Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: http://www.tandfonline.com/loi/plcp21

Preface

Jeffrey S. Bowers

To cite this article: Jeffrey S. Bowers (2017) Preface, Language, Cognition and Neuroscience, 32:3, 255-256, DOI: 10.1080/23273798.2016.1266370

To link to this article: http://dx.doi.org/10.1080/23273798.2016.1266370

Published online: 27 Jan 2017.

LANGUAGE, COGNITION AND NEUROSCIENCE, 2017
VOL. 32, NO. 3, 255–256
http://dx.doi.org/10.1080/23273798.2016.1266370

Routledge
Taylor & Francis Group

PREFACE

Within psychology there is a long history of connectionist models that include *localist representations*. That is, models that contain single units that represent meaningful "things" (such as letters, words, objects, and faces). Perhaps the classic example is the interactive activation (IA) model of visual word identification by McClelland and Rumelhart (1981): words are represented with individual units, and word identification occurs when a given unit is activated beyond some threshold. By contrast, in connectionist models that include distributed representations, often called *Parallel Distributed Processing* models, meaningful things are coded as a pattern of activation across many units, and each unit is involved in representing multiple things.

Over the past 30 years localist and distributed models have been advanced across a wide range of domains, including visual and spoken word identification, short-term memory, episodic memory, semantic memory, object perception, face perception, motor control, etc. This distinction is fundamental because the two classes of models represent, process, and learn information in qualitatively different ways, and as a consequence, they often make different predictions about behaviour within a given domain. Accordingly, within psychology, the primary evidence for localist vs. distributed representations is based on comparing the relative successes of these two classes of models in explaining human behaviour.

Within neuroscience a similar issue has been discussed, although it has played a smaller role in shaping theory. In this case, the question is whether single neurons represent specific meaningful things (grandmother cells) or meaningful things are coded by many neurons and each neuron represents many things (distributed representations). The primary evidence for and against grandmother cells comes from single-cell recording studies that assess how selectively neurons respond to various forms of inputs (e.g. faces, objects, words, etc.).

For over 50 years neuroscientists have been reporting cases of single neurons responding highly selectively to meaningful things, from oriented lines in primary visual cortex to images of faces in inferotemporal cortex and hippocampus. Nevertheless, grandmother cells are generally dismissed for a variety of reasons. Perhaps most critically, research has shown that many (millions) of neurons fire in response to a given input, and all neurons appear to respond to a range of different things. Indeed, even when a neuron in a given experiment responds to only one out of many images tested, it is assumed (quite plausibly) that it would fire to some other (untested) images (e.g. Waydo, Kraskov, Quiroga, Fried, & Koch, 2006). Accordingly, the most striking examples of selective neural responding are often taken as evidence for highly sparse and selective coding rather than for grandmother cells.

Does the neuroscience settle this issue? To many the answer is clear: both localist representations and grandmother cells are considered biologically implausible. However, a number of theoretical issues need to be addressed in psychology and neuroscience before any strong conclusions are warranted (Bowers, 2009). Let me highlight two examples.

First, the rejection of grandmother cells should not be taken as evidence against localist representations within psychology because neuroscientists mean something quite different by grandmother cells. Grandmother cells are generally characterised as neurons that respond to only one thing, with no firing above baseline to other things (e.g. Gross, 2002). By contrast, localist representations do not share this attribute. Consider the IA model of visual word identification (McClelland & Rumelhart, 1981). When a given word is input to the model many units become active across the feature, letter, and word levels of the network, and critically, a given word unit responds to multiple words. For instance, the localist unit for DOG is activated most strongly to the input <DOG> but it is also activated (to a lesser extent) by the visually similar words <HOG> and <LOG> by virtue of their shared letters. This unit does not behave like a grandmother cell, and nevertheless, it is the prototypical instantiation of a localist representation. Accordingly, there is no reason to reject localist models based on the observation that even highly selective neurons respond to a small number of related inputs.

Second, although various results from neuroscience rule out some versions of grandmother theories, there has been little consideration of whether the brain computes with localist representations. For example, do single neurons sometimes represent one specific thing by becoming active beyond some threshold while at the same time becoming partially active in response to similar things (like a localist word representations in the IA)? As noted above, the theoretical distinction between localist and distributed coding is critical, but there has been almost no consideration of the relative biological plausibility of these two approaches in neuroscience. The common rejection of grandmother cells has contributed to a failure to fully engage with many key issues in psychology and neuroscience, such as: Are some types of problems best solved with localist representations and others with sparse distributed representations, and still others with highly distributed representations? Are there computational limitations of distributed coding schemes that are overcome with localist models (or vice versa)?

My hope is that this special issue will provide a collection of papers from psychology and neuroscience that will advance our understanding of how information is coded in mind and brain. A key step in doing this is to avoid confusions that lead researchers from different fields from talking past one another, and to ensure that insights from psychology and neuroscience are shared.

References

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, *116*, 220–251.

Gross, C. G. (2002). Genealogy of the "grandmother cell". *The Neuroscientist*, *8*, 512–518.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407.

Waydo, S., Kraskov, A., Quiroga, R. Q., Fried, I., & Koch, C. (2006). Sparse representation in the human medial temporal lobe. *The Journal of Neuroscience*, *26*, 10232–10234.

Jeffrey S. Bowers
j.bowers@bristol.ac.uk