The Leabra architecture: Specialization without modularity

doi:10.1017/S0140525X10001160

Alexander A. Petrov,[a] David J. Jilk,[b] and Randall C. O'Reilly[c]

[a]*Department of Psychology, Ohio State University, Columbus, OH 43210;* [b]*eCortex, Inc., Boulder, CO 80301;* [c]*Department of Psychology and Neuroscience, University of Colorado, Boulder, CO 80309.*

apetrov@alexpetrov.com **david.jilk@e-cortex.com**
Randy.OReilly@colorado.edu
http://alexpetrov.com **http://www.e-cortex.com**
http://psych.colorado.edu/~oreilly

Abstract: The posterior cortex, hippocampus, and prefrontal cortex in the Leabra architecture are specialized in terms of various neural parameters, and thus are predilections for learning and processing, but domain-general in terms of cognitive functions such as face recognition. Also, these areas are not encapsulated and violate Fodorian criteria for modularity. Anderson's terminology obscures these important points, but we applaud his overall message.

Anderson's target article adds to a growing literature (e.g., Mesulam 1990; Prinz 2006; Uttal 2001) that criticizes the recurring tendency to partition the brain into localized modules (e.g., Carruthers 2006; Tooby & Cosmides 1992). Ironically, Anderson's critique of modularity is steeped in modularist terms such as *redeployment*. We are sympathetic with the general thrust of Anderson's theory and find it very compatible with the Leabra tripartite architecture (O'Reilly 1998; O'Reilly & Munakata 2000). It seems that much of the controversy can be traced back to terminological confusion and false dichotomies. Our goal in this commentary is to dispel some of the confusion and clarify Leabra's position on modularity.

The target article is vague about the key term *function*. In his earlier work, Anderson follows Fodor (2000) in "the pragmatic definition of a (cognitive) function as whatever appears in one of the boxes in a psychologist's diagram of cognitive processing" (Anderson 2007c, p. 144). Although convenient for a meta-review of 1,469 fMRI experiments (Anderson 2007a; 2007c), this definition contributes little to terminological clarity. In particular, when we (Atallah et al. 2004, p. 253) wrote that "different brain areas clearly have some degree of specialized function," we did *not* mean cognitive functions such as face recognition. What we meant is closest to what Anderson calls "cortical biases" or, following Bergeron (2007), "working."

Specifically, the posterior cortex in Leabra specializes in slow interleaved learning that tends to develop overlapping distributed representations, which in turn promote similarity-based generalization. This computational capability can be used in a myriad of cognitive functions (O'Reilly & Munakata 2000). The hippocampus and the surrounding structures in the medial temporal lobe (MTL) specialize in rapid learning of sparse conjunctive

