ity. Localist representations increase the number of neurons required for information processing but decrease the average activity of individual neurons. Distributed representations do the opposite. Since baseline synaptic activity represents up to 75% of resting glucose utilization of the brain (Phelps et al. 1979), it is likely that significant reductions in metabolic cost can be obtained by minimizing the number of neurons. Hence efficient distributed representations will minimize metabolic costs.

Page raises two additional objections to this notion of efficiency: comprehensibility and learnability. Presumably both will be addressed in other commentaries so we will limit our response to two brief comments. First, although localist representations are often transparent and therefore can be interpreted by an outside observer much more readily than distributed representations, the important point to remember here is that this is not the purpose of neural representations. Instead, their purpose is to offer the maximal adaptive advantage to an organism. Second, Page claims that learning distributed representations is both inefficient and implausible. However, if McClelland et al. (1995) theory of complementary learning systems is correct, then the metabolic costs of maintaining a hippocampus must be outweighed by the massive reduction in neocortex which it allows. Furthermore, although backpropagation may be biologically implausible, more realistic algorithms do exist (e.g., Hinton et al. 1995). Thus learning distributed representations need not be an insurmountable problem.

In conclusion, we contend that both efficiency and reliability lead one to adopt distributed, not localist, representations. Distributed codes minimize metabolic costs and therefore provide an adaptive advantage to an organism. Let us be clear. We are not suggesting that the brain uses an efficient coding scheme because it is theoretically optimal. Instead, our claim is that evolution has developed schemes to help minimize the metabolic cost of neural computation. This is achieved through the use of sophisticated encoding schemes resulting in the use of distributed representations. Page (sect. 8) claims "that if the brain doesn't use localist representations then evolution has missed an excellent trick." We would like to suggest, however, that if efficiency and reliability are important factors in neural information processing, then distributed, not localist, representations are evolution's best bet.

ACKNOWLEDGMENT

This commentary was supported by a McDonnell-Pew grant and an MRC Program grant to L. K. Tyler.

The case against distributed representations: Lack of evidence

Simon Farrell and Stephan Lewandowsky

Department of Psychology, University of Western Australia, Nedlands, W.A. 6907, Australia. **{simon; lewan}@psy.uwa.edu.au**
www.psy.uwa.edu.au/user/{simon; lewan}

Abstract: We focus on two components of Page's argument in favour of localist representations in connectionist networks: First, we take issue with the claim that localist representations can give rise to generalisation and show that whenever generalisation occurs, distributed representations are involved. Second, we counter the alleged shortcomings of distributed representations and show that their properties are preferable to those of localist approaches.

Page eloquently extolls the virtues of localist representations and their presumed superiority over distributed representations in connectionist networks. We focus on two aspects of the argument: First, we contend that generalisation cannot occur without involvement of distributed representations. Second, we refute six objections levelled against distributed representations.

Localist representations do not generalise. Page identifies a representation as localist if it is "possible to interpret the state of a given node independent of the states of other nodes" (sect. 2.2., para. 7). For example, the representations {0, 1} and {1, 0} for items A and B would be considered localist, whereas {0, 1} and {1, 1} would be considered distributed. Critically, Page advocates a hybrid approach that "supplements the use of distributed representations . . . with the additional use of localist representations" (sect. 1, para. 3). In support, he presents a generic "localist" network that exhibits a number of desirable properties, among them the ability to generalise a learned response to noisy input. Critics have often questioned whether localist representations are capable of generalisation, so its occurrence in a localist network deserves scrutiny.

We contend that the network's ability to generalise arises entirely from the use of *distributed* representations at the input layer which "reflect, in a graded fashion, the degree of similarity that the current input shares with each of those learned patterns" (sect. 4.3.1, para. 2). Localist representations, as defined by Page, are necessarily orthogonal to each other. Hence, the graded similarity that Page identifies as critical for generalisation is inextricably linked to the presence of distributed representations at the input layer.

Although this supports our claim that generalisation requires distributed representations, other research shows that they need not be confined to the input layer. Hinton (1986) presented a multilayer network in which representations at the input layer were strictly localised whereas the hidden layer used distributed representations. The network was found to exhibit meaningful generalization. Subsequent analysis of the activation profiles of the hidden layer confirmed the crucial role of distributed representations.

