**The Churchlands' Neuron Doctrine: both cognitive and reductionist**

Gold and Stoljar (G&S) allow 'biological neuroscience' to *include* study of the function as well as the structure of neuronal ensembles (section 2.1). But they think that invocations of function in actual neurobiological explanations already invoke nonbiological concepts, so that explanations of causal mechanisms fulfilling those functions are *not* purely neurobiological. Because even an apparently physiological notion such as the reflex is 'highly theoretical', G&S deny that it is a legitimate construct of physiology alone (note 40). It's as if the fact that neurophysiology, as Berent Enc put it, 'contains as an essential component a certain abstract level of description of the functional organization of the nervous system' (1983, p.298), automatically makes it a non-biological science! Thus, the 'radical neuron doctrine' (RND) as defined is ludicrously strong. G&S's purified definition, excluding all psychological, theoretical, or behavioral terms from neurobiology, allows only theorists who refuse to invoke concepts such as classical conditioning, information, and representation consistently to propose RND. There may be some such theorists among those who deny the utility of current concepts of representation, seeking instead to replace psychology with terms from dynamical systems theory (van Gelder 1995), or even quantum theory (Penrose 1994). G&S could persuasively argue that these attempts to unify cognitive science *directly* with physics, which are compatible with RND, do not have sufficient resources to explain mentality. Surprisingly, though, they are not the targets. Instead, G&S implausibly interpret the Churchlands as supporters of RND. However, neurocomputational models of learning and memory centrally invoke representations (P.S. Churchland and Sejnowski 1992, pp.141-237). They are pitched "at a decidedly abstract level": the two-pronged framework of transient, occurrent representations, and enduring, dispositional (distributed) representations can in principle be realized in many neurobiological systems. (Churchland and Churchland 1996, pp.224-30). Indeed recognizably connectionist frameworks of explicit and implicit memory representation were developed by early modern theorists such as Descartes and Hartley, who relied on quite different neurophysiological realizations, in animal spirits and in vibrations, respectively (Sutton 1998, 1999).

G&S have two responses. Firstly, they complain that the Churchlands do nevertheless, in confusion, often defend RND. A more charitable reading would focus less on hyperbolical rhetoric and more on the Churchlands' detailed proposals for specific neurocomputational explanations, where they rely on thoroughly *cognitive* theories, including the opponent process theory of color perception (P.M. Churchland and P.S. Churchland 1998, pp.168-172; compare the use of psychophysical and clinical data on vision in P.S. Churchland and Ramachandran 1993 and of the cognitive neuropsychology of emotion and decision-making in P.S. Churchland 1996).

More substantively, G&S see the only alternative to RND as the weak trivial neuron doctrine, by which 'cognitive neuroscience' will explain mentality. They see all versions of the neuron doctrine which allow for relations of integration (rather than exclusion, reduction, or replacement) between psychology and neurobiology as equally 'trivial', in allowing the retention of linguistic and other psychological terms in a 'complete' theory of mind, and in telling us 'nothing about the concepts that will be used in a successful theory' (section 2.2.1, para.2).

But where G&S's definition of 'biological neuroscience' was implausibly narrow, their definition of 'cognitive neuroscience' is too inclusive to be useful. As a label for a 'science of minimal commitments', their category includes a 'vast family of sciences' which might possibly contribute to an understanding of the mind (Stoljar and Gold 1998, p.130, p.111). Approaches as diverse as computational neuroscience and cognitive ethology *do*actively seek 'to synthesize biology and psychology in order to understand the mind' (sec.2.1, para.2): but many others who accept the basic materialism of the trivial neuron doctrine do not pursue this synthesis, and a definition of cognitive neuroscience that includes them is misleading. In, for example, Chomskian linguistics, psychoanalysis, and classical AI, many theorists study the *brain* only in the attenuated sense that, say, geologists or ecologists study particles. This is not yet a criticism; it might be, as the analogy makes clear, that direct study of the brain doesn't aid understanding of some mental phenomena. The Churchlands' targets are not the psychological and linguistic sciences per se, but only certain theories within those sciences. In context, P.M. Churchland's reference to "an alternative to, or potential reduction of" Chomskyan linguistics is clearly not a statement of RND (section 1.2, para.5) but an empirical bet that other (neuro)computational, thoroughly cognitive frameworks will better explain linguistic performance and competence.

