nothing to do with the specific contents, or therefore the specific concepts$_0$, meant by the expressions constituting that sentence. Consequently, such a sentence doesn't expose what is distinctive about those concepts$_0$ and thus doesn't provide any information about them. By the same token, if a sentence is non-semantically but analytically true, its truth does turn on the peculiarities of the meanings, and therefore the specific concepts$_0$, assigned to its constituent-expressions. Consequently, such a sentence does expose what is distinctive about those concepts, and it therefore does provide information regarding the structures of those concepts$_0$.

Here is a useful analogy. It is a point often made by philosophers of science that the term "explanation" is contextual.[3] To provide a *complete* explanation of anything, I would have to describe every event that led up to that event, starting with the Big Bang. In *this* sense of the word "explanation", it is impossible in practice, and probably in principle, to explain anything. But it is a datum that perfectly good explanations are given by statements like "Smith crashed the car because he was drunk", "Jones is limping because he twisted his ankle while playing squash", and so on. So given only that "complete" explanations cannot be given, it would be absurd to conclude that explanation *tout court* is impossible. It would be absurd to say: "the world is explanatorily amorphous, since nothing can be completely explained."

Similarly, it would be absurd to say: "the concept$_o$ of knowledge is explanatorily amorphous, since there is no *complete* account of its analytic structure." Given that there are partial characterizations of its structure, it follows that it is not analytically amorphous.

An objection to our assessment of Fodor's argument

There is an objection to this line of thought that we must consider.[4] And it is not just *an* objection: it is one that embodies a position that Fodor himself holds and that underlies his various arguments for conceptual atomism:

> According to the argument you just gave, C and C* have different structures exactly if they differ in their entailment-relations. With a few innocuous qualifications, statement:
>
> (S) "x is identical with Socrates"
>
> doesn't entail anything not also entailed by:
>
> (A) "x is identical with Aristotle."
>
> We know from Kripke that the sentence "Socrates was bald" is not analytic. We also know from Kripke that, if we replace the occurrence of "Socrates" in that sentence with any other proper name the result will not be analytic. Thus, "Plato was bald" is not analytic (even though it is true), and neither is "Napoleon was bald." In general, uniformly replacing proper names with other proper names never turns a non-analytic truth into an analytic truth or (for similar reasons) *vice versa*. This suggests that the concept$_o$ expressed by the word "Socrates" – this being the concept$_o$ *identical with Socrates* (or some such) – has no analytic structure not also had by the concept$_o$ expressed by the words "Plato" or "Napoleon." Your

3. See, for example, Carnap (1966).

4. It is similar to an objection that George Rey (1996) makes of Peacocke's (1992, 1996) analysis of conception. We will discuss both Peacocke's analysis, and Rey's objection to it, in Chapter 24.

position is that concepts$_o$ differ just in case their analytic liaisons differ. If you are right, then the concept$_o$ *Aristotle* (or *identical with Aristotle*) is structurally just like the concept$_o$ *Plato*. By itself, this doesn't mean that those concepts$_o$ have *no* structure. (It could be that *x is identical with Aristotle* entails *x is a spatiotemporal object*.) But it does show that, so far as those concepts$_o$ *differ*, that has nothing to do with structural differences between them. That, in its turn, shows that grasping the one concept$_o$ *as opposed to the other* cannot be understood in terms of grasping structural differences between them, and thus cannot be understood in terms of one's grasping some set of concepts$_o$ $C_1 \ldots C_n$ such that the one concept, but not the other, is constructed out of those concepts$_o$. So one's grasping the concept$_o$ *Plato* as opposed to the concept$_o$ *Aristotle* is not to be understood in terms of what other concepts$_o$ one grasps. That difference is conceptually primitive. What we just said about the concepts$_o$ *Plato* and *Aristotle* is true of countless other pairs of concepts$_o$. Indeed, what we just said holds of practically any two concepts$_o$. The pairs of concepts$_o$ that you chose seem to be more the exception than the rule. What you said about the pair of concepts$_o$ *knowledge* and *paint* seems not to apply universally or even usually. So, given only what you've said, Fodor's argument may go through, at least for the great majority of concepts$_o$. Further, it seems clear that *differences* between concepts$_o$ are often primitive, i.e. if C and C* are different concepts$_o$, that fact cannot typically be understood in terms of how those concepts$_o$ are composed out of other concepts$_o$. So it would seem that, in the end, Fodor's argument *does* go through, notwithstanding some of the technical problems that you have pointed out with it.

We have spoken a great deal about the difference between, on the one hand, what is literally meant by expressions and, on the other hand, what we learn in the process of ascertaining what is literally meant by them. The objection just stated involves a failure to appreciate this difference. As we've seen, there is some x such that what is literally meant by any token T of:

(HL) "Hesperus is lovely",

and also by any token T* of

(PL) "Phosphorous is lovely",

is simply:

(XL) x is lovely.

But one must compute the meanings of T and T* on the basis of background semantic and pre-semantic knowledge; and for this reason, those two tokens may diverge dramatically in respect of what they impart to one. For a moment, let us discuss the concepts$_o$ *Hesperus* and *Phosphorous* (or *identical with Hesperus* and *identical with Phosphorous*). Given the points that we will make in connection with these concepts$_o$, it will be clear what we must say about the concepts$_o$ *Socrates* and *Plato*.

First of all, which entity are we talking about when we talk about "*the* concept$_o$ *Hesperus*" or "*the* concept$_o$ *identical with Hesperus*"? If we are talking about the information *through which* one computes the meanings of sentence-tokens like T, then that concept$_o$ is replete with conceptual$_o$ structure. As we saw in Chapter 4, the information that one associates with T (in our technical sense of "associates with") is some existence-claim having the form:

(EHL) There is some luminescent beautiful x in such and such part of the morning sky...and what is meant by "Hesperus is lovely" is: x is lovely.