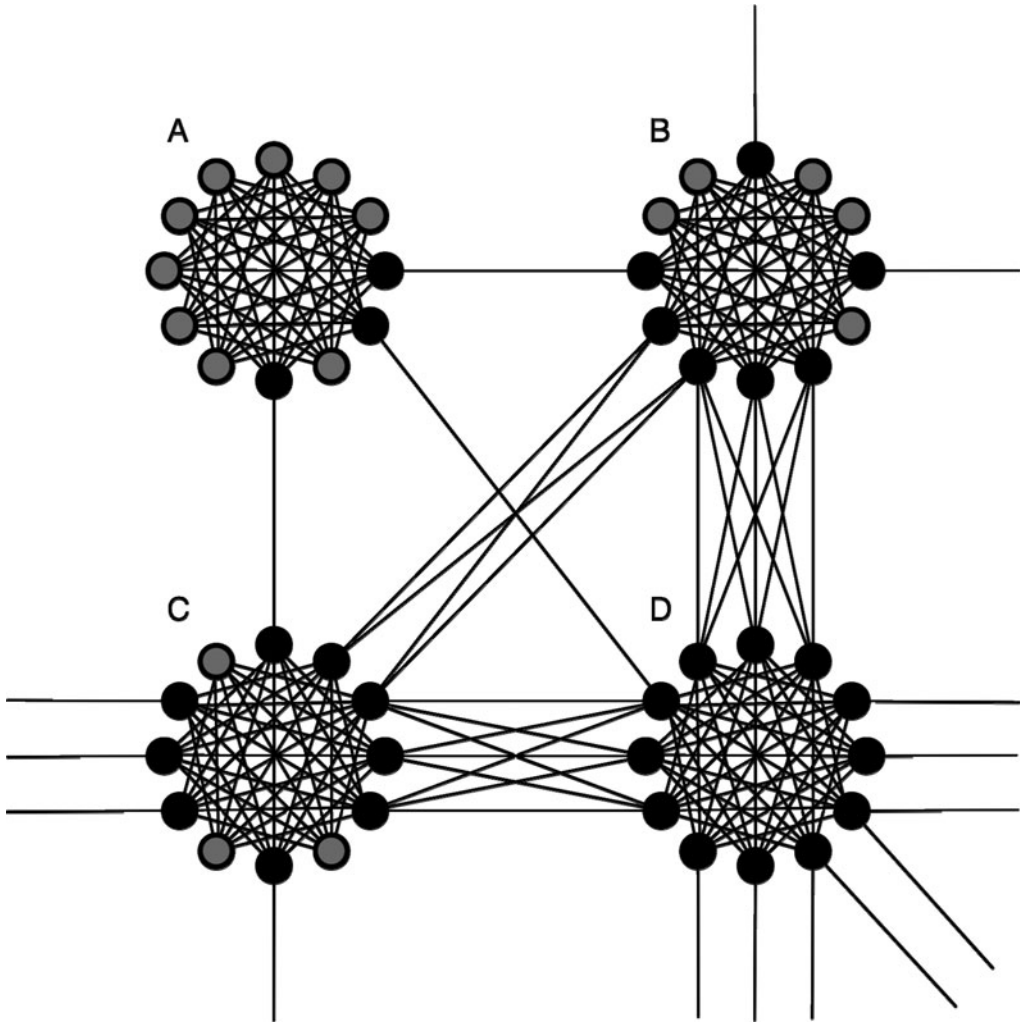

Figure 1 (Petrov et al.) Information encapsulation is a matter of degree. Four neuronal clusters are shown, of which A is the most and D the least encapsulated. Black circles depict exposed (input/output) units that make distal connections to other cluster(s); grey circles depict hidden units that make local connections only.

representations that minimize interference (e.g., McClelland et al. 1995). The prefrontal cortex (PFC) specializes in sustained neural firing (e.g., Miller & Cohen 2001; O'Reilly 2006) and relies on dynamic gating from the basal ganglia (BG) to satisfy the conflicting demands of rapid updating of (relevant) information, on one hand, and robust maintenance in the face of new (and distracting) information, on the other (e.g., Atallah et al. 2004; O'Reilly & Frank 2006). Importantly, most[1] of this specialization arises from parametric variation of the same underlying substrate. The components of the Leabra architecture differ in their learning rates, the amount of lateral inhibition, and so on, but not in the nature of their processing units. Also, they are in constant, intensive interaction. Each high-level task engages all three components (O'Reilly et al. 1999; O'Reilly & Munakata 2000).

We now turn to the question of modularity. Here the terminology is relatively clear (e.g., Carruthers 2006; Fodor 1983; 2000; Prinz 2006; Samuels 2006). Fodor's (1983) foundational book identified nine criteria for modularity. We have space to discuss only *domain specificity* and *encapsulation*. These two are widely regarded as most central (Fodor 2000; Samuels 2006).

A system is *domain-specific* (as opposed to *domain-general*) when it only receives inputs concerning a certain subject matter. All three Leabra components are domain-general in this sense. Both MTL and PFC/BG receive convergent inputs from multiple and variegated brain areas. The posterior cortex is an interactive multitude of cortical areas whose specificity is a matter of degree and varies considerably.

The central claim of Anderson's massive redeployment hypothesis (MRH) is that most brain areas are much closer to the general than the specific end of the spectrum. This claim is hardly original, but it is worth repeating because the subtractive fMRI methodology tends to obscure it (Uttal 2001). fMRI is a wonderful tool, but it should be interpreted with care (Poldrack 2006). Any stimulus provokes a large response throughout the brain, and a typical fMRI study reports tiny differences[2] between conditions – typically less than 1% (Huettel et al. 2008). The importance of Anderson's (2007a; 2007c) meta-analyses is that, even if we grant the (generous) assumption that fMRI can reliably index specificity, one still finds widespread evidence for generality.

