Contents Volume 13:4 December 1990

subliminal and neurophysiological methods. These unconscious fantasies appear to have aspectual shape.

The neurophysiology of consciousness and the unconscious

Christine A. Skarda

Molecular & Cell Biology, Life Sciences Addition, University of California, Berkeley, CA 94720

Electronic mail: *wfreeman@garnet.berkeley.edu*

Searle adopts the physiological point of view that there is no basis for the claim that behavior is driven by programs like those used by digital computers. But how well does he characterize the neural mechanisms of such conscious states as perceptual recognition, and such unconscious ones as memories?

Laboratory data (Freeman & Schneider 1982; Freeman & Skarda 1985; Skarda & Freeman 1987) suggest that each conscious mental state begins *within* the brain as an internally generated, self-organized process that is projected through the brain both as a motor command that orients exteroceptors optimally in the light of past sensations, and as reafferent messages to all sensory systems that prepare them for the consequences of expected action on the basis of past experience. Conscious mental states also require chemically mediated synaptic changes during learning that lead to the formation of nerve cell assemblies (NCAs) that sensitize sensory cortices to particular stimulus configurations, but these states should not be identified with NCAs (Skarda & Freeman 1988). In the olfactory system, a perceptual process is initiated by an inhalation that causes a volley of excitatory receptor input. When a critical threshold of excitation is reached in the bulb, a state change occurs in which the entire bulb abruptly changes to a globally distributed, stereotypical pattern of activity. The role of the NCA is to mediate the selection of this pattern, but it is the globally distributed activity pattern that constitutes perceptual recognition and that is transmitted to the limbic system, comprising neocortical and subcortical structures of the forebrain in an interactive hierarchy. Its self-organized, global activity patterns are the best candidates we have to identify with consciousness. To summarize: Conscious mental states are inaugurated from within, involve globally distributed dynamics, and are self-organized, hierarchical, interactive states of dynamic patterned activity.

Contrast this with such unconscious, nonmental states as reflex behaviors (Sherrington 1906). A reflex is stimulus dependent; it is initiated from *without* by the stimulus rather than by the brain. It involves a series of passive, feedforward transformations performed on the input pattern by effecting each link in the neronal chain in turn like a string of dominoes (Thach 1978). Most reflexes do not involve the cerebral cortex; those that do involve limited portions of it, not the whole. This neurophysiological difference is to be expected: For a state to be conscious and mental it has to have the right sort of neurophysiology, something reflexes and other kinds of nonmental, unconscious phenomena do not have.

Searle's target article makes a distinction between a neurophysiology that has the "capacity" for consciousness and that which does not, and in this respect is consistent with what we know about brain functioning. But what of his claim that unconscious, mental processes are "going on" in my brain and that they are mental precisely "because they are capable of causing conscious states" (step 6, last paragraph)? Is Searle's characterization of unconscious mental phenomena consistent with what we know from neurophysiological research? I think not.

What leads Searle astray in his attempt to characterize unconscious mental phenomena is his causal account of the mind/brain. Neurophysiological processes do not *cause* mental states, they *are* mental states at the neurophysiological level of description. Phenomena described at different levels of description are not causally related (Rose 1987). To assume otherwise is both bad science and bad philosophy. Searle is aware of this even if he waffles on the issue, but his causal account still gets him into trouble.

