between the activity difference δ'_t and the regular temporal difference δ_t never substantially affects the plasticity of a synapse. This happens, for instance, if there is only in fact one single reward right at the end of the trial [e.g. r_t=0, $t < T$; $r(T)$=1]. In a simplification of the version of this that Rao and Sejnowski consider, $r(T)$=1 comes from an action potential, caused by a privileged input to the postsynaptic cell, backpropagating up its dendritic tree. This makes positive the activity difference associated with pre-synaptic events initiated at time t=T–1 in a trial, thus engendering increases in synaptic efficacy. These, in turn, reduce the activity difference, by increasing P_{T-1}, until the difference reaches 0. This process can lead to the prior pre-synaptic events causing the post-synaptic cell to spike in a preditive manner. Rao and Sejnowski further suggest a specialized inhibitory connection architecture [18], which allows the predictive spike to cancel out the predicted spike (thus eliminating the effect of the difference between δ_t and δ'_t).

In the converse case, what happens if indeed δ'_t is used in the learning rule rather than δ_t? I don't know of compelling computational analyses of this case, other than the obvious point that the resulting learning rule looks like a correlational learning rule between the stimuli and the *differences* in successive outputs.

Rao and Sejnowski face the even trickier problem of making the learning rules work in the face of biophysically realistic timescales for synaptic currents and membrane potentials and the like. The most dangerous problem that arises is instability, that the learning rule can make the synaptic efficacies rise without bound. This happens when the biophysical mechanism for propagating information around the post-synaptic cell (backpropagating action potentials) lasts over a longer time scale than that involved in the derivative P_{t+1}–P_t. That can make the learning rule operate more like a regular correlational learning rule, and these are notoriously unstable. Synaptic saturation is suggested as a possible fix, although one might worry about a consequent loss of synaptic selectivity.

Altogether, the notion that temporally asymmetric Hebbian learning rules are best seen in predictive rather than correlational terms has been taken in various interesting directions. Rao and Sejnowski usefully add to our armoury of ways of approaching such rules, and remind us of an essential Yogic truth.

Acknowledgements

I am grateful to Raj Rao and Chris Watkins for helpful comments.
The author's research is funded by the Gatsby Charitable Foundation.

References

1 Sejnowski, T.J. (1999) The book of Hebb. *Neuron* 24, 773–776
2 Abbott, L.F. and Blum, K.I. (1996) Functional significance of long-term potentiation for sequence learning and prediction. *Cereb. Cortex* 6, 406–416
3 Sutton, R.S. (1988) Learning to predict by the methods of temporal difference. *Mach. Learn.* 3, 9–44
4 Sutton, R.S. and Barto, A.G. (1990) Time-derivative models of Pavlovian conditioning. In *Learning and Computational Neuroscience*, (Gabriel, M. and Moore, J.W., eds), pp. 497–537, MIT Press
5 Montague, P.R. *et al.* (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947
6 Schultz, W. *et al.* (1997) A neural substrate of prediction and reward. *Science* 275, 1593–1599
7 Schultz, W. (1998) Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27
8 Rao, R.P.N. and Sejnowski, T.K. (2001) Spike-timing dependent Hebbian plasticity as temporal difference learning. *Neural Comput.* 13, 2221–2237
9 Sutton, R.S. and Barto, A.G. (1981) Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* 88, 135–170
10 Rescorla, R.A. and Wagner, A.R. (1972) A theory of Pavlovian conditioning: the effectiveness of reinforcement and non-reinforcement. In *Classical Conditioning Vol. II: Current Research and Theory* (Black, A.H. and Prokasy, W.F., eds), pp. 64–69, Appleton-Century-Crofts
11 Sutton, R.S. and Barto, A.G. (1998) *Reinforcement Learning*, MIT Press
12 Bertsekas, D.P. and Tsitsiklis, J.N. (1996) *Neuro-Dynamic Programming*, Athena Scientific
13 Barto, A.G. *et al.* (1990) Learning and sequential decision making. In *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, (Gabriel, M. and Moore, J.W., eds), pp. 539–602, MIT Press
14 Grossberg, S. and Schmajuk, N.A. (1987) Neural dynamics of attentionally modulated Pavlovian conditioning: conditioned reinforcement, inhibition, and opponent processing. *Psychobiology* 15, 195–240
15 Abbott, L.F. *et al.* (1997) Synaptic depression and cortical gain control. *Science* 275, 220–224
16 Tsodyks, M.V. and Markram, H. (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. U. S. A.* 94, 719–723
17 Thomson, A.M. (2000) Facilitation, augmentation and potentiation at central synapses. *Trends Neurosci.* 23, 305–312
18 Rao, R.P.N. and Sejnowski, T.J. (2000) Predictive sequence learning in recurrent neocortical circuits. In *Advances in Neural Information Processing Systems* (Vol. 12) (Solla, S.A. *et al.*, eds), pp. 164–170, MIT Press

