Monaghan, P., Shillcock, R. & McDonald, S. (2004) Hemispheric asymmetries in the split-fovea model of semantic processing. *Brain and Language* 88(3):339–54. [rFP]
Neville, H. J., Mills, D. L. & Lawson, D. S. (1992) Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex* 2:244–58. [rFP, SJS]
Oliveri, M., Finocchiaro, C., Shapiro, K., Gangitano, M., Caramazza, A. & Pascual-Leone, A. (2004) All talk and no action: A transcranial magnetic stimulation study of motor cortex activation during action word production. *Journal of Cognitive Neuroscience* 16(3):374–81. [rFP]
Pulvermüller, F. (1999) Words in the brain's language. *Behavioral and Brain Sciences* 22(2): 253–336. [rFP, SJS]
(2002) A brain perspective on language mechanisms: From discrete neuronal ensembles to serial order. *Progress in Neurobiology* 67:85–111. [rFP]
(2003) *The neuroscience of language.* Cambridge University Press. [rFP]
Pulvermüller, F., Hummel, F. & Härle, M. (2001) Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language* 78(2):143–68. [rFP]
Pulvermüller, F., Lutzenberger, W. & Birbaumer, N. (1995) Electrocortical distinction of vocabulary types. *Electroencephalography and Clinical Neurophysiology* 94:357–70. [rFP, SJS]
Pulvermüller, F., Preissl, H., Lutzenberger, W. & Birbaumer, N. (1996) Brain rhythms of language: Nouns versus verbs. *European Journal of Neuroscience* 8:937–41. [rFP]
Pulvermüller, F., Shtyrov, Y. & Ilmoniemi, R. J. (2003) Spatio-temporal patterns of neural language processing: An MEG study using Minimum-Norm Current Estimates. *Neuroimage* 20:1020–25. [rFP]
Pulvermüller, F., Shtyrov, Y., Kujala, T. & Naatanen, R. (2004) Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology* 41(1):106–12. [rFP]
Schmauder, A. R. (1996) Ability to stand alone and processing of open-class and closed-class words: Isolation versus context. *Journal of Psycholinguistic Research* 25:443–81. [SJS]
Segalowitz, S. J. & Lane, K. C. (2000) Lexical access of function versus content words. *Brain and Language* 75(3):376–89. [rFP, SJS]
Shallice, T. (1988) *From neuropsychology to mental structure.* Cambridge University Press. [rFP]
Shtyrov, Y., Hauk, O. & Pulvermüller, F. (2004) Distributed neuronal networks for encoding category-specific semantic information: The mismatch negativity to action words. *European Journal of Neuroscience* 19(4):1083–92. [rFP]
Taft, M. (1990) Lexical processing of functionally constrained words. *Journal of Memory and Language* 29:245–57. [SJS]
Warrington, E. K. & Shallice, T. (1984) Category specific semantic impairments. *Brain* 107:829–54. [rFP]
Whitaker, H. A. (1983) Towards a brain model of automization: A short essay. In: *Memory and control of action,* ed. R. A. Magill, pp. 199–214. North-Holland. [SJS]

Commentary on **Ian Gold & Daniel Stoljar (1999). A neuron doctrine in the philosophy of neuroscience. BBS 22(5):809–869.**

Abstract of the original article: Many neuroscientists and philosophers endorse a view about the explanatory reach of neuroscience (which we will call the *neuron doctrine*) to the effect that the framework for understanding the mind will be developed by neuroscience; or, as we will put it, that a successful theory of the mind will be solely neuroscientific. It is a consequence of this view that the sciences of the mind that cannot be expressed by means of neuroscientific concepts alone count as indirect sciences that will be discarded as neuroscience matures. This consequence is what makes the doctrine substantive, indeed, radical. We ask, first, what the neuron doctrine means and, second, whether it is true. In answer to the first question, we distinguish two versions of the doctrine. One version, the *trivial* neuron doctrine, turns out to be uncontroversial but unsubstantive because it fails to have the consequence that the nonneuroscientific sciences of the mind will eventually be discarded. A second version, the *radical* neuron doctrine, *does* have this consequence, but, unlike the first doctrine, is highly controversial. We argue that the neuron doctrine appears to be both substantive and uncontroversial only as a result of a conflation of these two versions. We then consider whether the radical doctrine is true. We present and evaluate three arguments for it, based either on general scientific and philosophical considerations or on the details of neuroscience itself, arguing that all three fail. We conclude that the evidence fails to support the radical neuron doctrine.