So on one possible disambiguation, the term "the concept$_o$ *identical with Hesperus*" refers to this wide-scope descriptive information.[5]

What else might the expression "the concept$_o$ of Hesperus" refer to? Hesperus' existence is not a simple, ultimate metaphysical fact. Given any possible world W where Hesperus exists, its existence supervenes on various other facts. Certain mass-energy displacements, having certain origins, resulted in a large solid body having a certain trajectory. Of course, Hesperus is not modally frozen: there are possible worlds where, because of some cataclysm, it breaks free of its orbit around the sun. More generally, there are worlds where it has properties very different from those that it actually has. But there are no worlds where Hesperus has *totally* different origins from those that it actually has.

Given this last point, it would seem that any world W where Hesperus exists is one that has much in common with our world *apart* from its having the property of including Hesperus: some space-time region R of W must comprise innumerable sub-atomic events that are qualitatively much like the events comprised by some corresponding region in our world. Further, because Hesperus is presumably individuated by its origins, it exists in a world only if that world comprises at least some objects (other than Hesperus) that are *numerically* identical with objects in our worlds, the reason being that a

5. Or, more likely, it refers to some kind of idealization of this information. As we discussed, the information through which one computes the meaning of tokens of "Hesperus", and thus of sentence-tokens containing that term, varies from person to person. No two people will learn what is meant by "Hesperus" in quite the same way. Given any two people to whom that term is defined ostensively, their respective optical, and subsequent epistemic, relations to Hesperus will probably not be exactly the same; and the same is true of any two people to whom it is defined indirectly or non-ostensively. So far as there is *one* "concept$_o$ of Hesperus" – i.e. so far as the term "the concept$_o$ of Hesperus" is not an improper, and thus non-referring, definite description – that term seems to refer to some concept$_o$ that abstracts from all of these person-to-person variations, leaving only some core concept$_o$ like *unique last body to disappear from the morning sky*, or some such. (Of course, this is exactly what Frege and Russell, and indeed most pre-Kripke semanticists, would have identified as "the concept$_o$ of Hesperus.")

What is important here is that if by "the concept$_o$ of Hesperus" one means the information through which people think about Hesperus – or, more exactly, through which they compute the semantics of occurrences of expressions like T – then the concept$_o$ of Hesperus is by no means conceptually$_o$ simple, as it consists of the concepts$_o$ *morning*, *luminescent*, and so on.

world that comprised none of the same micro-particles as our world would *ipso facto* fail tom comprise anything that had the same origins as Hesperus. (And, of course, the same thing *mutatis mutandis* holds of those other objects – of those various micro-particles.) Further, if a world comprises those other objects, and those objects are arranged in the right way, then that world *ipso facto* comprises Hesperus. So there is some set of properties $P_1 \ldots P_n$ such that, for a world W to comprise Hesperus, it is necessary and sufficient, that each of these properties be instantiated in W, and such that none of these properties will be *identical with Hesperus* or *identical either with Hesperus or with a square circle* or any other property that must itself be understood *in terms* of the concept$_o$ of Hesperus.

It seems to me that, on one possible disambiguation, the term "the concept$_o$ of Hesperus" refers to this set of properties: the set of properties such that their joint instantiation is necessary (and sufficient) for Hesperus' existence in any possible world W. Relative to this disambiguation, the concept$_o$ of Hesperus is replete with conceptual$_o$ structure. For exactly similar reasons, "the concept$_o$ of Socrates" would refer to a conceptually$_o$ complex entity. And this, of course, would scuttle the objector's point that "the concept$_o$ of Hesperus" and "the concept$_o$ of Socrates" refer to conceptually$_o$ simple entities.

Inevitably, some will respond by saying:

> I have a perfectly good grasp of the concepts$_o$ *Socrates, Hesperus,* and *Aristotle.* But I haven't any knowledge as to the nature or identity of the sub-atomic events in virtue of whose occurrences these concepts$_o$ are instantiated. So it is not really an option to say that the concept$_o$ of being identical with Socrates is to be understood in the way you just proposed.

Supposing that this is right, what does it mean to say that one "grasps the concept$_o$ of being identical with Socrates"? It means that one is able to think about Socrates. Of course, to think about Socrates, it is *not* necessary to know much at all about his origins, let alone about their molecular basis. But as we've seen, to think about Socrates, one *does* have to know some rather complex description that Socrates alone uniquely satisfies. The objector is quite right that, in *this* sense of the term "concept$_o$ of Socrates", one doesn't have to know anything about the molecular basis of Socrates' conditions of origination to grasp a concept$_o$ of Socrates. But this is consistent with our point that anything fit to be denoted by the term "the concept$_o$ of Socrates" has conceptual$_o$ structure. Of course, what we just said about "Socrates" is true of "Aristotle", "Hesperus", and any other singular term. This shows that the thing denoted by "the concept$_o$ of Hesperus", on either disambiguation of that expression, has a structure very different from the thing denoted by "the concept$_o$ of Socrates."

Of course, the objector is entirely right to say that that, in terms of their literal *meanings*, "Socrates" and "Aristotle" are equally devoid of descriptive content. And he is right to say that, in terms of their literal meanings, ⌜Aristotle has phi⌝ will be analytic iff the same is true of ⌜Plato has phi.⌝ (This last statement assumes, of course, that phi is not a predicate like "identical with Aristotle" or "identical with Plato.")

But as we've seen (Chapters 1–4), nothing at all can be inferred from this fact concerning the structure of the concept$_o$ of Aristotle or of the concept$_o$ of Socrates. An exact analogue of the argument just given shows that we just said about singular terms is also true of terms denoting natural kinds ("water", "wood").

Does conceptual atomism follow from conceptual$_o$ atomism?