Peter Dayan
Gatsby Computational Neuroscience Unit, University College London, UK WC1E 6BT.
e-mail: dayan@gatsby.ucl.ac.uk

Debate

Neuroecology and psychological modularity

Jonathan I. Flombaum, Laurie R. Santos and Marc D. Hauser

In a recent review, Bolhuis and Macphail challenge the thesis that specialized systems mediate the learning, encoding and retrieval of different types of information – what they call a neuroecological approach to learning and memory [1]. In particular, they challenge 'the arbitrary assumption that different "problems" engage different memory mechanisms' (p. 426), and the idea that this fact can be used to motivate neurobiological studies. To substantiate their claims, they appeal to data dealing with the neural substrates of song learning and food storage in birds. Recently, Hampton *et al.* [2] pointed out how Bolhuis and Macphail misrepresent these data and set-up a 'straw neuroecologist' with respect to the functionalist/adaptionist perspective. Here, we take up a different problem.

 1364-6613/02/$ – see front matter PII: S1364-6613(02)01872-7

Namely, we argue that neurobiological data cannot be used to bring a case against the thesis of psychological modularity, whether in the case of memory and learning, or otherwise. Data of this sort, although relevant to a discussion of the principles underlying modularity, are orthogonal to a discussion of whether or not modularity, in the psychological sense, exists in the first place.

As Bolhuis and Macphail point out, much of the interest in the idea of modularity comes from the work of the philosopher Jerry Fodor [3]. Since their original conception in 1983, Fodor's ideas have generated many debates and undergone many revisions and reincarnations [4]. Evolutionary psychologists, for example, have argued that most psychological systems, not just the perceptual ones, are modular, and that these modules have been shaped by natural selection [5]. In this article, we stick with Fodor's original conception of modularity because this is the version to which Bolhuis and Macphail refer.

Modularity is the thesis that the mind contains independent input systems that are restricted in the types of information that they can consult [3]. Therefore, the fundamental characteristic of a modular system is that information that is accessible to one module is not necessarily available to all the other modules within that particular system. In other words, to use the Fodorian term, modules are 'informationally encapsulated' (Ref. [3], p. 71). A classic example of an informationally encapsulated module is the perception of line length. Fodor discusses the famous Müller-Lyer illusion: although many of us have seen this illusion a number of times and know that that the two middle lines are exactly the same length, we are still fooled into perceiving the line with the acute arrowheads as longer than the one with obtuse arrowheads. This illusion persists even after we measure the lines with a ruler because our visual system cannot make use of the knowledge we can derive from specific measurements. The lack of connection between these systems, their encapsulation, represents the signature of modules in the Fodorian sense. Because modules are encapsulated, we can expect them to have other characteristics as well. They are often (though not necessarily) mandatory, fast, shallow in output, susceptible to selective impairment [3] and, sometimes, executed in a highly specific neural locus [6,7].

Bolhuis and Macphail present a somewhat different and, we believe, incorrect view of modularity, one that leads them to question the usefulness of neuroecology. We argue that Bolhuis and Macphail make three incorrect claims about modules, which we address in turn.