G&S see Kandel's account of learning as a mere implemention, rather than a reduction, of psychological theory. This is a controversial, narrow picture of reduction, by which the reducing theory has to be entirely conceptually independent of the reduced theory. However, many philosophers of science hold that reductions can be *partial* (Bickle 1998). In a thoroughgoing discussion of Kandel's work, for example, Kenneth Schaffner (1992, pp.323-339) argues that reductive connections between psychology and neurobiology need not be simple. He acknowledges that the causal generalizations of theories such as Kandel's are 'typically *not*framed in purely biochemical terminology' but instead mix different levels. There is not even a single *neurobiological*level, as the model of molecular biological processes is integrated into, or 'seen as a more detailed expansion of the neural circuit for the gill-siphon reflex.' Genuine explanatory reductions will produce 'many weblike and bushy connections' across levels, with causal sequences described at many levels of aggregation. The generalizability of biological reductions is limited; some may be specific to the system in question. Thus, not even reductionists impressed by Kandel need claim that this kind of synaptic plasticity explains *all*forms of learning and memory, though Kandel himself seems tempted by RND (1987, p.viii). Reduction, on a range of more liberal views, is 'bound to be patchy' (Schaffner 1992, p.237; cf P.M. Churchland 1996, p.306 on 'objective knowledge of a highly idiosyncratic reality').

G&S rely on a sharp distinction between 'parasitic' theories, which merely specify implementing mechanisms for independent psychological functions, and genuinely reductive theories (such as the kinetic theory of heat) which render reduced items ('temperature') explanatorily redundant (section 5.3.3, para.2). In their view, explanations in neurobiology that rely on functional characterizations of the explananda are automatically (nonreductive) mere implementations. However, if a Schaffner-like picture of reduction is correct, this distinction breaks down, and many different relations *between*mere implementation and complete reduction are possible. A modified 'argument from unification as reduction' can then go through. G&S's strategy against this argument (section 4.2) is to set aside the 'enormous literature dealing with reductionism' and then to interpret reduction in a specific, implausibly strong way, as requiring direct and complete descent to the physical. If this was the only form of reductionism, then reductionists would refute themselves whenever they employ terms other than those of a completed fundamental physics, but it is not. The modified argument from reductive unification encourages close engagement, as exemplified by G&S, with the complex mesh of causal generalizations embedded in specific neurophysiological theories and importantly leaves open the possibility that, in some domains, psychological concepts may be (partially) revised. RND then becomes unnecessary; we get a modified conception of genuine reduction without inevitably dispensing with psychological concepts.

References

Churchland, P.M. and P.S. Churchland. 1998. Recent work on consciousness: philosophical, theoretical, and empirical. In *On the Contrary: critical essays, 1987-1997*. Cambridge: MIT Press.

Churchland, P.S. 1996. Feeling reasons. *In The Neurobiology of Decision-Making*, edited by A.R. Damasio et al. Berlin: Springer-Verlag.

Churchland, P.S. and V.S. Ramachandran. 1993. Filling in: why Dennett is wrong. In *Dennett and His Critics*, edited by B. Dahlbom. Oxford: Blackwell.

Enc, B. 1983. In defense of the identity theory. *Journal of Philosophy 80*:279-298.

Penrose, R. 1994. *Shadows of the Mind*. Oxford: Oxford University Press.

Schaffner, K.F. 1992. Philosophy of medicine. In *Introduction to the Philosophy of Science*, by Merrilee H. Salmon et al. Englewood Cliffs, N.J.: Prentice Hall.

Sutton, J. 1998. *Philosophy and Memory Traces: Descartes to connectionism*. Cambridge: Cambridge University Press.

Sutton, J. 1999. Distributed memory, coupling, and history. *Proceedings of the Fourth Australasian Cognitive Science Society*.