8. Conclusion

This discussion is the upshot of the application of two principles. Always ask yourself: What do you know for sure? and: What facts are supposed to correspond to the claims you are making? Now, as far as the inside of the skull is concerned, we know for sure that there is a brain and that at least sometimes it is conscious. With respect to those two facts, if we apply the second principle to the discipline of cognitive science, we get the results I have tried to present.

ACKNOWLEDGMENT

I am indebted to a very large number of people for helpful comments and criticisms on the topics discussed in this article. I cannot thank all of them, but several deserve special mention; indeed many of these patiently worked through entire drafts and made detailed comments. I am especially grateful to David Armstrong, Ned Block, Francis Crick, Hubert Dreyfus, Vinod Goel, Stevan Harnad, Marti Hearst, Elisabeth Lloyd, Kirk Ludwig, Irvin Rock, Dagmar Searle, Nathalie van Bockstaele, and Richard Wolheim.

NOTES

1. Chomsky, Noam (1976): "Human action can be understood only on the assumption that first-order capacities and families of dispositions to behave involve the use of cognitive structures that express systems of (unconscious) knowledge, belief, expectation, evaluation, judgment, and the like. At least, so it seems to me (p. 24). These systems may be unconscious for the most part and even beyond the reach of conscious introspection" (p. 35).

Among the elements that are beyond the reach of conscious introspection is "universal grammar" and Chomsky says: "Let us define universal grammar (UG) as the system of principles, conditions, and rules that are elements or properties of all human languages not merely by accident but by necessity – of course, I mean biological, not logical, necessity" (p. 29).

2. The argument here is a condensed version of a much longer development in Searle (1989). I have tried to keep its basic structure intact; I apologize for a certain amount of repetition.

3. I am indebted to Dan Rudermann for calling my attention to this article.

4. For these purposes I am contrasting "neurophysiological" and "mental," but in my view of mind/body relations, the mental simply is neurophysiological at a higher level (see Searle 1984a). I contrast mental and neurophysiological as one might contrast humans and animals without thereby implying that the first class is not included in the second. There is no dualism implicit in my use of this contrast.

5. Specifically, David Armstrong, Alison Gopnik, and Pat Hayes.

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Consciousness and accessibility

Ned Block

Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139
Electronic mail: *block@cogito.mit.edu*

Searle's Connection Principle says that unconscious mental states must be in principle accessible to consciousness. If deep unconscious rules, representations, states, and processes are not in principle accessible to consciousness, then they are not mental. I don't think that many in the cognitive science community care whether these phenomena are *mental* or not; the important point is that they are representational. But since Searle's argument applies as well to representationality as to mentality, we can move on to the real issues.

What does Searle mean by "accessibility in principle?" One of the real issues of which I speak is what "in principle" comes to. (Another, to be discussed later, is what consciousness is.) Searle clarifies his stance by describing people (I will call them the Less Conscious People) who have a desire for water, despite there being a "blockage" that prevents this desire from having any disposition to become conscious. What makes the Less Conscious People's desire in principle accessible to consciousness? The answer, as I read Searle, is that the Less Conscious People have the same brain state that is the "I want water" configuration in us. We have the "I want water" brain state just in case we want water, and we satisfy the Connection Principle since we can become conscious of our desire for water. So the presence of the "I want water" brain state in them justifies ascribing the desire for water to them. Though they are unable to become conscious of their desire for water, we should think of their desire as in principle accessible to consciousness because its brain state gives rise to awareness of the desire for water in us. As Searle notes, similar reasoning would justify ascribing visual knowledge to blind-sighted patients.[1]

Once Searle's point is set out in this way, it becomes clear that it is susceptible to a straightforward objection. For we can imagine a species of More Conscious People who bear the same relation to us that we bear to the Less Conscious People. The point calls for a concrete example.

Consider the dialect of English in which there is a difference between the pronunciation of the "ng" in "finger" and in "singer." In the dialect I have in mind, the "ng" in "finger" might be said to be hard, whereas the "ng" in "singer" is soft. (Actually, the "g" is deleted in "singer," and the "n" is velar.) In this dialect there is a rule that at least *describes* the phenomenon. For our purposes, we can take it as: Pronounce the "ng" as soft in "nger" words derived from "ng" verbs – otherwise hard (see Chomsky & Halle 1968, pp. 85–87; see Halle 1990 for other examples). One bit of weak evidence in favor of the hypothesis that such an internally represented rule (note, incidentally, that like many other cognitivists, I say "internally," not "mentally") actually *governs* our behavior is that this hypothesis predicts certain facts about the pronunciation of new words. If you tell a member of the dialect group in question that "to bling" is to look under tables, asking what you call one who blings, they say "blinger" with a soft "g." This result rules out the hypothesis that "nger" word pronunciation is simply a matter of a memorized list, and it also rules out certain alternative hypotheses of rules governing behavior. Nonetheless, I concede that there is no strong evidence for the hypothesis that an internal representation of the mentioned rule governs our behavior. But we need not tarry over this matter, since Searle's quarrel is with the very idea of the "deep unconscious," not with the strength of the empirical evidence.