Causal explanations are inappropriate not only when applied between explanatory levels, but also for phenomena *within* a level of description. Nonlinear dynamical systems theory has shown that brain dynamics preclude causal explanations for two reasons. Recent connectionist models of memory have focused on the mechanisms of neural network formation (Hinton 1985; Hopfield & Tank 1986; Kohonen 1984), with the result that many today believe that memory consists entirely of synaptic changes within an NCA. Searle may have had this in mind when formulating his causal account of unconscious mental states. Yet while activity in the NCA temporally precedes the globally distributed patterned activity we identify with conscious, mental phenomena, these microphenomena cannot explain – are not the "causes" of – the globally distributed activity patterns of the conscious state that follows (Skarda 1986). Brain dynamics are self-organized. Self-organized phenomena cannot be explained in reductionistic terms, as Searle claims (sect. 6, para. 3). Explanations of such phenomena can be given only in terms of the qualitative forms of behavior of the system as a whole, and not in terms of properties of its parts, whether these be neural networks or individual neurons. Second, the observation that neural dynamics are "chaotic" (Skarda & Freeman 1987) further undermines Searle's causal account of the unconscious (Skarda & Freeman 1988). Chaotic phenomena are inherently unpredictable because small uncertainties are amplified over time by the nonlinear interaction of a few elements (Crutchfield et al. 1987). The upshot for neurophysiology is that we cannot make strict causal inferences from the level of neural networks to that of neural mass actions. The impact of this explanatory revolution in neuroscience does not seem to have reached Searle. A causal account of conscious or unconscious phenomena is doomed to failure.

What sense can we give to the notion of unconscious mental phenomena if we reject the causal account, then? How do unconscious memories differ from reflexes? The globally distributed, self-organized activity patterns required for conscious recall do not persist in the brain when the state is unconscious. Memories arise new each time by self-organized processes and are not retrieved from a "memory store" as in a digital or analog computer memory. What remains when memories are unconscious is a "space of possibilities"; in mathematical terms, a structured, interdependent global system of "attractors." What we refer to as unconscious memories are "tendencies" or predispositions to engage in particular forms of patterned activity that are made available to the system at the point when it is destabilized by its interaction with the environment (Skarda & Freeman 1988). Unconscious mental phenomena are not "going on" in the brain and they do not "cause" consciousness, but they are available forms of dynamic activity that define what the system can do and that are constrained by each new conscious experience that further shapes this space of possibilities. And because these forms of patterned activity are merely available to the system whenever it is destabilized, they preserve our notion of the unconscious in that they may never be actualized.

The key point is this: The global structuring process in which past experience changes current neural dynamics does not occur with such unconscious, nonmental phenomena as reflex behaviors. The "structured space of possibilities" is how unconscious mental phenomena operate in brain dynamics. This process is unique to unconscious mental phenomena, distinguishing the neural dynamics of the unconscious. So there is neurophysiological reality to unconscious mental phenomena that

sets them apart from unconscious nonmental ones, but it is not the one Searle suggests. The unconscious cannot be understood in terms of causes; it must be understood in terms of future possibilities for the system to engage in self-organized patterned activity that is created anew each time. If Searle wants an account that is physiologically sound, he would do well to reject not only the digital computer paradigm, but also the notion of causality that philosophers still too often mistake as the hallmark of "scientific" explanation.

The possibility of irreducible intentionality

Charles Taylor

Department of Political Science, McGill University, Montreal, Quebec H3A 1G5 Canada

I agreed with much in John Searle's target article, in particular, with his insistence on the distinction between intrinsic and as-if intentionality. I think I can accept the Connection Principle under some formulation. But I'm still worried about whether there is not some a priori neurophysiology lurking in Searle's account.

I can get at this worry either by going at the issue of what it is that, inaccessible to consciousness, underlies our conscious thoughts and actions; or by questioning Searle's invocation of Darwinian explanation. Let me try both of these routes, in the order introduced.

I'll take an example that I don't really accept, but that could be true in some form: a Durkheimian explanation of religion. The sense of religious awe and allegiance "really" reflects the sense of dependence on society, not just for survival but for our being fully human agents operating on a moral level. Now Durkheim may have thought that we moderns come to recognize this consciously. But according to one interpretation, even the enlightened, lay citizens of the Third Republic didn't have first-person insight into what made them such fervent supporters of the republican tradition. Then why believe the theory? Because it *could* make sense of the whole history of religious development, including the rise of lay ideologies, better than any rival view.