As we noted earlier, Fodor's argument tacitly assumes that the truth of conceptual$_o$ atomism is sufficient for the truth of conceptual atomism. Is this assumption correct? No. While a case can be made that conceptual$_o$ atomism is *necessary* for conceptual atomism, no case can be made that it is sufficient.

For the sake of argument, let us suppose that (with the obvious exceptions that Fodor discusses) conceptual$_o$ atomism is correct, and let C and C* be two concepts$_o$. Given only that these concepts$_o$ are simple, i.e. that no concepts$_o$ (other than themselves) are constitutive of them, it doesn't follow that my mental representations of them are conceptually simple. Let X and Y be two qualitatively identical, perfectly homogeneous spheres that are located in different places. I think of X as *the sphere in Aunt Jenny's basement* and I think of Y as *the sphere in Uncle Fred's attic*. So even though X and Y are qualitatively identical, and even though they are both lacking in internal structure, it doesn't follow that my *concepts* of them are qualitatively identical or that my concepts of them are without internal articulations.

Given only that two different concepts$_o$ have no structure, and thus have the same structure, it doesn't follow that my concept of the one is qualitatively identical with my concept of the other. What follows is that I cannot distinguish those concepts$_o$ *on the basis* of their internal structure.

Of course, X and Y are spatiotemporal individuals, whereas concepts$_o$ are not; and, conceivably, one might therefore argue that where *non*-spatiotemporal entities are concerned, the structures of our concepts *do* mirror those of their objects. But Fodor provides no argument for such a view and, more importantly, such a view would be extremely implausible. Structurally identical spatiotemporal individuals can be distinguished by their differences in spatiotemporal location. But structurally identical concepts$_o$ could not be distinguished in this way. So far as concepts$_o$ can be distinguished from one another, it must therefore be on the basis of their structures. In any case, Fodor's tacit assumption (viz. that conceptual$_o$ atomism is necessary for conceptual atomism) is false.

Quine on analyticity

This is a good place to discuss Quine's (1951) analysis of the analytic-synthetic distinction. Quine holds that, with a few trivial exceptions, all sentences are synthetic. (The exceptions involve cases where one term has been *stipulated* to have the same meaning as another.) Our argument against Fodor assumes that there are analytic truths. Indeed, we have made liberal use of this assumption in this work, and we must therefore consider Quine's views on analyticity. It should be pointed out that Fodor (1998: 25, 45–46) agrees with Quine. Indeed, Fodor takes it as a foregone conclusion that Quine's arguments are cogent, give or take some nuances.

Here is the basic structure of Quine's argument. Analyticity is to be understood in terms of synonymy. But synonymy must be understood in terms of analyticity. Therefore the concept of analytic truth is a viciously circular and therefore incoherent one. Let us now state Quine's argument in full.

If indeed "bachelors are unmarried males" is analytic, that is because "bachelor" is synonymous with "unmarried male." Similarly, if "triangles are three-sided, closed, planar, straight-edged figures", that is because "triangle" is synonymous with "three-sided, closed, planar, straight-edged figure." In general, analyticity is to be understood in terms of synonymy.

But what is it for two expressions to be synonymous? It is not *merely* for all intersubstitutions of them to preserve truth-value. "Rhenates" and "chordates" can be intersubstituted *salva veritate*, but they are obviously not synonymous. For two expressions to be synonymous, intersubstitutions of them must preserve not only truth-value, but also *meaning*. But the concept of meaning is itself to be understood in terms of the concept of analyticity: 'x has phi' has the same *meaning* as 'x has psi' only if 'x has phi iff has psi' is analytic. So analyticity is to be defined in terms of synonymy, and synonymy is to be defined in terms of analyticity. Attempts to define analyticity are thus doomed to failure, since they are necessarily characterized by the same vicious circularity as attempts to inductively justify induction, and the concept of analyticity is therefore incoherent.

Having presented the preceding line of thought, Quine correctly notes that an objection can be made to it. Strictly speaking, "rhenates" and "chordates" *cannot* always be intersubstituted *salva veritate*. Even if "Smith believes that rhenates are chordates" is false, "Smith believes that rhenates are rhenates" may be true. In general, *intensional* contexts don't always tolerate intersubstitutions of non-synonymous co-referring (or co-extensive) terms. For this reason, contrary to what was alleged a moment ago, "synonymous" *can* be defined; for it can be defined as "intersubstitutable *salva veritate* in intensional contexts." And this breaks the definitional circle.

Quine has no trouble rebutting this point. The concept$_o$ of an "intensional" context is itself to be understood in terms of the concept$_o$ of meaning (and, therefore, in terms of the concept$_o$ of analyticity). C is an intensional context iff replacing any term occurring in it with a co-referring term is guaranteed to preserve the truth-value of the host-sentence only if that intersubstitution involves expressions having the same *meaning*.

As we saw a moment ago, the concept$_o$ of meaning is to be understood in terms of the concept$_o$ of analyticity. So "synonymous" cannot be non-circularly defined as "inter-substitutable *salva veritate* in intensional contexts."

Evaluating Quine's argument

If there are any analytic sentences at all, one of them is:

(*) "n is a unique even prime iff n is one less than three."

Quine explicitly assumes that the expressions flanking the biconditional must be *synonymous* if that sentence is to be analytic. But the expression "n is a unique even prime" obviously isn't *synonymous* with the expression "n is one less than three." If those two expresions were synonymous, i.e. if they had precisely the same linguistic meanings, then (*) would be as trivial as "a yard is a distance of three feet." Analytic truth isn't always definitional truth (that being why analytic truth isn't always trivial). The sentence "if an event begins at 3:00 p.m., then it doesn't begin at 3:02 p.m." is analytic but not definitionally true, since (for well-known reasons) "begins at 3:00 p.m" is not synonymous, or otherwise equivalent, with "doesn't begin at 3:01 p.m. or at 3:02 p.m. or..." or with any other comparable expression. Quine is therefore wrong to assume that there is analyticity only where there is synonymy, and his argument thus implodes.