MRH also predicts a correlation between the degree of generality and phylogenetic age. We are skeptical of the use of the posterior-anterior axis as a proxy for age because it is confounded with many other factors. Also, the emphasis on age encourages terms such as *reuse*, *redeployment*, and *recycling*, that misleadingly suggest that each area was deployed for one primordial and specific function in the evolutionary past and was later *re*deployed for additional functions. Such inferences must be based on comparative data from multiple species. As the target article is confined to human fMRI, the situation is quite different. Given a fixed evolutionary endowment and relatively stable environment, each human child develops and/or learns many cognitive functions simultaneously. This seems to leave no room for *re*deployment but only for deployment for multiple uses.

Anderson's critique of modularity neglects one of its central features – *information encapsulation*. We wonder what predictions MRH makes about this important issue. A system is encapsulated when it exchanges[3] relatively little information with other systems. Again, this is a matter of degree, as our Figure 1 illustrates. The degree of encapsulation depends on factors such as the number of exposed (input/output) units relative to the total number of units in the cluster, and the density and strength of distal connections relative to local ones. Even when all units are exposed (as cluster D illustrates), the connections to and from each individual unit are still predominantly local because the units share the burden of distal communication. Long-range connections are a limited resource (Cherniak et al. 2004) but are critical for integrating the components into a coherent whole. The Leabra components are in constant, high-bandwidth interaction, and parallel constraint satisfaction among them is a fundamental implicit processing mechanism. Hence, we eschew the terms *module* and *encapsulation* in our theorizing. This is a source of creative tension in our (Jilk et al. 2008) collaboration to integrate Leabra with the ACT-R architecture, whose proponents make the opposite emphasis (J. R. Anderson 2007; J. R. Anderson et al. 2004). Much of this tension is defused by the realization that the modularist terminology forces a binary distinction on what is fundamentally a continuum.

NOTES

1. There are exceptions, such as the use of a separate neurotransmitter (dopamine) in the basal ganglia.

2. Event-related designs do not escape this criticism because they too, via multiple regression, track contingent variation around a common mean.

3. *Encapsulation* on the input side is usually distinguished from *inaccessibility* on the output side. We discuss them jointly here because of space limitations. Also, the reciprocal connectivity and the task-driven learning in Leabra blur the input/output distinction.

Neural reuse and human individual differences

doi:10.1017/S0140525X1000107X

Cristina D. Rabaglia and Gary F. Marcus

Department of Psychology, New York University, New York, NY 10003.

rabaglia@nyu.edu

gary.marcus@nyu.edu

Abstract: We find the theory of neural reuse to be highly plausible, and suggest that human individual differences provide an additional line of argument in its favor, focusing on the well-replicated finding of "positive manifold," in which individual differences are highly correlated across domains. We also suggest that the theory of neural reuse may be an important contributor to the phenomenon of positive manifold itself.

Anderson's compelling case for neural reuse is well motivated by empirical results and evolutionary considerations and dovetails nicely with the "descent with modification" perspective put forward by our lab (Marcus 2006; Marcus & Rabagliati 2006). An important additional line of support comes from the study of human individual differences.

In an entirely modular brain, one might predict that individual differences in specific cognitive domains would be largely separate and uncorrelated, but the opposite is in fact true: An extensive literature has shown that performance on separate cognitive tasks tends to be correlated within individuals. This "positive manifold," first noted by Spearman (1904), is arguably one of the most replicated findings in all of psychology (e.g., Deary et al. 2006). At first glance, such correlations might appear to be a statistical by-product of the fact that any individual cognitive task draws on multiple underlying processes. However, even when the impurity of individual tasks is taken into account, using more sophisticated structural equation models that form latent cognitive constructs (representing a cognitive ability, such as short-term memory, by the shared variance among performance on diverse tasks with different specific task demands), clear correlations between cognitive capacities within individuals remain. Positive manifold is not an artifact, but a fact of human cognitive life. (Our point here is reminiscent of Anderson's observation that patterns of co-activation in fMRI remain even after subtraction, and are therefore not attributable solely to mechanistic impurities at the task level.)

These correlations between cognitive domains have now been shown in hundreds of separate data sets, and at many levels, ranging from parts of standardized tests such as SAT math and SAT verbal, to broad ability domains such as memory and spatial visualization (see Carroll 1993), to more specific links