Distributed representations resist objections. Page attributes six deficiencies to distributed representations (sects. 7.1–7.6), all of which revolve around the overlap of representations at the hidden layer. We counter these objections as follows.

7.1. Catastrophic interference. We concur with Page that interleaved learning, in particular as instantiated by McClelland et al. (1995), is not a preferred solution to catastrophic interference. We also agree that elimination of catastrophic interference requires minimisation of the overlap between representations at the hidden layer. However, it does not follow that localist representations are therefore preferable. First, as alluded to by Page, *distributed* solutions other than interleaved learning exist that reduce catastrophic interference (for a review, see Lewandowsky 1994). Second, localist solutions to the interference problem, as for example provided by ALCOVE (Kruschke 1992), have been shown to engender impaired generalisation (Lewandowsky 1994). By contrast, all available distributed solutions to interference are known to retain their ability to generalise (Lewandowsky 1994).

A careful consideration of catastrophic interference and generalisation therefore points to an advantage of distributed over localist representations.

7.2. Implausibility of the learning rule. This criticism rests entirely on the biologically dubious nature of the gradient-descent algorithm in back-propagation. However, other distributed learning rules, such as Hebbian learning, have been directly supported by biological research (e.g., Kelso et al. 1986). Moreover, at a psychological level, direct empirical support for distributed representations has been provided by the plethora of studies that have confirmed the predictions of the Rescorla-Wagner theory of learning (e.g., Shanks 1991). An essential element of the Rescorla-Wagner theory is that stimuli (e.g., in a categorisation task) are represented by ensembles of attributes or features.

7.3. The dispersion problem. Can distributed representations capture the similarity between sentences such as "John loves Mary" and "Mary loves John"? (sect. 7.3, para. 1). In agreement with Page, we find this question difficult to answer for all possible distributed schemes. However, we note that distributed linguistic parsers have been implemented that address this problem (e.g., Miikkulainen 1996). It follows that distributed schemes are not at a selective disadvantage in handling the dispersion issue.

7.4. Problems deciding "when" and "what." In many distributed networks, a response is identified by some extraneous process "done by the modeller rather than by the model" (sect. 7.4, para. 2). Page correctly identifies this as a serious problem. However, the solution to the problem need not be localist. Distributed networks that can unambiguously identify a response, without any extraneous mechanism or any of the other objections raised by Page, have been presented by Lewandowsky (1999), Lewandowsky and Farrell (in press), and Lewandowsky and Li (1994).

7.5. Problems of manipulation. Contrary to the claim in the target article, response suppression can demonstrably be accomplished in a distributed network using (Hebbian) "anti-learning" (e.g., Lewandowsky, in press; Lewandowsky & Li 1994). Page is correct in assuming that other items might be affected to the extent that they are similar to the suppressed target, but there is no evidence that this does not occur empirically. Indeed, this suppression of "neighbours" might explain why similar list items suffer more during serial recall than dissimilar ones.

7.6. Problems of interpretation. We agree that distributed models are more difficult to interpret than those with localist representations. This is because distributed models, unlike localist schemes, are capable of restructuring the input in interesting and novel ways that may at first glance escape interpretation.

Consider the distributed network presented by Hinton (1986). The network learned a set of input-output patterns whose semantic structure was not captured by the localist input and output representations. Through supervised learning alone, the network was found to organise its hidden layer into a distributed representation that captured the underlying semantic structure. While it required some analysis to visualise that distributed representation, the very fact that it was not immediately obvious implies that the network learned something novel and interesting.

Conclusion. We concur with Smolensky (1990) that representation is "crucial . . . , for a poor representation will often doom the model to failure, and an excessively generous representation may essentially solve the problem in advance" (p. 161). Unlike Page, we do not believe that localist representations are inherently preferable to distributed approaches. The alleged flaws of distributed schemes cited by Page are in fact desirable properties.

