

What Is The Connection Principle?

Jerry Fodor
Ernie Lepore
Rutgers University

*Euch Mach ihr's leicht, mir macht ihr's schwer.
-Die Meistersinger, Act 3.*

The Connection Principle (hereafter, CP) says that there is some kind of internal relation between a state's¹ having intentional content ("aspectual shape") and its being (at least potentially) conscious. Searle's argument for the principle is just that potential consciousness is the only thing he can think of that would distinguish original intentionality from ersatz (Searle, 1992, pp. 84, 155 and passim. All Searle references are to 1992). Cognitivists have generally found this argument underwhelming given the empirical successes recently enjoyed by linguistic and psychological theories with which, according to Searle, CP is not reconcilable. Our primary interest in this paper is *not*, however, to decide whether CP is true, but just to get as clear as we can about what exactly it asserts. Finding a reasonable formulation of the principle turns out to be harder than Searle appears to suppose; or so we claim.

A reasonable formulation of CP must be *plausible* -- that is, not susceptible to obvious prima facie objections -- and it must be *tendentious* -- that is, prima facie incompatible with standard Cognitivist views and claims (See pp. 151, 153, 173). Consider, for example, a formulation according to which CP says that *there can be no intentional zombies*² (an intentional zombie is a creature which has intentional states but has and can have no conscious states at all). We agree that it is philosophically interesting whether there can be intentional zombies, and we agree that it is prima facie plausible that there can't be. But we take it that the claim that there can't be, since it isn't tendentious, isn't an adequate formulation of CP. Prima facie, at least, nothing a Cognitivist wants to say about the psychology of the intentional requires the possibility of intentional zombies. Cognitivism³ is *neutral* on this since all the unconscious states that Cognitivists actually want to postulate are, as far as anyone knows, states of creatures that are conscious at least from time to time. It's true, of course, that nothing Cognitivists say about intentionality *explains why* intentional zombies are impossible (if they are); but nothing Searle says

¹ Or event or whatever; for present purposes, we have no views about what sorts of individuals the ontology of a theory of mind should acknowledge, so we'll speak as convenience suggests. Generally we'll assume that mental states and events are the tokenings of mental properties.

² This is, of course, an issue around which considerable discussion has turned. See, for example, Block (1980).

³ There are, metaphysical theories around which are committed on this issue. For example, all of the current approaches to "naturalizing" intentionality assume that sufficient conditions for content need not also be sufficient for consciousness. But, of course, the viability of the empirical psychological explanations that Cognitivist psychology offers does not depend upon the success of these metaphysical programs.

about intentionality does either. Maybe Searle is right that what distinguishes original intentionality from ersatz is potential consciousness. But he has no story about why it shouldn't be relative humidity or time of day instead. More on this later.

We're going to argue that Searle hasn't managed to formulate a version of CP that is both plausible and tendentious; and that, lacking some detailed proposals about just why it is that intentionality is supposed to presuppose consciousness, it's unlikely that anyone will. Before we get to the main argument, however, we need to consider two questions that any formulation of CP must take a stand on, but about which Searle is surprisingly noncommittal in his book. We think that once these questions are made explicit, it's pretty clear how they have to be answered. We hope Searle would agree, since the arguments we'll offer later depend on this.

Q1. *What's the modality of CP?* That is, what's the force of the "can't" in "There can't be an intentional state that isn't potentially conscious"?

We're going to take it for granted that the modal in CP must express *nomologically* or *metaphysical* necessity. In particular, CP can't be merely the claim that it is always *conceptually* possible that a de facto unconscious state should be conscious, for that claim too is untendentious. Suppose, for example, that Chomsky⁴ is right: everybody who learns English must unconsciously contemplate (indeed, must unconsciously endorse) the proposition that NP is a bounding node. Clearly, all the events of considering and accepting this proposition that have thus far transpired in the course of language learning have been unconscious. But, equally clearly, it isn't part of Chomsky's story to deny that they might have been otherwise; that is, it is untendentious that it is intelligible that they *might* have been otherwise.

