Matter, Levels, and Consciousness

Jerry R. Hobbs

SRI International

Menlo Park, California

Searle's ontology is at once richer and more barren than that of most cognitive scientists. He says repeatedly that matter exists (in particular, neurons), and conscious, intentional experience exists, and that is all. If we accept this, it follows trivially that there are no deeply unconscious rules operating at a symbolic level intermediate between neurophysiology and intelligent activity. There are no such rules because they are neither neurons nor potentially conscious intentional experience. The six carefully argued steps in his demonstration could have been dispensed with entirely. Throughout his long argumentation, however, we find no justification for an ontology so barren.

For example, in Step 4 (p. 5) he says, "Well, the only facts that could exist while he is completely unconscious are neurophysiological facts. The only things going on in his unconscious brain are sequences of neuro-physiological events occurring in neuronal architectures. At the time when the states are totally unconscious there is simply nothing there except neurophysiological states and processes." This is repetition, not argument.

Levels. Such an ontology runs counter to the spirit of the entire scientific enterprise, for Searle apparently would not admit such complex, large-scale, diffuse organizations of matter as geological faults, immune systems, data structures, and nations. There is nothing but matter and conscious experience. To a scientist's arguments that such entities are useful descriptively and that their reality is at least plausible, he would apparently reply in triumph that these intermediate levels can be dispensed with, and for the sake of parsimony, should be.

Science is organized by levels, a strategy that is successful probably because nature is organized by levels. There are at least two ways we can view these levels. First, we can view them as *levels of description*. Nature cannot usefully be described solely in terms of the motions of elementary particles. We have found it convenient to define or hypothesize larger-scale entities and to couch our theories in terms of them. We then try to account for the behavior of these entities in terms of the entities provided by the theory of the phenomena one or two levels down. Thus, chemists seek to understand in quantum theoretic terms why molecules react as they do.

Second, we can view the levels as *levels of organization*. That is, they are not merely convenient fictions that allow our poor, finite minds to understand what is going on. There is something in nature that actually corresponds to these large-scale entities and actually behaves approximately in the manner that our theories describe. The argument for assuming these things are really out there in the world is what has often been said: We should adopt the ontology implied by our most successful theories. The reality of the ontology is the best explanation for the success of the theory. Molecules, cells, tissues and organs, organisms, herds, and nations are not merely stories we tell. They really do exist.

Levels of description are not necessarily levels of organization. Explanations based on a metaphor of intentional action may once have been useful in plant physiology; they may have corresponded to a level of description. But this did not come to be viewed as a level of organization. The biochemical mechanisms involved turned out to be simple enough that this intermediate level was not required.

Searle seems to lack this notion of entities at different levels. His unwillingness to accept the notion of two different entities operating at two different levels but realized in exactly the same portion of matter is, in fact, the reason for his failure to accept the "Systems Reply" to the Chinese Room Argument (the debate over which rages on largely because at least one of its participants rages on—what else can one say to someone who can write a sentence like "Strong AI is now primarily of historical interest, though, of course, it survives as a sociological phenomenon.").

What I take to be the standard view, or hope, in cognitive science today is the following: intelligent activity is implemented in a symbolic level, which in turn is implemented in a connectionist architecture, which in turn is implemented in neurophysiology. Variations on this view dispense with the symbolic or with the connectionist level. (It may have been a personal discovery for Searle that the symbolic level could be eliminated in favor of a more powerful connectionist level, but it is a rather commonplace view among workers in the field.) The reasons for this strategy are clear. We can observe intelligent activity and we can observe the firing of neurons, but there is no obvious way of linking these two together. So we decompose the problem into three smaller problems. We can formulate theories at the symbolic level that can, at least in a small way so far, explain some aspects of intelligent behavior; here we work from intelligent activity down. We can formulate theories at the connectionist level in terms of elements that behave very much like what we know of the neuron's behavior; here we work from the neuron up. Finally, efforts are being made, with modest success, to implement the key elements of symbolic processing in connectionist architecture. If each of these three efforts were to succeed, we would have the whole picture.

What success has been experienced at the symbolic level in modeling intelligent activity has been due to hypothesizing operations that are very close to conscious intentional mental operations—idealized, cleaned up versions of them. In the most successful symbolic theories, the processor at the symbolic level does such things as manipulate symbols, follow rules, draw inferences, and plan. (And contrary to what Searle says, this does not require a homunculus. There are no little men in our machines.) Symbolic operations are for the most part based on metaphors drawn from folk psychology, and their success is largely parasitic on the success of folk psychological explanations of intentional behavior. This "anthropomorphizing" is not a "mistake" Searle has discovered. It is a quite deliberate strategy.