To return to the point, we can now imagine a species of people, the aforementioned More Conscious People, some of whom speak the dialect of English just mentioned, and are also conscious of using the "nger" rule mentioned. Let us further

suppose that in the More Conscious People, the use of this rule is coextensive with a certain brain state, call it the applying-the-"nger"-rule brain state. To complete the analogy, let us now suppose that you and I also have the applying-the-"nger"-rule brain state just in case we belong to the dialect group that makes the mentioned distinction between "singer" and "finger." Here is the punch line: The very reason that Searle gives for postulating blockage in the Less Conscious People applies to us. We have a blockage that keeps us from becoming conscious of our application of the "nger" rule. Since a similar story can be told for any "deep unconscious" phenomenon, this point can be used to legitimize any of the cognitivist's favorite rules, representations, states, or processes. Thus Searle's clarification of his notion of in principle accessibility undermines his overall claim against the deep unconscious.

What does Searle mean by "Conscious?" One way of stating Searle's argument for the Connection Principle is this: Mentality requires aspectual shape, but there is no matter of fact about aspectual shape without (potential) consciousness; hence mentality requires (potential) consciousness. The main line of cognitivist reply should be to challenge the second premise, arguing for a different theory of aspectual shape, namely, a language of thought theory. This line of reply will be no surprise to Searle. I prefer to follow a less traveled path, taking seriously Searle's point that consciousness is a neglected topic.

The word "conscious" is notoriously ambiguous. In one sense of the term, for a state to be conscious there must be something "it is like" to have it. This is the sense that figures in the famous inverted spectrum hypothesis: Perhaps, what it is like for me to look at things that we agree are red is the same as what it is like for you to look at things that we agree are green.

There are many other senses of "consciousness, "including Jaynes's (1977) "internal soliloquy" sense in which consciousness was actually discovered by the ancient Greeks.[2] There is one sense of 'consciousness' that is particularly relevant for our concerns, one in which a state is conscious to the extent that it is accessible to reasoning and reporting processes. In connection with other states, it finds expression in speech. Something like this sense is the one that is most often meant when cognitive science tries to deal in a substantive way with consciousness, and it is for this reason that consciousness is often thought of in cognitive science as a species of *attention* (see Posner 1978, Chapter 6, for example).

Now there is some reason to take Searle's notion of consciousness to be the last sense, the accessibility sense. It is only in this sense that the phenomena postulated by Freud and by cognitive scientists are *clearly and obviously* unconscious. However, in this last sense of "conscious," Searle's Connection Principle is implausible and the argument for it is question-begging. (I am assuming here that Searle will tighten his notion of "in principle accessibility" to avoid the conclusion of the last section that all of the "deep unconscious" is in principle accessible.) If consciousness is simply a matter of access to reasoning and reporting processes, then two states could be exactly alike in all intrinsic properties, yet differ in that one is situated so that reasoning and reporting processes can get at it, whereas the other is not. Yet the state that is badly situated with respect to reasoning and reporting processes might be well situated with respect to other modules of the mind, and may thus have an important effect on what we think and do. Searle would have to say that the first (well situated) state is mental, whereas the second is not. But who would care about such a conception of mentality? This is why I say that according to the present sense of "conscious," the Connection Principle is implausible.

Recall that Searle's argument for the Connection Principle involves the premise that there is no matter of fact about aspectual shape without consciousness. But what reason would there be to believe this premise if all that consciousness comes to is a relation to reasoning and reporting processes? Two states might have exactly the same aspectual shape despite the fact that one can be detected by certain mechanisms and the other can be detected only by different mechanisms.

Further, for the access sense of "conscious," the metaphor of the fish that Searle rejects is quite appropriate. Just as the same type of fish can be below or above the water, the same type of mental state (with the same aspectual shape) can be either accessible or inaccessible to mechanisms of reasoning and reporting. If Searle's argument is to get off the ground, he must take consciousness to be an intrinsic property of a conscious state, not a relational property.