In fact, it's not clearly tendentious even that it is *nomologically* possible that such events should be conscious. Suppose that Johnson and Johnson were to invent a consciousness pill: one that makes the mental processes involved in learning and using a language all introspectively accessible. This would, no doubt, occasion massive technological unemployment among linguists and psycholinguists, since you could then determine whether NP is a bounding node by just asking an English speaker. But it is, to put it mildly, not clear that a consciousness pill would refute cognitivist linguistics. As far as we can tell, cognitivist linguistics is simply *neutral* about whether a consciousness pill is nomologically possible.⁵ If so, then it is *neutral* about whether it is nomologically necessary that the thinking about bounding nodes that goes on in learning English is unconscious.

It would, in any case, clearly be untendentious for CP to assert only the *conceptual* possibility that any (de facto unconscious) intentional event could be conscious. So we assume that CP must claim at least that it is a matter of nomological or metaphysical

⁴ . The Chomsky whose views we'll be concerned with isn't the famous linguist but an arbitrary Cognitivist whose opinions are typical of the kind. Since this Chomsky didn't exist, we had to invent him.

⁵ Compare, however, Freudian theories, according to which the unconsciousness of at least some intentional states is essential to their proper functioning. Perhaps Freudians would claim that such states are ipso facto unconscious in all the nearby worlds in which our psychology is like what it's like in this world. It may be, then, that some cognitivist theories hold that the unconsciousness of some kinds of intentional states is nomologically necessary not qua intentional states but qua states which play a certain role in mental life. This sort of possibility will be important later in the discussion.

necessity that this is so, and we assume that Searle would want us to assume this. As far as we can tell, Searle's view is that it is conceptually necessary that it is nomologically possible that any (de facto unconscious) intentional state should be conscious.

Q 2. *What's the ontology of CP?*

If CP is to be tendentious, it must apply to mental state *tokens*; merely applying to types would be too weak. Chomsky, for example, makes a living by having conscious thoughts that are type identical to the unconscious ones that are supposed to go on in language learning and language use; such thoughts as that NP is a bounding node in English, for example. He would therefore find untendentious a formulation of CP which requires only that *for each de facto unconscious token of a certain thought type, it must be nomologically possible that some other token of that type ---specifically, some other token with the same intentional content--- should be conscious.*⁶ So then, to summarize: to be tendentious, CP needs to be a claim about counterfactual (i.e., "cross-world") token identities. Specifically, it has to claim that it is nomologically possible that *any de facto unconscious token intentional event (/state) might be (/have been) conscious.* In what follows, we'll refer to this sharpened formulation of CP as CPF.

CPF will, no doubt, cause alarms to go off, since it makes CP hostage to the conditions for counterfactual token identity, and these are notoriously buckets of worms. But CPF isn't a polemical contrivance; we truly can't think of a way to formulate CP that would succeed in being tendentious without raising these issues. And, knowing Searle as we do, we're sure that tendentious is something he intends CP to succeed in being.

We now propose to argue that CPF, though tendentious, lacks plausibility. It is *prima facie* unenforceable for reasons that have nothing to do with whether there is an internal connection between consciousness and intentionality.

3. Main argument

step 1: counterfactual intentional event identities conserve (some) mental kinds.

Some things about the conditions for counterfactual token identities among intentional states and events are straightforward. For example, it's clear that event e1 isn't counterfactually identical to event e2 unless e1 and e2 have the same intentional content.⁷ There is thus no possible world in which my current yen for a chocolate ice is a yen for a T-bone steak. The best you could get along these lines is a world in which I have a T-bone yen instead of a chocolate ice yen.

⁶ Searle does sometimes write as though it is only the weak, "type" version of CP he intends to defend. "...we understand the notion of unconscious mental state only as a possible content of consciousness, only as the sort of thing that, though not conscious, and perhaps impossible to bring to consciousness for various reasons, nonetheless is the sort of thing that could be or could not have been conscious." See the discussion pp. 155-162, *passim*.

Just as cross-world token identity constrains the content of a mental event, so too it constrains (what we'll call) its "mental kind"; mental kind properties include: being a belief; being a desire; being a whim; being a yen; being a doubt; etc. So, for example, I suppose my current doubt that it's going to rain couldn't have been a belief that it's going to rain. If I had believed (rather than doubted) that it's going to rain, I would have been in a different token state than the one that I actually am in.