I find incomprehensible Searle's statement that these symbolic operations have no causal powers. When they are written down on paper, of course they have no causal powers. Just like real intentions, they have causal powers if and only if they are implemented in neurophysiology or in electronics or in some other adequate material medium. Moreover, most computer programs implementing this strategy do not merely *behave* as though they were following rules. The rules really are there in the data structures, and the following really is there in the procedures that manipulate the data structures.

The symbolic level is at least a level of description. Whether it will turn out to be a level of organization we simply don't know today, although many cognitive scientists believe it successful enough already to adopt the ontology it implies, and say that symbolic operations really do exist. It may be, as Searle suggests and many connectionists believe, that the symbolic level will wither away as better direct accounts relating neurophysiology and intentional behavior become available. Or it may be that a symbolic level will come to be viewed as useful a concept and as real in psychology as tissues and organs are in physiology.

There *is* a notion of levels that Searle makes use of, but this is a quite different notion and should probably be thought of not as levels but as perspectives one can take at any level. There are exactly two of these, and they may be called the *structural* ("hardware") perspective and the *functional* perspective. From a structural perspective, we attempt to decompose the entity into its constituent parts and tell a causal story about how the properties of the entity emerge from the properties of its parts. From a functional perspective, we consider the entity as a whole, undecomposed, and ask how it influences and is influenced by its environment. It is not an intrinsic feature of the functional perspective, as Searle says, that it be related to human interests, although, unsurprisingly, most functional accounts that we have bothered to construct are.

Consciousness. I said that Searle's ontology is not only more barren but also richer than that of most cognitive scientists. He believes not merely in the existence of conscious, intentional experience, but in its special, fundamental, explanatory power. It seems to play a role equal to that of matter: matter can cause, and conscious intentions can cause. Moreover, this is a position that he attributes to cognitive science, in what can only be characterized as a fundamental misunderstanding of the first order.

The assumption is so deeply ingrained in what he has written that it is not easy to spot. But near the end of Section V (p. 15), he says, "*as-if* intentional states, not being real, have no causal powers whatever. *They explain nothing*," and "*as-if* intentionality ... simply restates the problem *which the attribution of real intentionality is supposed to solve*" (italics mine). It looks very much as if Searle believes that the postulation of deeply unconscious intention-like operations at the symbolic level is an attempt to appropriate the causal, explanatory power that he supposes conscious intentionality to have. **But this is no part of the strategy of cognitive science.**

It really is true of Freudian psychology that the hypothesis of subconscious intentions is meant to borrow the causal, explanatory power of conscious intentions. In folk psychology, conscious intentions have explanatory power; explanation can often stop there. In the social sciences, where we often take folk psychology as a background science and hence unproblematic, conscious intentions can have explanatory power; explanation can again stop there. However, cognitive science, in its use of deeply unconscious intentions, is not seeking "intentionalistic *explanations*" in the sense of Section III (p. 7). In cognitive science, intentions do not have special explanatory power. Explanation cannot stop there, because intentional behavior is precisely what is to be explained. Conscious intentions have no causal powers except insofar as they are viewed as complex processes implemented, through who knows how many levels, in neurophysiology. They cause only because matter causes and they are implemented in matter.

Searle seems to believe that intentionalistic explanations are valid in psychology when applied to potentially conscious intentions, and accuses cognitive scientists of illegitimately extending them to "deeply unconscious intentions". But in fact cognitive scientists have rejected intentionalistic explanations altogether, and use the hypothesis of deeply unconscious intentions as a way station to a structural, computational explanation. If deeply unconscious rules and representations could not be realized computationally, and hence materially, they would be of no value.

Searle closes his article with the statement, "We know for sure that there is a brain and that at least sometimes it is conscious." Could it be that he believes, since each of us knows for sure that his or her brain is sometimes conscious, that that fact requires no further—structural—explanation?

Consciousness, in any case, seems a very shaky foundation to try to build a science on. I know I'm conscious. I can't be sure about other people, but it's a courtesy I'm willing to extend. I would extend it to orangutans as well, but probably not to birds or to any computer program I've ever encountered, and certainly not to amoebas.

Although the intelligent activity that cognitive science seeks to model is normally exhibited by conscious people, there is a sense in which Searle is quite right in saying that the field has largely "neglected" consciousness. There are some aspects of consciousness that can be modeled in symbolic terms, such as focus of attention and knowledge of one's own beliefs. But how the subjective experience of consciousness could emerge from complex arrangements of neurons is a mystery apparently inaccessible to present-day cognitive science. The neglect is not because it is in bad taste to study it, but because there are no very good ideas about it.

I can imagine Searle replying to all this with, well then, the symbolic operations are just *as-if* intentionality, and not intrinsic. I believe it is possible to justify attributing intrinsic intentionality to them on the basis of their place in a large system of similar rules, some of which are deeply unconscious and some of which are quite conscious and frequently verbalized. However, in cognitive science, since intrinsic intentionality confers no extra explanatory power, the issue is not very important.