Conceivably, Quine might deny that (*) is analytic on the grounds that the expressions flanking the biconditional are not synonymous. But in that case, he would simply be redefining the term "analytic", and his argument would have force only against an artifact of his own definitions.

If we are to make this argument more precise, we must note that the term "analytic" is ambiguous. Sometimes when a sentence is described as "analytic", what is meant is that it encodes a conceptually$_o$ true proposition: a proposition that holds entirely in virtue of the structures of the concepts$_o$ composing it. An example of a sentence that is analytic in this sense would be: "there is no law where there is no government." Surely it is not an empirical fact that there is law only where there is government, it being part of the concept$_o$ of law that instances of it presuppose the presence of government.

But sometimes when a sentence is described as "analytic", what is meant is that it is *formally* true. Even though no empirical knowledge, above such as is needed to grasp its meaning, is needed to determine the truth-value of "there is no law where there is no government", that sentence is not *formally* true. By contrast, a token of "I am here now" *is* formally true, as is a token of "if snow is white, then snow is either white or green." As we discussed earlier, a sentence-token T belonging to language L is *formally* true iff it is a theorem of the semantic rules of L that T is true.

Quine's statement that analyticity is to be understood in terms of synonymy is therefore ambiguous. Let us consider each disambiguation of Quine's statement.

The concept$_o$ of formal truth is *not,* at least not on the face of it, to be defined in terms of that of synonymy. (A sentence or sentence-token T is formally true if its truth is a theorem of the relevant semantic rules. There is no mention of synonymy here.) So Quine's argument implodes if by an "analytic" sentence, Quine means one that is formally true.

To be sure, a defender of Quine's position has what initially appears to be a compelling response to this. If the concept$_o$ of a semantic theorem, and thus of a formal truth, is to be understood in terms of the concept$_o$ of analyticity, then in defining an "analytic" truth as one that is formally true, one has defined analytic truth in terms of itself. This, of course, would vindicate Quine's view that the concept$_o$ of analyticity is a viciously circular and therefore incoherent one. And, at first, it does seem that the concept$_o$ of logical truth is to be understood in terms of the concept$_o$ of analyticity. After all, it seems at least approximately correct to say that S's truth is a theorem of the semantic rules of language L just in case it is analytic that, given the semantic rules of L, S is true.

But here we must be careful. Remember that there are two kinds of analyticity: there is formal truth and there is also conceptual$_o$ truth. Given this, suppose that $s_1...s_n$ are the semantic rules of L. Even if S is formally true in L, and is therefore analytic in that sense, the following sentence is not formally true in L:

(S_1) "given $s_1...s_n$, S must be true."

S_1makes a true statement *about* the semantic rules of L. For many reasons, it follows that S_1 cannot *itself* be assigned truth by those semantic rules. (Some of these reasons are strictly philosophical in nature, and others have to do with well-known results in formal semantics.) At the same time, S_1 obviously isn't empirical or a *a posteriori.* (S_1 is obviously analytic one of the two meanings of that word.) Thus, S_1 is either a formal truth of some language L_1 that is distinct from L, or S_1 is a non-formal, conceptual$_o$ truth. Let us consider each case. If S_1 is *not* a formal truth of any language, then S_1 is not analytic in the same sense as S. Whereas S is analytic in the sense of being formally true, S_1 is analytic in the entirely distinct sense of being conceptually$_o$ true. And, of course, there isn't anything circular about defining "analytic" (formally true) in terms of "analytic" (conceptually$_o$ true).

What about if S_1 is a formal truth of some language L_1? In that case, everything that we just said about S_1 is true of the sentence:

(S_2) "given $s^*_1...s^*_m$, S_1 must be true",

where $s^*_1...s^*_m$ are the semantic rules of L_1. And, of course, what we just said about S and S_1 might also hold of comparable sentences S_2, S_3, and so on. But this process cannot go on forever. Given any one language L_i, it can be coherently supposed that its expressions are given meaning by the expressions of some other language L_{i+i}. But it cannot be coherently supposed that this is true of *every* language. Linguistic meaning ultimately has non-linguistic sources. So it is incoherent to suppose that a formally true sentence, e.g. "either is is snowing or it isn't", is given meaning by expressions belonging to a language whose expressions are given meaning by those of yet another

language - and so on *ad infinitum*. This means that the sequence of sentences described earlier ($S, S_1, S_2...$) must come to an end, which in turn means that, for some finite n, S_n is analytic (on one disambiguation of that word) but *not* formally true.

Formal truth presupposes conceptual$_0$ truth. The concept$_0$ of a semantic theorem is to be understood in terms of conceptual$_0$ truth. The defender of Quine's argument is right to hold that the concept$_0$ of a semantic theorem is to be understood in terms of the concept$_0$ of analytic truth. But we are not on that account guilty of circularity in analyzing the concept$_0$ of analytic truth in terms of the concept$_0$ of a semantic theorem. For there is nothing viciously circular about analyzing analytic (formal) truth in terms of analytic (conceptual$_0$) truth.

Let us sum up before we move on. The concept$_0$ of analytic (formal) truth is to be understood in terms of the concept$_0$ of a semantic theorem. The concept$_0$ of a semantic theorem is to be understood in terms of the concept$_0$ of analytic (conceptual$_0$) truth. Obviously there is no circularity in analyzing the concept$_0$ of analytic (formal) truth in terms of the concept$_0$ of analytic (conceptual$_0$) truth.

But now the question arises whether the concept$_0$ of analytic (conceptual$_0$) truth must be understood in terms of itself. If so, then the concept$_0$ of analytic (conceptual$_0$) truth is viciously circular. And in that case, the concept$_0$ of analytic (formal) truth is also viciously circular, since that concept$_0$, as we just saw, is to be analyzed in terms of the concept$_0$ of conceptual$_0$ truth.