It is time to move to the obvious candidate, the "what it is like" sense. Understanding Searle this way, we enter deep and muddy waters where Searle cannot so easily be refuted, but where he cannot so easily make his case either. I think we can see that his argument depends on a point of view that cognitive scientists should not accept, however. An immediate problem for Searle with this sense of consciousness is this: How does Searle *know* that there is nothing it is like to have the rules and representations he objects to? That is, how does he know that what he calls the "deep unconscious" *really is* unconscious in the what-it-is-like sense? Indeed, how does he know that there is nothing it is like to have Freudian unconscious desires? (Recall that in the present sense of "unconscious" there is nothing it is like to be in any unconscious state.) Our reasoning and reporting mechanisms do not have direct information about these states, so how are we to know whether there is anything it is like to have them? Suppose you drive to your office, finding when you arrive that you have been on "automatic pilot," and recall nothing of the trip. Perhaps your decisions en route were not available to reasoning and reporting processes, but that does not show that there was nothing it was like to, say, see a red light and decide to stop. Or consider a "deep unconscious" case. If subjects wear headphones in which different programs are played to different ears, they can obey instructions to attend to one of the programs. When so doing, they can report accurately on the content of the attended program, but can report only superficial features of the unattended program, for example, whether it was a male or a female voice. Nonetheless, information on the unattended program has the effect of favoring one of two possible readings of ambiguous sentences presented in the attended program. (See Lackner & Garrett 1973.) Does Searle know for sure that there is nothing it is like to understand the contents of the unattended program? Let us be clear about who has the burden of proof. Anyone who wants to reorient cognitive science on the ground that the rules, representations, states, and processes of which it speaks are things that there is nothing it is like to have must show this.

The underlying issue here depends on a deep division between Searle and the viewpoint of most of cognitive science. Cognitive science tends to regard the mind as a collection of semiautonomous agencies – modules – whose processes are often "informationally encapsulated," and thus inaccessible to other modules (see Chomsky 1986; Fodor 1983; Gazzaniga 1985; and Pylyshyn 1984) Though as Searle says, cognitivists rarely talk about consciousness (and to be sure, many cognitivists – Dennett, Harman, and Rey, for example – explicitly reject the what-it-is-like sense), the cognitivist point of view is one according to which it is perfectly possible that that there could be something it is like for one module to be in a certain state, yet that this should be unknown to other modules, including those that control reasoning and reporting. Searle will no doubt disagree with this picture, but his conclusion nonetheless depends on a view of the organization of the mind that is itself at issue between him and cognitivists.

Suppose Searle manages to refute the point of the last paragraph by showing that there is nothing it is like to be in states that are unavailable to reasoning and reporting. That is, suppose that the access sense and the what-it-is-like sense of 'conscious'

apply to exactly the same things. Still, the issue arises as to which is *primary*. Searle will no doubt say that our states are accessible to reasoning and reporting mechanisms precisely when and because there is something it is like to have them. But how could he know that this is the way things are, rather than the reverse, that is, there being something that it is like to have a state is a byproduct of accessibility of the state to reasoning and reporting mechanisms. If the latter is true, once again the metaphor of the fish would be right. For an unconscious thought would have its aspectual shape, whether or not reasoning and reporting processes can detect it; it would be only when and because they detect it that there would be anything it is like to have the thought.

The upshot is this: If Searle is using the access sense of "consciousness," his argument doesn't get to first base. If, as is more likely, he intends the what-it-is-like sense, his argument depends on assumptions about issues that the cognitivist is bound to regard as deeply unsettled empirical questions.

NOTES

1. It is worth mentioning that it is easy to arrange experimental situations in which normal people act like blind-sighted patients in that they give behavioral indications of possessing information that they *say* they do not have. See the discussion below of the Lackner & Garrett (1973) experiment. Thus in one respect we are not very different from the Less Conscious People.

2. See Block, 1981, for a critique of Jaynes (1977), and Dennett, 1986, for a ringing defense of the importance of Jaynes's notion of consciousness.

Intention itself will disappear when its mechanisms are known

Bruce Bridgeman

Professor of Psychology, Clark Kerr Hall, University of California, Santa Cruz, Ca 95064

Electronic mail: *psy160@c.ucsc.edu*

The problem of mentalistic explanation is both more and less than meets the eye. It is less because some of Searle's as-if examples were never meant to be taken literally. The problem is more serious than Searle implies, however, because intentional language for brain processes is always metaphorical; intention is a result, not a process or state.

