<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>A more substantive neuron doctrine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Lau, JYF</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Behavioral And Brain Sciences, 1999, v. 22 n. 5, p. 843-844</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>1999</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/43153">http://hdl.handle.net/10722/43153</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Behavioral and Brain Sciences. Copyright © Cambridge University Press.</td>
</tr>
</tbody>
</table>
A more substantive neuron doctrine

Joe Y. F. Lau
Department of Philosophy, The University of Hong Kong, Hong Kong.  
jyllau@hkusua.hku.hk  www.hku.hk/philodep/joelau

Abstract: First, it is not clear from Gold & Stoljar’s definition of biological neuroscience whether it includes computational and representational concepts. If so, then their evaluation of Kandel’s theory is problematic. If not, then a more direct refutation of the radical neuron doctrine is available. Second, objections to the psychological sciences might derive not just from the conflation of the radical and the trivial neuron doctrines. There might also be the implicit belief that, for many mental phenomena, adequate theories must invoke neurophysiological concepts and cannot be purely psychological.

In presenting the radical neuron doctrine, Gold & Stoljar (G&S) did not explicitly say whether computational and representational concepts (CRCs, for short) fall within their definition of biological neuroscience; but this is important because these concepts seem to be indispensable in understanding the function of neural mechanisms. Without them, we cannot understand how neurons contribute to information processing in the brain. As a matter of fact, even the Churchlands appeal to notions such as content-addressable memory, distributed representations, parallel processing, and vector transformation in articulating their favorite research program. Such concepts obviously cannot be reduced to neurophysiology, however, as they can also apply to nonbiological systems. Thus, if CRCs are indeed indispensable, and they fall outside biological neuroscience, then this is already sufficient to refute the radical neuron doctrine.

Perhaps G&S meant to include CRCs within biological neuroscience. However, such a move is likely to weaken their argument that Kandel’s theory of learning cannot provide a reduction of the concept of classical conditioning. According to G&S, the current conception of classical conditioning involves the learning of relations among represented events. However, this involves the notion of information about relations that they think cannot be captured in Kandel’s theory. This might be so, but the issue is whether biological neuroscience in principle has the resources to fill the gap. Insofar as CRCs are ideally suited for capturing informational concepts, proponents of the radical doctrine might reply that Kandel’s theory (or an improved version) can provide a reduction of classical conditioning when embedded within a suitable computational framework, and this enriched theory can still be part of biological neuroscience in the broad sense. Whether the radical neuron doctrine is true on this reading would then depend on whether there are psychological concepts that cannot be reduced to CRCs plus other concepts in biological neuroscience. I think that there are indeed many such concepts, but this is not the place to go into the arguments.
A related issue arising from G&S’s discussion concerns the relationship between psychological and neurophysiological theories. G&S seem to think that the latter can at most provide implementational forms of the former, and they illustrate their point using the theory of color opponency and David Marr’s theory of vision. A common feature of both examples is that there is a level of psychological theory that can be specified independently of neural implementation. In the first case it is the theory of the opponent character of color perception; in the second case it is a theory of what the visual system computes and why. Interestingly enough, however, Marr himself cautions that the distinction between computational and implementational theories might not be applicable to all problems of biological information processing. He says that “this can happen when a problem is solved by the simultaneous action of a considerable number of processes, whose interaction is its own simplest description” (Marr 1977, p. 38 [his emphasis]).

If I understand him correctly, I think his point is that in such situations, which he calls “Type II” situations, it might be impossible to find an informative abstract description of what a system does without mentioning the complex mechanisms involved.

The relevance of Marr’s remark is that it raises the following possibility. There might be many mental phenomena for which it is impossible to devise informative and explanatory theories that are purely psychological and that do not make use of neurophysiological concepts. Let the “substantive neuron doctrine” be the claim that this possibility does in fact obtain. Of course, even if this doctrine were true, it would not vindicate the radical neuron doctrine, insofar as the mixed theory can contain irreducible psychological concepts, but this substantive doctrine is not trivial either; it has the methodological consequence that for some mental phenomena it would be misguided to try to develop a purely psychological theory.