But the concept$_0$ of conceptual$_0$ truth is not to be analyzed in terms of itself, and is not otherwise viciously circular. The sentence "there is no law where there is no government" is analytic because it is conceptually$_0$ true. And it is conceptually$_0$ true because, *though true, it doesn't register any empirical facts*.

Thus, a conceptually$_0$ true proposition is one that is non-empirically true; and the concept$_0$ of conceptual$_0$ truth need not, and ought not, be understood in terms of the concept$_0$ of synonymy or of analyticity.

Quine might, and probably would, deny that there are non-empirical truths. But, in this context, such a denial would be question-begging. After all, the very thing that Quine is trying to show, in his attempt to eliminate the analytic-synthetic distinction, *is* that there are non-empirical truths. A sentence is analytically true just in case it is non-empirically true. Analytic truth is non-empirical truth. According to nominalism, all facts are spatiotemporal and there is thus no non-empirical truth. So in assuming the truth of nominalism, Quine is simply assuming that there is no analytic truth and is thus question-beggingly assuming the very thing that he wishes to establish.

Let us sum up. The thesis that there are no analytic truths is ambiguous, as it can mean either (a) that there are no formal truths or (b) that there are no conceptual$_0$ truths. Given either disambiguation, Quine's argument fails. Conceptual$_0$ truths are non-empirical truths. So the concept$_0$ of "analytic" (conceptual$_0$) truth isn't to be understood in terms of synonymy or, more importantly, in terms of itself; and that concept$_0$ is therefore free of any vicious circularity. The concept$_0$ of analytic (formal) truth is to be understood in terms of the concept$_0$ of a semantic theorem, and that concept$_0$, in its

turn, is to be understood in terms of the concept$_o$ of conceptual$_o$ truth. Given that, as we just saw, the latter concept$_o$ is free of vicious circularity, it follows that the same goes for the concept$_o$ of formal truth. Thus, Quine's thesis fails on each interpretation of it.

A brief look at philosophical history may shed some light on this topic. Leibniz held that all the laws of logic can be reduced to the law of identity (now called "Leibniz's Law"): if x and y are identical, then x has a given property if y has that same property. So, in Leibniz's view, all logical truths – e.g. *if p, and if p then q, then q* – are really instances of Leibniz's Law.

But Arthur Pap (1958) showed that in order to reduce truths of logic to instances of Leibniz's Law, one must presuppose the truth of logical principles *other* than Leibniz's Law and, consequently, that principles other than that law are ineliminable components of logic. Leibniz's thesis is therefore wrong.

In holding that all logical truths can be modeled on "all bachelor's are married", Quine is guilty of a fallacy similar to Leibniz's. (Chomsky (1988: 525) makes a similar point.) Not all logical (or analytic) truths express identities ("the class of bachelors is identical with the class of unmarried adult males") or partial identities ("the class of bachelors is a subset of the class of unmarried entities"). There is no non-contrived sense in which "legal obligation presupposes governmental protection" holds in virtue of some *synonymy*-relation between two expressions; and any equivalence between that sentence and an identity (or partial identity) could be shown only on the basis of analytic truths that are not identities. So, contrary to what Quine assumes, it is *not* the case that, if there are analytic truths, they can all be modeled on "all bachelors are unmarried" or on any other sentence whose analyticity derives from a relation of synonymy.

It is not hard to find independent corroboration for this criticism of Quine's argument. "All bachelors are unmarried" is trivial. The same is true of any sentence whose analytic status depends entirely on some instance of synonymy: "there are three feet in a yard", "women are adult female humans", "banks are financial institutions."

But many analytic sentences are not trivial, and non-trivial analyticities never hold merely in virtue of synonymy relations. Consider the following sentences: "there are continuous functions that cannot be differentiated at any point", "laws are narrow-scope governmental assurances of protections of moral rights"[6], "a symbol-token is an instance of a relation between morphology and socio-psychological context", "the minimum number of sides that a three-dimensional closed figure can have is five", "there is no complete and consistent formalization of arithmetic." The analytic status of any given one of these sentences cannot be understood in terms of the fact that one term is synonymous with another. Unlike "all bachelors are unmarried", these sentences do not merely register synonymies, and that is why they are all non-trivial.

There are trivial analyticities that *do not* merely register synonymies, e.g. "green is a color." (Attempts were made to show that "green" could be defined in terms of "color";

6. This was the thesis of my doctoral dissertation.

but these attempts failed, and it is clear that they must fail.[7]) But all non-trivial analyticities do more than register synonymies.

Quine's mistaken thesis that analyticity is to be understood in terms of synonymy can be understood as a distortion of the correct point that a sentence is analytic, on one disambiguation of that term, iff it is *semantically* true. Synonymy-based analyticities are among the more obvious cases semantic truths: "there are three feet in a yard" is analytically true because a "yard" is *defined* as a distance of three feet. But they are not the only examples of analytic truth.

Additional problems with Quine's argument

We said that, on one disambiguation of the term "analytic", a sentence is analytic iff it expresses a logically true proposition. Quine did not believe in propositions or in meanings generally, and would thus reject this characterization of analyticity.

At least part of the reason that Quine rejects the notion of meaning (and, thus, of propositionality) is that he is a nominalist: he believes in spatiotemporal individuals, but not in platonic abstracta – there are instances of properties, but no properties (except in so far as properties can be identified with, or otherwise reduced to, instances of properties). In Quine's view, there are noises and ink-marks, but no propositions (except in so far as propositions can be constructed out of noises, ink-marks, and the like).

My own view is that nominalism is wrong. (We've already seen that, in assuming the truth of nominalism, Quine is begging the question against those who believe in analytic truth. But we are going to set this aside here, so as to develop a new criticism of Quine's view.) But we don't have to refute nominalism to show that Quine's position is indefensible.