Searle gives a particularly clear explanation of the contrast between intentionalistic and mechanical/functional explanations with his examples of anthropomorphosed plants and the successful mechanical/functional explanation of the VOR (vestibular ocular reflex). But he exaggerates in assigning intentionality wherever cognitive scientists use intentionalistic language. Searle admits that "We often make . . . metaphorical attributions of as-if intentionality to artifacts," but he maintains that the as-if character is lost when descriptions of brain processes are concerned. Not so – in fact, the contrast between functional and as-if explanation is quite explicit in the neurosciences, and is taught to students of physiological psychology.

A recent textbook (with which I am particularly familiar) makes this clear in words that almost echo Searle's:

> Biologists often speak, in a kind of verbal shorthand, as though useful traits were evolved purposefully, using statements such as: Fish evolved complex motor systems to coordinate their quick swimming movements." This is the intentionalistic statement. The textbook goes on: "What they really mean, though, is that the fish that by chance happened to have a few more neurons in the motor parts of their brains (Searle's mechanical hardware explanation) . . . survived in greater numbers and had more offspring than those that happened to have fewer neurons or less effectively organized ones (Searle's functional explanation). . . . The shorthand of purposeful language will sometimes be used in this book, though the more biologically valid interpretation should always stand behind it. (Bridgeman 1988, p. 10, parenthetical comments added)

Thus purposeful language in neuroscience explanations should always be taken metaphorically, a colorful and compact means of exposition that can always be unpacked to the two-step mechanical/functional argument by the informed student.

In this context, Rock (1984) receives somewhat of a "bum rap" from Searle. The processes of perception work as if they were intelligent, and Rock makes clear even in the quoted passage that the intelligence describes processes in brains, not conscious insights. The perceptual parts of the brain merely process information in ways that in other contexts are interpreted as intelligent.

The second part of my argument, that the intentionalistic explanation is more of a problem than Searle makes it out to be, comes from a closer look at the processes that Searle is still willing to describe in intentionalistic terms, that is, those explicitly identified as conscious.

Perhaps the primary problem with the concept of unconscious mental state is not the "unconscious" or the "mental," but the assumption of a static "state." The state, of course not Searle's invention, is a problem because it derives ultimately from introspection and nothing else. At the start of his target article, Searle questions the relatively small role of consciousness in cognitive science, noting that mention of the term was completely suppressed in respectable circles until recently. This was not always so, however – psychology as a separate discipline was in fact founded on the basis of introspection, or careful examination of the contents of consciousness. Implicit in this effort was the assumption that mental life was indeed accessible to consciousness. The assumption turned out to be false. Freud formalized the insight that some aspects of brain function were unconscious, and neurophysiology has revealed more and more nonconscious processing in the brain. Everywhere we look, nonconscious processes such as early vision, parsing of sensory tasks to different cortical areas even within a modality, or coding of different sorts of memory dominate the brain.

Psychology has been understandably wary of returning to consciousness, and has done so only with an array of new techniques. But it is already clear that the role of consciousness in mental life is very small, almost frighteningly so. The aspects of mental life that require consciousness have turned out to be a relatively minor fraction of the business of the brain, and we must consider consciousness to be a brain system like any other, with particular functions and properties. It looms large only in our introspections.

More specifically, whenever we examine an aspect of what seems to be conscious it turns out to be made of simpler parts. The process of seeing, for instance, is made of a great cascade of neural processing based on a welter of relatively simple algorithms. What had seemed like visual intelligence turns out to be only processing after all, when we look empirically at how it works. (A similar fate has befallen seemingly intelligent AI efforts.) Wherever we look for intentionality, we find only neurons, as Searle laments. We can predict that intentionality will evaporate in the twenty-first century, as certainty did in the twentieth.

My final comment is on the problem of aspectual shape. The resolution, seen from psychology, is that just the fact that a conscious manifestation has aspectual shape is no reason that the memories on which it is based should have any such property. Why not construct the aspectual shape during the process of bringing memory from storage? With apologies to Searle, the computer analogy applies here, in a kind of lilies-of-the-field argument – if the humble computer has a given capacity or property, why not us? The information stored on my computer's disc has no margins, lines or paragraphs; it looks nothing like the form it will have when it is displayed on my terminal. When it is needed, the bare-bones disc information is recoded, formatted,

Young in various ways question my notion of consciousness; and several others, specifically **Chomsky, Limber, Piattelli-Palmarini** and **Rey,** think I am relying in some way on the notion of introspection.