In other words, the theory points to some factor, here an attachment to something, which is posited as what is really moving people, even though these people can't call it up to first-person consciousness. What I mean by this last phrase is that they couldn't see the desirability of the objects of their religious or moral fervor under the aspect, "the social bond that makes us human"; although clearly they could come to accept the *theory* as a third-person account of their and others' behavior. It is clear that a Freudian-type theory could also be constructed with this feature, as could any theory supposing an "unthought."

Now this factor has a very definite aspectual character, in Searle's sense. The theory purports to identify a desirability-characterization. We don't have the kinds of cases Searle considers, where some being seeks water, and we rightly argue that we have no way of determining whether he seeks it under the aspect "H_2O." Here we have triangulated, as it were, to a determinate aspect from out of conscious, full-blown, intentional behavior.

If Searle could be got to accept this social bond factor as a genuine example of the in-principle unconscious, then his answer would be to ask what facts correspond to the claim that it explains religion. Since by hypothesis we don't have a conscious thought, we must have a brain state. And so why can't we settle for some brain condition which always produces social-bond-cathecting behavior, along with conscious thoughts identifying a certain range of things (including God, Republican principles, but not the social bond) as objects of devotion? This would be the underlying mechanism, analogous to the VOR in the vision case, and we could dispense with unconscious desire.

Maybe we should settle for this, thus espousing the Connection Principle. But this brings us to the second line of approach, Searle's appropriation of the Darwinian analogy. What does it mean to make a hard and fast distinction between hardware explanations and functional explanations? Searle wants to argue that the functional consideration tells us nothing about the causation of the phenomenon. The plant turns because of the secretions of auxin, not because the turn maximizes sun exposure. (Of course, over time the survival effects of sun exposure ensure that this kind of plant proliferates, but Searle's point concerns the explanation of the individual plant's behavior.)

But what do we mean by a hardware account? One plausible interpretation is that we mean accounts that have recourse only to factors recognized by the disciplines of neurophysiology or organic chemistry as they now exist. The firings of different neurons will be explained by local chemical changes, and the larger patterns of firing will be accounted for by the concatenation and mutual interactions of such local effects. Or we might admit larger field effects into our causal story, but these fields would be defined anatomically. Imagine, however, that a further step is necessary: that we need to invoke field effects where the "fields" in question cohere just in being embodiments of some state defined in intentional terms – a thought of something, or a desire for something.

Now the latter step would take us well beyond the bounds of mainstream neurophysiology, by incorporating intentionalist concepts into the science. Is the notion of a hardware account meant to exclude this possibility? If so, then it may be excluding a priori a promising line of advance. What is more, if all neurophysiological function is explained in hardware accounts so defined, and if all conscious states are realized in neurophysiological states (and the latter premise is pretty universally shared), then it is hard to see how accounts of conscious action can avoid becoming derivative of these hardware explanations. Epiphenomenalism looms.

If, on the other hand, "hardware accounts" can be enlarged to include any such useful intentionalistically defined factors, then it is hard to see how the rigid separation between hardware and the functional can be maintained. To play my sub-Durkheimian fantasy out to the end, what if the underlying neurophysiological account of our cathexis of the social bond had recourse to field effects of neuronal clusters, where these were identified at least in part inescapably in terms of what they related to – for example, the social bond? Would it be so clear that we shouldn't speak of in-principle unconscious states? Perhaps indeed, it would, and we would be wise to follow Searle's warnings about the confusions that can arise here. But the grounds for this exclusion would no longer be that the intentional, the aspectual, the functional cannot extend beyond what can be called to consciousness. To lay that principle down at this stage is to slide into a priori theorizing about the future development of science.

The causal capacities of linguistic rules

Alice ter Meulen

Department of Philosophy, Indiana University, Bloomington, IN 47405
Electronic mail: *atm@ucs.indiana.edu*

1. Degrees of unconsciousness and the nature of linguistic rules. In conversation we are usually aware of the meaning of the sentences uttered, but not of the syntactic, phonological, phonetic, and semantic principles and rules we use in computing it. We have conscious access only to products of linguistic rules and principles, not to the rules themselves nor to the computational