Consider the following definition of analyticity:

(*) a sentence (or sentence-token) S, belonging to language L, is analytic exactly if, given the semantic rules of L, and given no other empirical information, S's truth follows.

Notice that (*) doesn't involve the notion of a proposition, except in so far as the concepts *semantic rule* and *following from* involve that notion. But it would be absurd to deny that there are semantic rules, even though there are different reasonable views as to what semantic rules are. And it would be absurd to deny that statements ever follow from other statements. Somebody who denies that statement is thereby affirming the statement: *the statement that nothing follows from S follows from the statement that S is a statement*. Since there is analytic truth if anything follows from anything, it is therefore self-defeating to deny its existence, and somebody who denies that statements

7. Wittgenstein (1975) deserves credit for making this point. In Kuczynski (2006) I discuss why, in my judgment, Wittgenstein's insight warrants a non-atomistic conception of conception

ever follow from statements is affirming the negation of his own view. Thus, contrary to what Quine says, it is possible to give a non-circular analysis of the concept$_o$ of analyticity, and it is incoherent to deny that some truths fall into that category. Notice, by the way, that (*) covers both conceptual$_o$ and formal truth, and is thus a non-disjunctive, completely general analysis of analytic truth.

Most importantly, it is clear at an intuitive level that some truths are analytic. To know that there is pain only where there is sentience, one doesn't need empirical knowledge, over and above such as is required to grasp that proposition (or, if you don't believe in propositions, to understand the corresponding sentence).[8] Quine's argument is too wobbly to demand that we repudiate this intuition.

Quine would say this intuition is deceiving us, just as Kant's intuitions were deceiving him when they led him to repudiate the very idea of a non-Euclidean space, or a non-Newtonian mechanics, or a non-Aristotelian logic. But as Laurence Bonjour (1998: 5–7) and Michael Dummett (1973: 592–595) show, it is not hard to find demonstrable support for our intuitions. Bonjour and Dummett provide the same argument, which I will now state.

For the sake of argument, suppose that no statements are analytic. (The term "statement" may be taken to mean either *sentence* or *sentence-token* or *proposition*.) In that case, no statement of the form S_1 *confirms* S_2 *to degree* n is analytic: every such statement is synthetic, and can be established only on the basis of experience. By the same token, for any statement S_3 whose truth is empirically established, it is up to future empirical findings to determine what degree of probability S_3 gives to S_1 *confirms statement* S_2 *to degree* n. For an exactly similar reason, any future empirical finding S_4 will, without the assistance of yet *other* empirical discoveries, be incapable of confirming the statement that S_1 *confirms statement* S_2 *to degree* n *is confirmed to degree* m *by statement* S_3. And so on *ad infinitum*.

Nothing can confirm anything unless at least some confirmation-relations are analytic. This doesn't mean that this or that *specific* statement (e.g. "there is no government where there is no law", "there is no pain where there is no sentience") is analytic. (Actually, it does show that the statement "given that some statements confirm other statements, it is analytic that there are analytic truths" is analytic. But with a few exceptions of this sort, this argument provides no special support for the view that this or that specific statement is analytic.) But it does show that the class of analytic truths is non-empty.

This brings us to another argument against Quine's position. If the statement:

(NAT) there are no analytic truths

is analytic, then Quine's position is false. If (NAT) is synthetic, then empirical research is needed to show that it is true. So there can be no *demonstration* or non-empirical proof of (NAT) if that statement is true. Therefore that statement is non-demonstrable if

8. An argument similar to the one just given is found in Dummett (1978: Chapter 22).

true. Quine's argument is obviously not *empirical*. So if Quine is right to deny that there is analytic truth, then we must regard his argument for that position as fallacious.

Hempel on analyticity

Carl Hempel (1952: 26–28, 1965: 113–116) has provided an argument against the analytic-synthetic distinction that, although ultimately fallacious, is considerably more compelling than Quine's and that, unlike Quine's, gives us real insight into the concept$_o$ of analyticity.

Scientific concepts$_o$ are capable of partial (though almost never complete[9]) operational definitions. "x has length L" can be partially operationally defined as

(1) "x can be made to coincide with object y",

(where y is some other object, e.g. the standard meter-rod in Paris) or as

(2) "it takes light n nanoseconds to travel from one end of x to the other."

Of course, stipulative definitions are analytic. So supposing that O has length L, it is analytic, given (1) and (2), that:

(3) it takes light n nanosecond to travel from one end of O to the other,

and it is also analytic that

(4) O can be made to coincide with y.

But (3) and (4) are not analytically (or logically) equivalent. Given that the one is true, it does not analytically follow that the other is true. If (3) and (4) have the same truth-value, that is entirely a matter of *a posteriori* fact.

Any one operational definition is purely analytic. But when one attempts to give multiple (partial) operational definitions of a concept$_o$, one has to make sure that those definitions are empirically consistent with one another – otherwise the concept defined is incoherent. For example, there is nothing incoherent in supposing that light takes m (≠n) time to travel from one end of O to the other, but that O can be brought into coincidence with y. For the sake of argument, suppose that to be true. In that case, given both (1) and (2), the statement "O has length L" entails both that L can, and also cannot, be brought into coincidence with y. To avoid saddling the term "length" with a broken and incoherent concept$_o$, we must make sure that our various operational definitions of that term are mutually consistent: and that can be done only through empirical work.