So then: there are cases where cross-world mental event pairs fail to be identical tokens because they belong to different mental event kinds. Just how counterfactual token identity is constrained to respect mental kinds we don't know.⁸ Could my current yen for a chocolate ice have been a raging lust for a chocolate ice, compatible with its being the very same event it is? Search us. But the general (Geachian) point holds: sameness of individuals is intricately involved with sameness of kinds; and the type identification of token intentional states depends both on their content and on what kind of mental states they are.

step ii: Mental kind properties are functional (at least inter alia). It is generally supposed, and it seems to us entirely plausible, that the mental kind of a mental state is (at least in part) determined by its causal powers. What makes a state a yen for rain, rather than a belief that it's going to rain or a doubt that it's going to rain, is something about its actual and possible causal role in a creature's mental life. More generally: At least some differences in causal powers are sufficient for at least some differences of mental kind.

Notice that this is not to claim that the intentional content of a mental state is in any way determined by its causal role. Functionalism about content Searle would surely consider question-begging; and we, in fact, don't believe it to be true. For present purposes, we allow any account of the determination of content that you like, including the one according to which what determines content is sui generis and irreducible.

So the metaphysics of the difference between believing and doubting that P is functional inter alia; but the metaphysics of the difference between believing that P and believing that Q is what you will. It is, of course, a famous Searlean doctrine that what distinguishes a belief from a desire with the same intentional content is their "direction of fit". We take this to be compatible with holding that what distinguishes them is functional role. In fact, it's hard to imagine what the direction of fit of an intentional state could amount to except the use to which its content is put: If one believes that P, one is disposed, ceteris paribus, to act on the assumption that P is true; whereas, if one desires that P, one is disposed, ceteris paribus, to act so as to make P true. You can't, we suppose, run "inverted fit" arguments

corresponding to inverted qualia arguments; a doubt that P that is functionally indistinguishable from a conviction that P wouldn't be a doubt.

step iii: Consciousness isn't epiphenomenal. Unless you are an epiphenomenalist, and whether or not you believe in CP, you have to suppose that the consciousness of a state does (or, anyhow, can) affect its causal powers. We won't argue for this claim, since Searle is surely the last person in the world who would wish to deny it (Searle, 1993, p. 126). However, we remark in passing that what we are assuming is stronger than what we need. It would do for our purposes that the determinants of consciousness are also determinants of causal powers in at least some cases.

So, according to us, it's nomologically necessary that whatever (neural?) property of a state makes the difference between its being +conscious and its being -conscious, also does, or can, determine aspects of the causal role that the state plays in mental life.

step iv: CPF isn't plausible. We now have the following situation: token mental states can depend for their cross-world identity on their mental kind. Mental kinds are constituted by their causal powers. Epiphenomenalism is false, so making an unconscious state conscious typically affects its causal powers. Hence it is open that making an unconscious state conscious could alter its mental kind. Hence it is open that the token identity of a de facto unconscious intentional state might not survive its becoming conscious even if there is an internal connection between consciousness and intentionality. If there is just one case in which the cross-world identity of an intentional state token depends on its consciousness in this way, then CPF cannot be enforced.

Is it plausible that there actually are such cases? Absolutely. It will do, to show that it is, if we can find an example where the difference between +/- conscious is associated with very substantial differences in the causal powers of an intentional state; for, the more substantial this difference in causal powers is, the more likely it is to constitute a difference in mental kind that is incompatible with cross-world token identity.

There are many such cases. Perhaps, in fact, they are the norm; "consciousness does matter," (p. 18) as Searle is forever reminding us. Consider, for example, that old psycholinguistic standby the "tip of the tongue" effect (TOT). You're trying to remember a word but, you can't quite. So you guess. It turns out to be much more likely than chance that your guesses will have the same initial letter, the same number of syllables, and the same stress pattern as the word that you're "looking for", and that they will belong to the same semantic field. (See Brown and McNeill (1966)).

Here is a standard etiological story about TOT.⁹ Each word you know there is a

corresponding mental representation, a "logogen". Logogens have real numbered excitation values between, say, 0 and 1.

Their effects on one another, and on other mental entities/states, vary as a function of variation of their excitation levels. Each logogen also has a "consciousness threshold". If its excitation level is above threshold, its activity is consciously accessible; if not, then not. TOT effects are caused by logogens whose level of excitation, though subthreshold, are nonnegligibly higher than their rest states. What happens when you find the word that was on the tip of your tongue is that the corresponding logogen crosses its consciousness threshold.