The point is not just that one has to keep in mind the issue of neural implementation when devising psychological theories for these phenomena. Rather the claim is that one cannot begin to formulate an adequate theory without explicitly bringing in neural details, “getting one’s hands dirty” as it were. It seems to me that a lot of the rhetoric directed against the psychological sciences might have to do with the implicit acceptance of this substantive doctrine and not just the conflation of the radical and the trivial doctrine. This is one way to interpret what the Churchlands have in mind when they criticize “autonomous psychology” (McCauley 1996, p. 220). They give the example that the structure of the periodic table remains a mystery until quantum mechanics enter into the picture. Likewise, the suggestion might be that many distinctive features of the mind can be explicated only if we bring in neurophysiological findings. Whether this is true is of course an empirical matter. There can be no a priori route to the conclusion that, say, theories of syntactic principles must somehow bring in neurophysiological concepts if they are to be viable. As with the rest of science, the ultimate justification for any particular approach lies in its success, but, whatever the case may be, on this interpretation we need not see those who defend the neuron doctrine as defending a view that either has no defense or that needs none.

The assumption that neurons are the appropriate level of description for cortical information processing and mental phenomena in general (the neuron doctrine) is usually regarded as valid. It is important, however, to question once in a while the very foundation and scope of this doctrine, as Gold & Stoljar (G&S) have done.

The ultimate reductionist approach to cortical information processing would only point to physics, and the ultimate level of description would be that of elementary particles. From this perspective, as G&S remark, neurobiology would be only a “local stop” (sect. 4.2), so the privileged status of neurons in today’s brain science cannot be derived from reductionism itself.

How then is neural firing the appropriate level of description in neuropsychology? From the dynamics point of view, neural activities are special because of the nonlinearity and all-or-none character of action potential generation. No subneural processes are known at present that show the same degree of macroscopic nonlinearity. In addition, in most cases, synaptic interaction is invoked only when a neuron fires. These are the rationales for treating neural firing as the only relevant explicit variable in cortical information processing. All other variables (including those describing the subcellular processes) can be treated as implicit variables, affecting cortical information processing only indirectly; the effect on the eventual neural firing. The reductionist would only have to go as far as neural activity; the rest would be details. Neurobiology might be a “local stop,” but it suffices. Treating neural firing as an explicit variable does not necessarily entail a grandmother cell-type coding and is, in fact, a generic assumption behind any model of neural coding. It is in this modern sense that the neuron doctrine (Barlow 1972) should be interpreted.

The rather simplified but effective treatment of cortical information processing in terms of neural activities given above does leave some very important issues unanswered, as G&S rightly point out. The main difficulties are in the field of “cognitive neuroscience” as opposed to “biological neuroscience” (sect. 2.1). Here, there is indeed an “ambiguity” in what the neuron doctrine means (sect. 1.4). If it is claimed that the neuron doctrine is relevant only for the biological neuroscience, fine; it is plausible but not radical. If it is claimed that the neuron doctrine supersedes the psychological sciences as well, then it is surely radical, but does not necessarily sound plausible. What is the neuron doctrine really supposed to mean in this view?

In my interpretation, the ambiguity could be resolved by considering the “supervenience” of mental events on neural activities. Davidson (1970) introduced the concept of supervenience thus: “Mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.” To paraphrase, we could hypothesize that there cannot be two events alike in all neural activities but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some neural activities. This hypothesis does sound plausible, and in this sense it is plausible that mental events should supervene on neural activities. In other words, it should in principle be possible to explain mental events in terms of neural activities only, with no extraneous elements needed.

Qualia (Chalmers 1996) come into the picture here. Qualia are the hallmark of our mental activities, at least as far as conscious mental activities are concerned. It seems plausible to assume that a certain quale is invoked in our mind when a certain pattern of neural firing occurs in the brain. There is certainly the difficult question of comparing the qualia that two individuals have. We cannot ever be sure whether the qualia of the red that two subjects have are identical, nor whether such a comparison is meaningful at all. However, it does seem to be plausible that once we have a specific neural firing pattern in individual subjects’ brains they will have a certain quale corresponding to that neural activity. In this sense, qualia would supervene on the neural activities.