Supposing that we (partially) define "length" in terms of (1) and (2), but we haven't yet established the empirical equivalence of those definitions, we are left with a situation where a true statement ("O has length L") analytically entails two other statements,

9. See Hempel (1965: 123-134).

namely (3) and (4), but where it remains an *empirical* question whether both of those statements are compatible. But, if it is supposed that there is a sharp distinction between analytic and synthetic statements, this is not a tenable situation. If S_1 is true, and analytically entails S_2 and S_3, then it is *ipso facto* settled whether both of S_2 and S_3 are true. The distinction between the analytic and the synthetic appears to have broken down; and the statement "if O has length L, then it takes light n nanoseconds to travel from one end of O to the other" isn't non-arbitrarily categorized as either analytic or synthetic.

What was just said about "length" is true of each member of an extensive and important class of expressions. Many terms begin as expressions of lay-discourse and are later appropriated by science. Having been thus appropriated, they are redefined, and usually thereby rendered both more precise and more comprehensive in extension, in light of scientific theories. Terms like "heat" and "temperature" have existed as long as language itself. But, as they are used by the lay-person, those expressions seem to have no determinate meanings outside the extraordinarily narrow corridors of day-to-day human experience.

Physics has established that, for any number n that corresponds to a temperature that can be directly experienced by human beings, a substance's having certain molecular properties is necessary and sufficient for its having a temperature of n°. (This is actually not quite accurate. But in this context that slight innaccuracy is irrelevant.) On this basis, the term "temperature" can be identified with, and thus defined in terms of, certain molecular properties. This greatly increases the range of things to which determinate temperatures can be assigned. It also brings statements of the form ⌜x has temperature n°⌝ into systematic connection with powerful and well-organized bodies of scientific theory, thereby increasing the predictive and explanatory import of such statements. It also gives greater precision to such statements: previously they were defined either in terms of instrumental operations whose outcomes were not a function only of the relevant object's temperature (whether a column of mercury rises by a certain amount in a thermometer is only partly determined by the temperature of the substance in which the thermometer is immersed) or, even worse, were understood in strictly phenomenological terms (in terms of what the object in question "felt like").

The prescientific usage of the term "temperature" lacked the uniformity of usage characteristic of rigorous discourse and, at least arguably, failed to determine a unique concept. The process of assimilating that term, and other terms in its category (e.g. "heat", "freezing"), into science generated a situation exactly parallel to that described a moment ago in connection with "length." What we have said about "length" and "heat" is true of "mass", "acid", and practically any expression denoting an explanatorily significant class of phenomena (or objects). Thus, given any one of these terms, an argument exactly analogous to that given in connection with "length" points to the existence of a breakdown of the distinction between analytic and synthetic truth.

Evaluating Hempel's argument

Let us start with a purely exegetical point. Hempel's primary concern is not to discuss the viability of the analytic-synthetic distinction, but only to describe scientific methodology. (His remark about the analytic-synthetic distinction is almost made in passing.) From that viewpoint, Hempel's analysis is unexceptionable. But we can accommodate Hempel's correct methodological point without rejecting the traditional view that all truths are either analytic or synthetic (and without rejecting the exceedingly plausible view that at least some truths are analytic) and without therefore embracing the subsequent revisions of classical logic that would be demanded by a rejection of that view.

First of all, we must distinguish reference-fixing from meaning-giving. An operational definition can be seen not as giving the meaning of "length" (or "temperature" or "mass"...) but rather as fixing its referent. Suppose that Smith is the one person who came to our party yesterday wearing a ridiculous looking plaid sportcoat; and suppose that Smith is also the one person who gave money to the charity that Brown started. In that case, the statements:

(5) "Smith is the one person who gave money to the charity that Brown started"

and

(6) "Smith is the one person who came to our party yesterday wearing a ridiculous looking plaid sportcoat"

fix the referent of Smith. There is some x such that, given *either* (5) or (6), "Smith is bald" means: *x is bald*. At the same time, (5) does, whereas (6) does not, entail that Smith gave money to Brown's charity; and (6) does, whereas (5) does not, entail that that Smith wore a plaid sportcoat on a certain occasion. Each of (5) and (6) is a correct definition of "Smith." At the same time, given the statement "Smith is bald", it follows from (5), but not from (6), that a bald person gave money to Brown's charity. So we seem to have a situation where "Smith gave money to Brown's charity" both does, and does not, entail "a bald person gave money to Brown's charity", and we therefore have a situation that precisely parallels that described in connection with "length."

But obviously there is no breakdown of the analytic-synthetic distinction. The content of (5) is perspicuously given by the statement:

(5*) Somebody x uniquely gave money to Brown's charity and "Smith" names x,

and the content of (6) is perspicuously given by the statement:

(6*) Somebody x uniquely wore a ridiculous plaid sportcoat on a certain occasion, and "Smith" names x.

What (5*) and (6*) make clear is that, where each of (5) and (6) is concerned, the definition of "Smith" actually being given is simply: *"Smith" names x*, for some x. The reason that (5) does, whereas (6) does not, entail that somebody gave money to Brown's charity

is that (5) is *not* just a definition: it consists of a synthetic statement ("somebody x uniquely gave money to Brown's charity") conjoined (to put it loosely) with a definition ("x is Smith"). The purely definitional component of (5) is given by a bare singular proposition (*x is named "Smith"*); the remaining part of (5) is part of the *pre*-definition, as opposed to the definition *per se*, of "Smith." What we just said about (5) is *true mutatis* mutandis of (6). So, ultimately, each of (5) and (6) defines "Smith" in the same way, namely: *"Smith" names x*, for some x. But (5) and (6) convey different *pre*-definitional information, and this creates the conundrum, or illusion thereof, just described.

For obvious reasons, expressions cannot be defined in a vacuum; definitions must be made in terms of phenomena that are known to both speaker and auditor. But it doesn't follow that those phenomena always enter constitutively into the definition itself. In some cases, they are among the means used to give the definition, and are not internal to the definition itself.