Suppose that some such story is true. Then our point is that, concomitant with the difference between being a logogen below its consciousness threshold and being the same logogen above its consciousness threshold, there is a comprehensive difference in the logogen's causal role. Conscious states of logogens obey radically different laws from the corresponding unconscious ones, and they predict (hence explain) radically different sorts of behaviors. Suppose, as we've been claiming, that cross-world token identities must preserve at least some mental kinds. And, suppose, as we have also claimed, that mental kinds are constituted by causal powers. It then seems perfectly possible that being below its consciousness threshold is an essential property of the states that have it; i.e., that any event token that consists of a logogen's having an excitation value that is below the consciousness threshold, will have an excitation value that is below that consciousness threshold in every world that it occurs in. A fortiori, TOT would violate CPF. No event that actually did consist of thinking of a word unconsciously could have consisted of consciously remembering that word. No event that actually did consist of having the word on the tip of one's tongue could have consisted of having that word on one's tongue tout court.

Two comments before we conclude: The first is that nothing hangs on the example. We don't know whether an event which actually consisted of a logogen's being at a subthreshold level of activation could have consisted of that logogen's being at a superthreshold level of activation. Such questions about cross-world identities make our heads hurt; we aren't at all sure that they have answers. Our point is that CPF requires that they do; in particular, it requires that the difference between being +conscious and being -conscious can never be sufficient for a difference of causal powers that is in turn sufficient for a difference of mental kind of the sort that cross-world token identity doesn't tolerate. (See Appendix, where this is spelled out.) But there is simply no reason at all to suppose that this can't happen. And if it can, then there are unconscious intentional state tokens which could not have been conscious. This is true independent of the question whether the link between consciousness and content is intrinsic.

Compare: A shout that P and a whisper that P both have the content that P; yet, quite likely, some events that were the one could not have been the other. (If Paul Revere had only whispered "The British are coming," he would never have gotten into a poem.) This is so independent of the question whether there are any intrinsic connections between content and volume.

The second comment is that it would be natural, indeed it would be correct, to complain that the argument we've been pursuing raises only a technical objection to CP; that all it really shows is that CPF won't do. Precisely so. But technical objections are nonetheless objections, and, as we remarked above, we see no way of formulating CP that avoids the worries that we've raised for CPF. Once you admit, as even Searle is apparently prepared to do, that some mental happenings, though really intentional, are de facto unconscious¹⁰, the problem of unpacking the modal force of CP will have to be faced. And with it comes the sort of nasty puzzles about cross-world token identities that we've been considering. It's a recurrent theme in Searle's book that the metaphysics of mind is easy. We should all live so long.

Conclusion: The moral isn't, of course, that the connection principle can't be stated. No doubt it can be; no doubt it eventually will be; and, not much doubt, when it finally is, it will prove not to be true. Rather, the moral is that you can't back into stating the connection principle in the way that Searle apparently wants to do. Searle has on offer no positive account at all of the putative connection between intentionality and consciousness; as we remarked, his argument for CP is just that he can't think of anything else that original intentionality has that ersatz intentionality can't have. But since that sort of argument offers no clue as to what it is about consciousness and intentionality that makes CP true, it offers no help in formulating a constraint on unconscious intentionality that's both plausible and tendentious.

Here's another way to put much the same point. What Searle really wants to claim is perhaps something epistemological: that you can't explain intentionality unless you understand consciousness. "...any discussion of intentionality that leaves out the question of consciousness will be incomplete," (p. 132, our emphasis, see, also, p. 84). The trouble with putting CP that way is, however, that it rests more weight on explanation than that notoriously pragmatic concept will bear. Maybe you would need a story about consciousness to explain intentionality completely, but so much the worse for complete explaining. Completely explaining anything requires completely understanding everything, for all that we know.

On our reading, Searle's way out of this was to formulate CP as an ontological thesis, a thesis about counterfactual event identity: Nothing could be an unconscious intentional state token unless that very same intentional

state token could have been conscious. This helps in one way; identity, unlike explanation, doesn't come in more or less. But the price, as we've seen, is paid in plausibility. It seems entirely possible that some intentional state tokens have their (un)consciousness essentially; i.e., preserving their (un)consciousness is a constraint on their cross-world identification. If this ever happens, then CPF fails as a formulation of CP; and it's not at all obvious what one might replace it with.