By consciousness I simply mean those subjective states of awareness or sentience that begin when one wakes in the morning and continue throughout the period that one is awake until one falls into a dreamless sleep, into a coma, or dies, or is otherwise, as they say, unconscious. On my account, dreams are a form of consciousness (this answers the queries of **Young** and **Hodgkin & Houston**), though they are of less intensity than full blown waking alertness. Consciousness is an on/off switch: You are either conscious or not. Though once conscious, the system functions like a rheostat, and there can be an indefinite range of different degrees of consciousness, ranging from the drowsiness just before one falls asleep to the full blown complete alertness of the obsessive. There are lots of different degrees of consciousness, but door knobs, bits of chalk, and shingles are not conscious at all. (And you will have made a very deep mistake if you think it is an interesting question to ask at this point, "How do you know that door knobs, etc., are not conscious?") These points, it seems to me, are misunderstood by **Block.** He refers to what he calls an "access sense of consciousness." On my account there is no such sense. I believe that he, as well as **Uleman & Uleman,** confuse what I would call peripheral consciousness or inattentiveness with total unconsciousness. It is true, for example, that when I am driving my car "on automatic pilot" I am not paying much attention to the details of the road and the traffic. But it is simply not true that I am totally unconscious of these phenomena. If I were, there would be a car crash. We need therefore to make a distinction between the center of my attention, the focus of my consciousness on the one hand, and the periphery, on the other. William James and others often use the notion of consciousness to mean what I am referring to as the center of conscious attention. This usage is different from mine. There are lots of phenomena right now of which I am peripherally conscious, for example, the feel of the shirt on my neck, the touch of the computer keys at my finger tips, and so on. But as I use the notion, none of these is unconscious in the sense in which the secretion of enzymes in my stomach is unconscious.

A remarkably large number of commentators think that introspection plays some role in my account. They are mistaken. I make no use of the notion of introspection at all. **Chomsky** in particular thinks that I assign some special epistemic priority to introspection, that I am committed to the view that we have some special knowledge of our own mental states by what Chomsky calls an "inner eye." That is not my view at all. Except when quoting Chomsky, I was very careful never to use the word "introspection" in the course of my article, because, strictly speaking, I do not believe there is any such thing. The idea that we know our conscious mental states by introspection implies that we spect intro, that is, that we know them by some inner perception. But the model of perception requires a distinction between the act of perceiving and the object perceived, and that distinction cannot in general be made for our own conscious states. This point was already implicit in our discussion of Objection 3.

I assign no epistemic privilege to our knowledge of our own conscious states. By and large, I think our knowledge of our own conscious states is rather imperfect, as the much cited work of Nisbett and Wilson (1977) shows. My points about consciousness, to repeat, had little or nothing to do with epistemology. They were about the ontology of consciousness and its relation to the ontology of the mental.

Chomsky, by the way, is mistaken in thinking that the discussion in this article is the same as our dispute of several years ago. That issue was genuinely epistemic. This one is not.

IV. Program explanations: Reply to Objection 5

One of the most fascinating things in this discussion is the extent of the disagreement among the objectors about the nature of cognitive science explanations. Most of the commentators accept my claim that cognitive science typically postulates deep unconscious rules and that *as if* intentionality explains nothing. But several have a different conception of cognitive science explanation. They see the computational paradigm not as an implementation of intrinsic intentionality but as an alternative to it or even a rejection of it. **Matthews, Higginbotham, Glymour** and – to some extent – the **EDITORIAL COMMENTARY** take it that computational forms of explanation might be a substitute for intentionalistic explanations, and **Hobbs** and **McDermott** think we have already superseded intentionalistic explanations, that the ascriptions of intentionality in cognitive science are entirely *as if,* but that this does not matter because the causal account is given by the program explanation. We substitute an algorithmic explanation in terms of formal symbol manipulation for the intentionalistic explanation.

The logical situation we are in is this: We are assuming that the Chinese Room Argument shows that the program level is not sufficient by itself for intentionality and that the Connection Principle argument shows that there is no deep unconscious intentionality. Well, all the same, the question remains open whether or not the brain processes might still be in part computational.

And the underlying intuition is this: As a matter of plain fact, there are a lot of physical systems out there in the world that are digital computers. Now maybe, as a matter of plain fact, each brain is one of those. And if so, we could causally explain the behavior of the brain by specifying its programs the same way we can causally explain the behavior of this machine by specifying its programs.

I can, for example, causally explain the behavior of the machine on which I am writing this by saying that it is running the vi program and not the emacs program. **Hobbs** and **McDermott,** more strongly than the others, concede my points about the nonexistence of the deep unconscious and think that cognitive scientists in general should also, but they think the Darwinian Inversion could be avoided because they think we might just discover that the brain is a biological computer.

Notice that this point is logically independent of the main argument in the article. I could, in principle, just concede it as irrelevant to the present issue but I want to discuss it, at least briefly, because I think the hypothesis of computationalism as a causal explanatory model of cognition has some real difficulties.