What we just said about "Smith" is true of "length", "mass", "heat", "hardness", and other such terms. Each of (5) and (6) can be seen as *fixing* the referent, not giving the meaning, of the term "length." If this is right, then the content of (1) is given by the statement:

(1*) There is some property P such that anything x (of the appropriate shape) has P exactly if x can be brought into congruence with object y, and "length L" refers to P.

And the content of (2) is given by the statement:

(2*) There is some property P such that anything x (of the appropriate shape) has P exactly if it takes light n nanoseconds to travel from one end of x to the other, and "length L" refers to P.

So where each of (1) and (2) is concerned, the strictly definitional component is given by the bare singular proposition: *the expression "length L" refers to P*. It is true that (1) doesn't have the same analytic consequences as (2). But this doesn't indicate that a statement can simultaneously entail, and fail to entail, some other statement. It merely corresponds to the banal fact that different synthetic statements have different consequences.

The distinction between reference-fixing and meaning-giving has far-reaching implications as regards scientific methodology. One example of this concerns a doctrine called "operationalism." According to this doctrine, a term T is scientifically meaningful iff it is operationally definable, i.e. if for any object (or ordered n-tuple of objects) x, there is some observable phenomenon E and some observable operation O such that T describes x iff E results when O is applied to x.

Hempel (1965: 123–134) argued compellingly that operationalism is untenable. But the term "operationalism" is ambiguous, depending on whether the term "define" is taken in the reference-fixing or meaning-giving sense. If it is taken in the latter sense, Hempel's arguments against operationalism are entirely cogent. But those arguments are not cogent if "define" is taken in the former sense. At the same, "operational-

ism" is eviscerated if it is taken to be the view that an expression T denotes a scientifically respectable concept exactly if observable outcomes of observable procedures may *fix* the referent of T. It is easily seen that, relative to that definition of "definition", there is *no* unique operational characterization of *any* phenomenon that has observable effects and that, consequently, operationalism is reduced to the triviality that, in some cases, a theoretical (or otherwise unobservable) phenomenon may have an observable effect. So Hempel's arguments against "operationalism" *are* cogent, so long as that term is taken in a way that does not eviscerate it (and this of course means that Hempel's arguments are cogent *tout court*).

The ambiguity of "ambiguity"

But even if we leave aside everything just said – even if we set aside the distinction between meaning-giving and reference-fixing – we must still reject Hempel's argument concerning the analytic-synthetic distinction. If "length L" is defined in two non-equivalent ways, then 'x has length L' will be ambiguous between two non-equivalent propositions. Obviously there will be statements that follow from the one disambiguation that do not follow from the other. But this would no more warrant a revision of our views concerning the analytic-synthetic distinction than is warranted by the fact that, on exactly one of its two disambiguations, "John went to the bank" entails that John went to a financial institution. Hempel's argument thus embodies a failure to take into account the pedestrian fact that some expressions are ambiguous.

The term "language" is ambiguous between "diachronic" and "synchronic" meanings, to use Saussure's (1966) terminology. The statement:

(7) "The English language came into existence after the year 1100 A.D."

is true on one disambiguation, but false on another. If by "the English language" is meant the set of semantic rules that, at this instant in time, characterize the language spoken in Australia, England, and so on, then that statement is false. But that statement is true if by "the English language" is meant a series of sets of semantic rules such that, first, the set of rules operative in Australia, New Zealand, and so on, is a member of that series and such that, second, there is a certain kind of spatiotemporal continuity between any two installments in that sequence. So (7) is false on the "synchronic" meaning of "the English language", but true on the "diachronic" meaning of that expression.

Words change their meanings. (There is "semantic shift", as linguists put it.) This inevitably creates ambiguity since, during the period of semantic shift, some people continue to use the term in the old way. (In fact, semantic shift usually, though not necessarily, occurs in consequence of an absence of uniformity of usage.) But while the ambiguities associated with words like "bank" and "dumb" are easily recognized as such, and thus tend not to lead to confusion, the ambiguities created by semantic shift are less likely to be recognized as such, and what are in fact purely terminological dif-

ferences come to be seen as having a substantive dimension. (We might say that "bank" is an instance of "synchronic ambiguity", since its ambiguity is not a reflection of *change* in its usage, and that "temperature" is an instance of "diachronic" ambiguity, since its ambiguity between molecular, phenomenological, and other concepts *is* a reflection of language-change.) Synchronic ambiguities are (relatively) stable: "bank" has meant both *river's edge* and *financial institution* for hundreds of years. By contrast, diachronic ambiguities are unstable, almost by definition, and they typically quickly give way to uniformity of usage and thus to synchronic *non*-ambiguity.

For obvious psychological reasons, if a term is used non-uniformly on Monday and uniformly on Tuesday, people will tend to regard as deviant those Monday-usages of the term that don't conform to its Tuesday-usage. Semantic shift thus tends to create the illusion that, during the unstable interim-period, people were simply using the expression in question wrongly.

A related point is that people tend to mistake continuous changes for identities. So, because semantic shift usually happens continuously, there is often little recognition that it has occurred. This creates the illusion that the term has *not* changed meaning, and the differences between pre- and post-shift usage are then seen as embodying changes of a substantive, as opposed to purely verbal, nature.

Hempel's argument overlooks the ambiguity of the word "ambiguity." Scientific terms are in a state of flux, since they are redefined in consequence of changes in scientific theory. This creates ambiguity. But because, for the reasons just described, the ambiguities in question are not comparable to those characteristic of paradigms like "bank" and "dumb", we don't recognize them as such. As a result, what is actually an innocent fact (e.g. that, at a certain point in time, "temperature" was ambiguous between optical and tactile concepts) appears to have far-reaching ramifications for logic ("no sharp line between analytic and synthetic truth").

We must also keep in mind that even if Hempel has shown that the analytic-synthetic distinction breaks down where *some* expressions are concerned, nothing he has said warrants the view that it breaks down universally or that it otherwise breaks down in a manner that would vitiate our use of that distinction.