Modules are not always innate

Bolhuis and Macphail claim that modules 'are innately specified – that is, they are species specific.' (p. 427, Box 1). Because the fundamental characteristic of modularity is informational encapsulation, we see no *a priori* reason to believe that the eventual manifestation of a module must be innately predetermined. Some have argued, in fact, that modules can arise exclusively from experience [6–8]. Simply put, restrictions on information access can be implemented through experience; they need not be predetermined. Processing written words, for example, is a modular system that is not innate; an illiterate person will not automatically read words, show characteristic Stroop effects, and so forth. By contrast, those who have learned to read cannot avoid reading every word that they see; that is to say, the reading module is acquired. There is no reason, therefore, to assume that modules must be innate. Similarly, 'species-specificity' is neither the same as 'innately specified,' nor is it a requirement of modularity as Bolhuis and Macphail suggest. This is clear when we again consider Fodor's line perception example, a modular system that is most probably shared across the primate order. Perhaps, what Bolhuis and Macphail mean by 'species-specific' is that all members of a particular species share a particular module, not that the module is absent in other species. If this is the case, however, then one need only reconsider the reading example to see why this characteristic is not necessarily true of all modules.

Modules need not be domain-specific

We also disagree with the claim made by Bolhuis and Macphail that modules are 'domain-specific' (p. 427). Although there is certainly controversy within the cognitive sciences with respect to defining a domain, most would agree that it represents a finite computational problem space in which a given system operates. Thus, a domain-specific mechanism becomes engaged only when faced with particular types of problems, and operates by picking out certain relevant features, rejecting others, and using specialized learning mechanisms. The domain of song learning, for example, refers to the system (or set of systems) that processes information associated with recognizing and producing conspecific song. The essence of domain-*specificity*, then, is the context-dependent application of a set of psychological mechanisms. The contexts, or domains, for which domain-specific mechanisms emerge are typically those that play an ecologically relevant role for a given organism [5,9], although domain-specific mechanisms can certainly be shared across different animal species. For example, the computations that underlie the approximation of large numbers are shared among birds, rodents, non-human primates and humans [10].

'...neurobiological data cannot challenge the claim that the psychological mechanisms ... are modular.'

Although the notion of a domain restricts the class of computational problems to be considered (acting like a filter) it does not *necessarily* make any restrictions on the inputs that are considered while solving particular problems [3,6,7]. This differs from modules, therefore, in that domain-specific mechanisms need not be informationally encapsulated. Modules, on the other hand, make no restriction on the specific type of computational problem being solved. Line-length modules, for example, are used to solve problems from a variety of different domains, including resolving visual illusions, recognizing words, and distinguishing between faces; line-length perception is thus a domain-general module. For these reasons, domain-specificity is not a fundamental criterion of modularity, as Bolhuis and Macphail have suggested. Although modules are often domain-specific, as Fodor [3] and others [7] have pointed out, they need not be.

We raise these problems with Bolhuis and Macphail's view of domain-specificity

and modularity because they lead them to conclude that song learning and food storage in birds are not domain-specific computations. They argue this in two ways. First, they appeal to the fact that memory is not modular by Fodor's account and is, instead, what he considered a central system [3]. If a module is domain-specific and memory is not modular, then, so their reasoning goes, memory is not domain-specific. The logic here is flawed, however, because Fodor's view of modularity does not require domain specificity as a necessary or sufficient criterion. In particular, birds may not be limited in the information that they access while learning song or retrieving food; nevertheless, birds might well deploy specialized learning mechanisms in these contexts. Second, Bolhuis and Macphail use neurobiological evidence to argue that the mechanisms of song learning and food retrieval are not modular, and therefore, not domain specific. However, since domain-specific mechanisms need not be modular (in the sense of information encapsulation), these data cannot challenge the claim that these mechanisms are domain-specific. This leads us to our final point about modularity, and our main critique.

Modules are not always localized in a specific part of the brain

We disagree with Bolhuis and Macphail's claim that modules are 'hardwired, and located in specific brain regions' (p. 427, Box 1). More specifically, we argue that if modules do not need to be located in specific brain areas, then neurobiological data cannot challenge the claim that the psychological mechanisms underlying song learning and food retrieval are modular. Bolhuis and Macphail's error arises in assuming that the execution of a psychological module in a similarly 'modular' piece of brain is, like information encapsulation, a necessary component of modularity. Their logic is that if a system is *psychologically* modular, it must have a modular *neural* substrate as well: 'If functional requirements lead to the evolution of different cognitive modules, so the reasoning goes, then there should also be neural modules that are specialized for a particular function' (p. 426). By this reasoning, evidence that a hypothesized psychological module, such as a mechanism for learning song or remembering food locations, does not have a perfectly modular neural substrate falsifies the claim that the psychological module exists in the first place. This is the type of evidence presented by Bolhuis and Macphail. It is not, however, the case that a psychological module must have a corresponding neural module in the sense of a specified brain nucleus.