As far as we can see, as things now stand, CPF is implausible and the connection principle simply isn't clear enough to be worth the bother of arguing about. Making it clear enough would involve making a positive proposal about what the putative relation between intentionality and consciousness consists in. But, of course, neither Searle nor anyone else has a hint, smidgen or suggestion of such a proposal on offer. And we're not, actually, expecting to hear of one by the next post.¹¹

APPENDIX

CPF rules out the following possibility:

(i) There is a world in which the conscious intentional event e' corresponds to the unconscious intentional event e in the actual world.

(ii) Events e and e' have the same intentional content, but e has causal powers p and e' has causal powers p' .

(iii) Having causal powers p is sufficient for being of mental kind k ; having causal powers p' is sufficient for being of mental kind k' .

(iv) $(x) (y) (e \text{ is of kind } k \text{ in World } x \text{ and } e' \text{ is of kind } k' \text{ in World } y)$
---> $e \neq e'$

Since there is no reason to suppose there can't be cases in which (i)-(iv) are all true, there is correspondingly no reason to suppose that CPF can be enforced.[PB]

FOOTNOTES

1.

2.

3. There are, of course, metaphysical theories around which are committed on this issue. For example, all of the current approaches to "naturalizing" intentionality assume that sufficient conditions for content need not also be sufficient for consciousness. But, of course, the viability of the empirical psychological explanations that Cognitivist psychology offers does not depend upon the success of these metaphysical programs.

4. The Chomsky whose views we'll be concerned with isn't the famous linguist but an arbitrary Cognitivist whose opinions are typical of the kind. Since this Chomsky didn't exist, we had to invent him.

5. Compare, however, Freudian theories, according to which the unconsciousness of at least some intentional states is functional. Perhaps Freudians would claim that such states are ipso facto unconscious in all the nearby worlds in which our psychology is like what it's like in this world. It may be, then, that some cognitivist theories hold that the unconsciousness of some kinds of intentional states is nomologically necessary not qua intentional states but qua states which play a certain role in mental life. This sort of possibility will be important later in the discussion.

6. Searle does sometimes write as though it is only the weak, "type" version of CP he intends to defend. "...we understand the notion of unconscious mental state only as a possible content of consciousness, only as the sort of thing that, though not conscious, and perhaps impossible to bring to consciousness for various reasons, nonetheless is the sort of thing that could be or could not have been conscious." See the discussion pp. 155-162, *passim*.

The trouble is that if this is what Searle is arguing for, he won't find anyone to argue with.

7. The same constraint holds, of course, for the cross-world counterpart relation. In general, nothing we say bears on this.

8. For the moment, we're leaving open whether the difference between +conscious and -conscious could itself be (or be a sufficient condition for) a difference of mental kind. We're going to argue that there is no reason to suppose that it can't.

9. For an early version of this sort of story, see Morton (1969). For recent elaborations, see MccLelland and Elman (1986). There are other stories about the etiology of TOT, some of which we actually prefer. But this will do for expository purposes.

10. Searle is unequivocal about this: "intentional states, conscious or unconscious, have aspectual shapes..." (Searle, 1993, p. 161). The view is

therefore NOT that only the former have original content, the intentionality of the latter being merely derived.

Searle is wise not to hold that unconscious intentionality is ipso facto ersatz, for that would entail that only what is within the specious present can be literally an intentional cause. It would thus NOT be literally the case that I put the sugar in to make the coffee sweet unless the thought "this will sweeten it" flitted through my consciousness as I did so.[PB]

REFERENCES

Block, N. (1980), "Are absent qualia impossible?" *Philosophical Review*, 89, 257-274.

Brown, R. and McNeill D. (1966) "The tip of the tongue phenomenon," *Journal of Verbal Learning and Verbal Behavior*, 5, 325-327.

MccLelland, M. and Elman, J. (1986) "Interactive processes in speech production," *Parallel Distributed Processes*, Vol. 2, 58-121. MIT Press, Cambridge.

Morton, J. (1969) "Interaction of information in word recognition," *Psychological Review*, 76, 165-178.

Searle, J. (1993) *The Rediscovery of the Mind*, MIT Press, Cambridge.