Could the neural ABC explain the mind?

Maurice K. D. Schouten[a] and Huib Looren de Jong[b]
[a]*Faculty of Philosophy, Tilburg University, 5000 LE Tilburg, The Netherlands;* [b]*Department of Psychology, Vrije Universiteit, 1081 BT Amsterdam, The Netherlands.* **m.k.d.schouten@uvt.nl** **h.looren.de.jong@psy.vu.nl**

Abstract: Gold & Stoljar are right in rejecting the radical neuron doctrine, but we argue that their distinction between determination and explanation is not principled enough to support their conclusion. We claim that the notions of multiple supervenience and screening-off offer a more precise construal of the dissociation between explanation and determination that lies at the heart of the antireductionist position.

Neuron doctrine: Trivial and radical. Gold & Stoljar (1999; henceforth G&S) distinguish two varieties of neuron doctrine. One option they consider is the radical neuron doctrine (RND), a view in which explanation and determination (wrongly, in their opinion) hang together: neural properties determine and ipso facto explain psychological properties. The alternative is the trivial neuron doctrine (TND), which states that although mental properties are *determined* by neurobiological properties, they are not necessarily also *explained* by them.

We agree that the difference between the trivial and radical neuron doctrines lies in separating determination and explanation. However, we also believe that G&S's distinction between explanation and determination is not principled enough to support the TND-RND distinction. G&S analyse Kandel's experiments as support for their antireductionist position. However, Kandel is a self-confessed reductionist, and is interpreted in this light by philosophers like Schaffner (1993) and Bickle (1998). Kandel and colleagues claim that the neural plasticity paradigm offers "surprising reductionist possibilities" (Kandel et al. 1995, p. 389). They employ the metaphor of a molecular or neural alphabet (see also Hawkins & Kandel 1984): What is necessary and sufficient to understanding learning and memory in all of their varieties is a full specification of the letters of this alphabet and of the possible combinations in which these letters can be strung together. Contrary to what G&S claim, Kandel's view implies that the molecular and neural letters in the end not only determine, but also explain and replace the psychological phenomena of learning and memory. What is suggested here is reductionism of a very special sort, namely, *combinatorial reductionism.* Thus, without a principled distinction between determination and explanation, the TND may well collapse into the RND. Two arguments are discussed below to suggest that determination and explanation can still be kept apart.

Multiple supervenience. There is a well-known argument that higher-level sciences, such as biology and psychology, deal in functional explanations. The notion of function allows generalisations, and even laws, abstracting over lower-level mechanisms that would be wildly heterogeneous from a physical perspective. The

https://doi.org/10.1017/S0140525X04330074 Published online by Cambridge University Press

function of spatial navigation, for example, offers a grouping that may be implemented in many different substrates. It may however still be claimed that a function in a particular organism is *determined* by a specific substrate and that the substrate also *explains* the function, so that the gates to reductionism (and the RND) are still wide open (Kim 1998). We suggest the argument from multiple supervenience (MS) to block the interference from determinism to reductionism (see also, Schouten & Looren de Jong 1999). The psychoneural supervenience thesis states that psychological properties are *determined* by neural properties. However, a single neural property determines not just a single higher-level (dispositional) property, but a multitude of them. To pick out from this set of supervenient properties the one particular property that is explanatorily salient (often a functional property), requires information that goes beyond what is present in a full specification of the microlevel details. What is explanatorily interesting about the neural "letters" that compose the "words" of learning and memory is something that cannot be read off from the neural alphabet alone. In order to make this selection of causally relevant properties from the supervenience base, a higher-level perspective is needed. MS thus honors determinism of psychological properties and yet it grants an irreducible role to higher-level, often functional explanation.

