DISCUSSION

LESION STUDIES, SPARED PERFORMANCE, AND COGNITIVE SYSTEMS

Jack C. Lyons

(Department of Philosophy, University of Arkansas)

Lesion studies are often thought to be important because they seem to indicate that the mind contains a number of distinct modules or cognitive systems.

The term 'module' has – to my ear – too many associations with Fodor's (1983) seminal book, and I will concentrate here on the more general notion of a cognitive system. The latter, as I will understand the term, is – roughly – a computational mechanism which can operate independently of all other computational mechanisms (for a much fuller and more precise treatment, see Lyons, 2001). To say that there is a face recognition system, for example, is to say, at least in part, that there is a mechanism which by itself is capable of effecting a transformation from some set of inputs to face identification outputs. If there is one such system, there are likely to be several. Since systems may contain various subsystems, it is generally impossible to specify a system uniquely without specifying a set of inputs. The largest system that would count as a face recognition system would be the one that takes retinal irradiation arrays as inputs and delivers face identifications as outputs, but the last subsystem in this system would map high level representations to face identifications. For any task (where a task is construed as an input/output mapping), take away all cortical regions whose absence does not affect the ability of what is left to perform the task, and you are left with the system that performs that task.

Researchers typically focus on lesion data that display selective impairment. It is possible, however, that the chief significance lies in the *spared* performance in the face of damage to known brain regions. Suppose an idealized case where a patient's face recognition (FR) ability is completely unimpaired, even though the entire left hemisphere is destroyed. Even without evidence of impaired performance or double dissociation, we would know that any computational mechanisms residing in the left hemisphere are not part of the system that performs FR, since this system can operate independently of them. Suppose now that in a second patient visual object recognition for non-face stimuli (VOR) is entirely spared, though the entire right hemisphere is destroyed. Assuming premorbid uniformity across brains (cf. Caramazza, 1986), we would know that there are at least two systems: one which performs (at least) VOR and another which performs (at least) FR, and which are independent of each other. This approach assumes that the locus of a performance-sparing lesion is outside the boundary of the relevant system. It does not follow that the locus of a

Cortex, (2003) 39, 145-147

performance-*impairing* lesion is within the boundary of the relevant system. The more performance-sparing lesions we examine, the more we can narrow down the cortical mechanism of the system and the exact task the system performs (i.e., face recognition, within-category discrimination, or what have you).

It is important to separate the claim that two systems are *distinct* from the claim that they are *disjoint*. In actual fact, FR and VOR systems are likely to overlap quite a bit, sharing a number of early visual subsystems. If so, they are distinct but not disjoint. Cutting very carefully, we could physically remove from the head a functioning FR system or a VOR system, but we would not get one of each from a single brain.

Talk about complete independence introduces an element of idealization, of course, but there is no real problem there. Science is rife with frictionless planes, perfectly elastic spheres, and the like. Galileo's theory of idealized free fall was superior to the Aristotelian theory not because it got the observable phenomena right – it didn't. But Galileo could specify the interfering factors (chiefly air resistance), thus explaining the observations' departure from the idealizations. Actual lesions rarely completely spare performance, but the potentially interfering factors are clear enough: a small part of the system received collateral damage, certain inputs are no longer getting in, the mechanisms mediating behavioral output are damaged, etc. In addition to interfering factors, it is expected that independence will actually come in degrees. Two weakly interacting systems may only approximate independence but nevertheless deserve to be counted as distinct systems.

Focusing on spared performance may be important. There have been some rather convincing connectionist demonstrations of doubly dissociated tasks that do not appear to involve distinct systems (e.g., Plaut, 1995; Juola and Plunkett, 2000). However, it would be hard to convince anyone that a pared down subset of a network can do exactly what the larger network did but is somehow not independent. Thus if lesioning different parts of what seemed like a unitary network results in the sparing of distinct tasks, then that is a better argument for thinking that the network consisted of two partially overlapping systems than for thinking that there is something wrong with the methodology.

This is particularly important given that systems can overlap and that independence can come in degrees. Imagine a connectionist network that represents word meanings via an output vector involving a number of microfeatures, some of which are concrete, and some of which are abstract (a simplification of Plaut and Shallice, 1993). The meanings of concrete nouns will be represented mainly by the activation of the concrete microfeatures and the abstract nouns by the activation of the abstract microfeatures. In the limit, the abstract noun system would be unaffected by the loss of the concrete microfeatures, and vice versa. We would have two distinct systems with a large degree of overlap in the rec of the network. To the extent, however, that some concrete nouns contain abstract microfeatures or vice versa, we would have a lesser degree of independence, but still two fairly distinct systems.

The present criteria for systemhood might seem too weak. What saves this approach from triviality, however, is the fact that, in principle, at least, the degree of independence between various systems can be quantified. Thus, even

if we are left with few interesting questions about *whether* systems s_1 and s_2 are distinct, there can be very interesting question about *how* distinct they are.

References

CARAMAZZA A. On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and Cognition, 5:* 41-66, 1986

FODOR JA. *The Modularity of Mind*. Cambridge, MA: MIT Press, 1983.

JUOLA P and PLUNKETT K. Why double dissociations don't mean much. In G Cohen, RA Johnston and K Plunkett (Eds), *Exploring cognition: Damaged brains and neural networks*. Philadelphia, PA: Psychology Press/Taylor & Francis, 2000, pp. 319-327.

LYONS JC. Carving the mind at its (not necessarily modular) joints. *British Journal for the Philosophy of Science, 52:* 277-302, 2001.

PLAUT DC and SHALLICE T. Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10:* 377-500, 1993.

PLAUT DC. Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17:* 291-321, 1995.

Jack Lyons, Department of Philosophy, University of Arkansas, Fayetteville, AR 72701 USA. e-mail: jclyons@uark.edu