The thesis that a system is limited in the types of information that it can access does not imply that this system has a requisite form of physical implementation. We can imagine the general case easily if we consider a computer. We might argue that a particular program is modular if we limit the program to accessing certain databases in the computer, but not others. However, this claim tells us nothing about how we expect the computer hardware to be organized. Consider again the example of line-length perception. Would we argue that length perception was not modular if it were found that length is realized jointly by a neuron in the occipital lobe, a neuron in the prefrontal cortex, and a neuron in the forebrain? The obvious answer to this question is a resounding 'No'. By this logic, then, the data presented by Bolhuis and Macphail simply cannot falsify the conclusion that birds have modular systems for song learning and food retrieval, nor can data of this sort challenge the more general claim that there are specialized mechanisms for learning and memory.

Distributed neurobiological architectures

To be fair, Bolhuis and Macphail are not alone in their misconceptions of modularity, particularly in the notion that psychological modularity assumes neural modularity. Many cognitive scientists, for example, have either implicitly or explicitly presupposed a necessary link between modularity at the cognitive and neural levels [8,10]. Our point is simply that Bolhuis and Macphail's assumptions about modularity are just that – assumptions. And sadly, they are assumptions that detract from what we feel is the most important point of their review, namely, that seemingly domain-specific cognitive systems *can* in fact be implemented in a distributed neural substrate. The data Bolhuis and Macphail review on song learning and food storage in birds suggest that distributed neurobiological architectures can subserve specified cognitive functions. These data are important in light of the fact that cognitive neuroscientists still know relatively little about how neuroanatomical structures constrain cognitive computations, and, particularly, about how specificity at the neural level impacts the implementation (and evolution) of cognitive specificity. In the same way that cognitive scientists working on humans should avail themselves of the existing comparative literature in order to understand whether the modules that subserve human thought evolved recently or in the distant past, ethologists working on non-human animals must avail themselves of the appropriate cognitive theories in order to ensure that they are shooting at the right targets.

References

1 Bolhuis, J.J. and Macphail, E.M. (2001) A critique of the neuroecology of learning and memory. *Trends Cogn. Sci.* 5, 426–433
2 Hampton, R.R. *et al.* (2002) 'Neuroecologists' are not made of straw. *Trends Cogn. Sci.* 6, 6–7
3 Fodor, J. (1983) *The Modularity of Mind*, MIT Press
4 Coltheart, M. (1999) Modularity and cognition. *Trends Cogn. Sci.* 3, 115–120
5 Cosmides, L. and Tooby, J. (1994) Origins of domain-specificity: the evolution of functional organization. In *Mapping the Mind: Domain Specificity in Cognition and Culture* (Hirschfeld, L.A. and Gelman, S.A., eds.), pp. 85–116, Cambridge University Press
6 Scholl, B.J. (1997) Reasoning, rationality, and architectural resolution. *Philos. Psychol.* 10, 451–470
7 Scholl, B.J. (1997) Neural constraints on cognitive modularity? *Behav. Brain Sci.* 20, 575–576
8 Karmiloff-Smith, A. (1992) *Beyond Modularity: A Developmental Perspective on Cognitive Science*, MIT Press
9 Duchaine, B. *et al.* (2001) Evolutionary psychology and the brain. *Curr. Opin. Neurobiol.* 11, 225–230
10 Gallistel, C.R. and Gelman, R. (2000) Non-verbal cognition: from reals to integers. *Trends Cogn. Sci.* 4, 59–65

Jonathan I. Flombaum*
Laurie R. Santos
Marc D. Hauser
Primate Cognitive Neuroscience Laboratory, Harvard University, Dept of Psychology, 33-Kirkland Street, Cambridge, MA 02138, USA.
*e-mail: flombaum@fas.harvard.edu