Why localist connectionist models are inadequate for categorization

Robert M. French and Elizabeth Thomas

Psychology Department (B32), University of Liège, 4000 Liège, Belgium; Institut Léon Frédéricq, University of Liège, 4000 Liège, Belgium. **{rfrench; ethomas}@ulg.ac.be** **www.fapse.ulg.ac.be/Lab/cogsci/rfrench.html**

Abstract: Two categorization arguments pose particular problems for localist connectionist models. The internal representations of localist networks do not reflect the variability within categories in the environment, whereas networks with distributed internal representations do reflect this essential feature of categories. We provide a real biological example of perceptual categorization in the monkey that seems to require population coding (i.e., distributed internal representations).

Despite Page's bold frontal assault on distributed connectionism, we wish to point out what appear to us to be two significant problems with this type of localist network.

The problem of category variability. Consider two categories, "fork" and "chair." The variability within the first category is very low: there just aren't that many different kinds of forks. Chairs, on the other hand, come in all different shapes, sizes and materials: they range from beanbag chairs to barstools, from overstuffed armchairs to rattan chairs, from plastic lawn chairs to that paragon of ergonomic design, the backless computer chair that you kneel on; some have four feet, some three, some none; some have backs, some don't; some are made of metal, others plastic, others wood, others, cloth and Styrofoam pellets, and so on. In other words, the variability within the category of *chair* is enormous.

But in the localist model proposed by Page, and in localist models in general, *this information about category variability is lost.* In distributed models, it takes more hidden nodes to encode a category with high-variability than one with low variability. In other words, the internal representations reflect external category variability. However, the category nodes in localist networks are unable to reflect this differential variability-in-the-environment of various categories. The one-node internal "representation" corresponding to the extremely low-variability category "fork" is precisely the same as the one-node internal representation corresponding to the highly variable category "chair."

Why is this a problem? Most significantly, because of the well-documented fact of category-specific losses: in general, naming of inanimate objects is found to be better preserved than naming of animate objects (Farah et al. 1996; Funnell & Sheridan 1992; Warrington & Shallice 1984). A model with distributed internal representations can handle this problem quite simply: low-variance categories (e.g., many natural kinds categories, like *cat, horse,* etc.) are encoded over fewer "units" than high-variance categories (e.g., many artificial kinds categories, like *chair, tool,* etc.) Random lesioning of the model will be more likely, on average, to destroy the representation of a category with low-variability (e.g., natural kinds categories) that is coded over a small number of units than a high-variability category (e.g., artificial kinds categories) coded over a large number of units. Localist models in which all the category nodes are the same will have considerable problems explaining category-specific deficits of this kind, especially when the featural inputs to the internal category representations remains intact. If, on the other hand, we assume differing degrees of variance associated with the internal encoding of different categories, these kinds of deficits can be predicted in a straightforward manner, as French (1997b) and French and Mareschal (1998) have shown using a dual-network architecture based on the hippocampal-neocortical separation proposed by McClelland et al. (1995).

As Page points out in his target article, we have argued for the necessity of "semi-distributed" representations in connectionist models for many years. But "semi-distributed" does not mean localist. "Semi-distributed" representations preserve category variance information; localist representations do not. Further, it seems crucial to us that these semi-distributed representations *emerge* as a result of learning.

Biological category representations. Page is right in pointing out that some of what is called population or ensemble coding in biological systems can be viewed as localist. For example, even though broadly tuned, cells of the motor cortex have their maximum activity tuned to a particular direction (Georgopoulos et al. 1993). One should therefore be able to ascertain the direction being represented by looking at the activity of individual neurons (or very small groups of neurons). However, an example of a cognitively relevant task that cannot be achieved in this fashion can be found in the anterior temporal cortex. Vogels (1999) reports on the responses of cells in this area during a tree, non-tree categorization task by a monkey. Most of the cells were stimulus selective, (i.e., they did not respond to all of the presented stimuli) and responded to both trees and non-trees. The maximum response of these neurons was not tuned to either category. Even though it was the case that certain (category-selective) neurons responded to particular subsets of tree exemplars, *no individual neuron (or small set of neurons) responded to all of the presented trees,* while not responding to any non-tree. These category-selective neurons alone did not appear to play an important role in the categorization performance of the monkey (Thomas et al. 1999). In other words, population coding was necessary for the monkey to correctly categorize all exemplars in the test set.