It may be objected, however, that functional ascriptions will turn out to be nothing more than heuristics. That is, when all the data are in and all the mechanisms are known, they drop out of the scientific picture as mere convenient fictions. When neuroscience would finish its job, all the explanatory weight shifts to the letters of the neural or molecular ABC. This suggests that the proffered functional higher-level explanation refers to something that may be causally irrelevant; hence, these higher-level explanations may not count as bona fide explanations. So, a further argument is required in order to uphold a more objective and qualitative distinction between level of explanations.

Screening off. We believe such an additional argument, required to establish the objective, and not merely heuristic, relevance of higher-level explanations, can be found in the application of the so-called screening-off (S-O) rule. Stated in the language of conditional probabilities, the formula for one cause M screening off another cause P from its outcome R is

$$M \text{ screens off } P \text{ from } R \text{ iff}$$
$$\Pr(R \mid P \,\&\, M) = \Pr(R \mid M) \neq \Pr(R \mid P).$$

To put it colloquially, a property M screens off a property P if adding P does not improve the prediction or explanation of outcome R, whereas M does improve the prediction or explanation of R. Brandon (1990) employed this screening-off relation to substantiate the claim that phenotypes (M) and not the genes or the genotypes behind the phenotypes (P) are relevant in explaining reproductive success (R). Because the genotype is asymmetrically dependent on the phenotype with respect to natural selection (for its effect on R, P depends upon M being present or absent, whereas M contributes to R irrespective of P's being present or absent), it is the phenotype that offers the best causal explanation of reproductive success (Brandon 1990, pp. 83–85). The phenotypic level has a causal efficacy and explanatory legitimacy of its own, even if the phenotype is determined by the genotype (among other things). Identifying phenotypic traits is not a merely heuristic, free-for-all, essentially void kind of explanation, but rather, it taps real causal factors in an organism's chances of survival. In the same way, behavior may be explained in terms of information processing, screening off the underlying neural processes that determine it. Although cognition is *dependent* on neurons, it cannot be exhaustively *explained* in neuronal terms.

Screening-off thus serves as a criterion to distinguish causally relevant levels of explanation from causally irrelevant ones; it affords a criterion for "picking levels" (McClamrock 1995). While admitting that cognition is determined by neurons (and ultimately by atoms, molecules, bosons, superstrings, and what have you), we can maintain that higher functional levels genuinely and irreducibly explain behavior that is determined by lower neural levels. G&S are right in rejecting RND, but we submit that the notions of multiple supervenience and screening-off offer a more precise construal of the dissociation between explanation and determination that lies at the heart of the antireductionist position.

Editor's Note: There is no Author's Response for this commentary.

References

Bickle, J. (1998) *Psychoneural reduction: The new wave.* MIT Press. [MKDS]

Brandon, R. N. (1990) *Adaptation and environment.* Princeton University Press. [MKDS]

Hawkins, R. D. & Kandel, E. R. (1984) Is there a cell-biological alphabet for simple forms of learning? *Psychological Review* 91:375–91. [MKDS]

Kandel, E. R., Schwartz, J. H. & Jessell, T. M. (1995) *Essentials of neural science and behavior.* Appleton & Lange. [MKDS]

Kim, J. (1998) *Mind in a physical world: An essay on the mind-body problem and mental causation.* MIT Press. [MKDS]

McClamrock, R. (1995) *Existential cognition: Computational minds in the world.* The University of Chicago Press. [MKDS]

Schaffner, K. F. (1993) *Discovery and explanation in biology and medicine.* The University of Chicago Press. [MKDS]

Schouten, M. K. D. & Looren de Jong, H. (1999) Reduction, elimination, and levels: The case of the LTP-learning link. *Philosophical Psychology* 12:237–62. [MKDS]

https://doi.org/10.1017/S0140525X04330074 Published online by Cambridge University Press