The Churchlands' neuron doctrine: Both cognitive and reductionist

Article *in* Behavioral and Brain Sciences · October 1999

DOI: 10.1017/S0140525X99462193

1 author:

John Sutton
Macquarie University
139 PUBLICATIONS **1,356** CITATIONS

SEE PROFILE

The Churchlands' neuron doctrine: both cognitive and reductionist

John Sutton
Macquarie University, Sydney, NSW 2109, Australia

Behavioral and Brain Sciences, 22 (1999), 850-1.
john.sutton@mq.edu.au
http://johnsutton.net/

Commentary on Gold & Stoljar, 'A Neuron Doctrine in the Philosophy of Neuroscience', *Behavioral and Brain Sciences, 22* (1999), 809-869: full paper and commentaries online at: http://www.stanford.edu/~paulsko/papers/GSND.pdf

Abstract: According to Gold and Stoljar, one cannot both consistently be reductionist about psychoneural relations *and* invoke concepts developed in the psychological sciences. I deny the utility of their distinction between biological and cognitive neuroscience, suggesting that they construe biological neuroscience too rigidly and cognitive neuroscience too liberally. Then I reject their characterization of reductionism: reductions need not go down past neurobiology straight to physics, and cases of partial, local reduction are not neatly distinguishable from cases of mere implementation. Modifying the argument from unification-as-reduction, I defend a position weaker than the radical, but stronger than the trivial neuron doctrine.

Gold and Stoljar (G&S) allow 'biological neuroscience' to *include* study of the *function* as well as the structure of neuronal ensembles (sec.2.1). But they think that invocations of function in actual neurobiological explanations already invoke non-biological concepts, so that explanations of causal mechanisms fulfilling those functions are *not* purely neurobiological. Because even an apparently physiological notion like the reflex is 'highly theoretical', G&S deny that it is a legitimate construct of physiology alone (n.40). It's as if the fact that neurophysiology, as Enc says, '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! So 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 like classical conditioning, information, and representation consistently to propose RND. There may be some such 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. But, surprisingly, they are not the targets. Instead, G&S implausibly interpret the Churchlands as supporters of RND. But 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.224-230). Indeed recognizably connectionist frameworks of explicit and implicit memory representation were developed by early modern theorists like 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' (sec.2.2.1, para.2). G&S's definition of 'cognitive neuroscience' is too inclusive. As a label for a 'science of minimal commitments', their category includes a 'vast family of sciences' which might contribute to an understanding of mentality (Stoljar and Gold 1998, pp.130,111). 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 which 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, since 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 (sec.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. But 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 like 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, as some may be specific to the system in question. So 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.337; compare 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 (like the kinetic theory of heat) which render reduced terms ('temperature') explanatorily redundant (sec.5.3.3, para.2). On their view, explanations in neurobiology which rely on functional characterizations of the explananda are automatically (non-reductive) mere implementations. But if a Schaffner-like picture of reduction is right, 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 (sec.4.2) is to set aside the 'enormous literature dealing with reductionism', and then to interpret reductionism 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 use 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, since we get a modified conception of genuine reduction without inevitably dispensing with psychological concepts.

References (additional to target article)

Bickle, John. 1998. *Psychoneural Reduction: the new wave.* Cambridge, MA: MIT Press.

Churchland, P.M. 1996. Flanagan on moral knowledge. In *The Churchlands and Their Critics,* edited by R. McCauley. Oxford. Blackwell.

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.

Kandel, E.R. 1987. Preface: molecular neurobiology and the proper study of humankind. In *Molecular Neurobiology in Neurology and Psychiatry,* edited by E.R. Kandel. New York: Raven Press.

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*.

Van Gelder, T. 1995. What might cognition be, if not computation? *Journal of Philosophy* 92:345-381.