How many Neurons for your 'Grandmother' ? Three Arguments for Localised Representations

Mayor, Julien; Plunkett, Kim

How many Neurons for your 'Grandmother' ?
Three Arguments for *Localised* Representations

Julien Mayor (julien.mayor@unige.ch)
FPSE, University of Geneva
1211 Genève 4, Switzerland

Kim Plunkett (kim.plunkett@psy.ox.ac.uk)
Department of Experimental Psychology, University of Oxford
Oxford OX1 3UD, United Kingdom

Abstract

In a recent article, Bowers (2009) argues that local representations are more consistent with neuro-biological data than distributed representations, as typically generated in Parallel Distributed Processing (PDP) models. We present three reasons why *localised* neural representations are good candidates for supporting mental representations, as they provide a solution to the trade-off between combinatorial arguments that favour fully-distributed representations and metabolic arguments which favour localist representations.

Keywords: distributed representations, local representations, self-organising maps, synaptic pruning, brain metabolism

Introduction

Over the last thirty years, hypotheses concerning the nature of mental representations have essentially been polarised to two interpretations: some researchers argue that brain representations are distributed (among them, proponents of the Parallel Distributed Processing (PDP) approach: e.g., Rumelhart, McClelland, & the PDP Research Group, 1986; Seidenberg & McClelland, 1989; McClelland & Rogers, 2003; Plaut & McClelland, 2010), while others suggest that local representations fit neuro-physiological data more accurately (e.g., Page, 2001; Bowers, 2009). Most of the arguments in favour of distributed representations fall into one of the following two categories: high combinatorial power and robustness with respect to lesions. In contrast, Bowers (2009) reviews neurobiological evidence for relatively sharply tuned neurons reminiscent of localist representations and argues that distributed approaches fail to provide unambiguous representations under superposition.

We attempt to clarify the role of combinatorial power, in light of the superposition problem. We then introduce two metabolic arguments to the debate. We suggest that a potential solution to the debate relies on *localised* representations, capitalising on robust representations that span only a limited number of neurons, thereby minimising the energy expenditure associated with mental representations. Finally, we discuss the implications of this proposal and highlight examples already using localised representations.

The combinatorial argument

The idea that distributed representations can code many more patterns than localist coding scheme is well established. Traditional binary coding, in which a neuron is either active or silent, emphasises this difference. We will therefore reiterate this combinatorial argument using binary activation levels and comment on its validity in the context of decoding superposed patterns. The extension to continuous encoding will then be discussed.

The case of binary encoding

The coding advantage Elementary calculus shows that 2^n patterns can be encoded over n neurons, corresponding to the case of fully-distributed representations (see Fig. 1). With localised representations, the combinatorial power decreases rapidly. Suppose, for example, that each pattern can use at most n neurons out of a total system of N neurons. The number of patterns that can be stored in n neurons is then 2^n in each of the subset of n neurons picked from the total pool of neurons. If $n = N/2$, for example, there are two subsets of n neurons, each coding 2^n patterns. The total number of patterns p with a degree of localisation of n stored in $N = 2n$ neurons would then be $p = 2 \cdot 2^n$. Fig. 1 depicts the number of patterns p that can be stored among N neurons (maximum 20 neurons in the simulation), for different levels of localisation n. A purely localist encoding (n=1) can only store as many patterns as there are neurons. At the other end of the spectrum, fully-distributed representations can store 2^n patterns. In between, the number of patterns one can store is directly related to the number of neurons that are involved in the coding of an individual pattern. As a consequence, localist representations have a limited capacity to store only as many separate representations as the number of neurons. Orders of magnitude can be gained by coding each patterns over a few neurons. With only 30 neurons, a fully-distributed approach would be able to store more than a billion different representations, a number that exceeds by orders of magnitude the likely capacity for human mental representations: "even if the distinguishable visual items are larger than the number of the different types of objects (< 100000) that humans are able to discriminate, cortical visual neurons are certainly so numerous that there would be enough sets of them to represent each single object (or property)"(Pareti & De Palma, 2004, p.45). On the other hand, localised encoding ($n > 1$) can rapidly reach the combinatorial power required to represent a very large number of different representations.

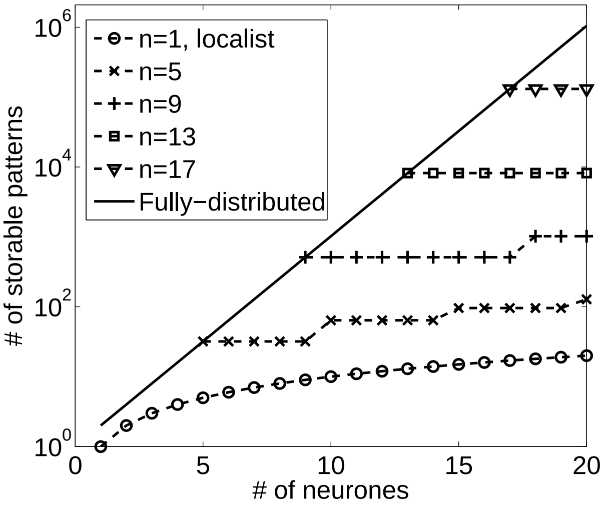

Figure 1: Number of patterns that can be stored as a function of the total number of neurons, using binary encoding. The different curves correspond respectively to a purely localist encoding, different levels of localisation and a fully-distributed encoding.

The decoding problem Encoding many representations is a necessary requirement for human cognitive performance. In many situations, multiple representations need to be encoded simultaneously. This leads to the potential ambiguity introduced by a superposition of representations. For example, visual scenes will contain a configuration of independent objects. A difficult task for the brain is therefore to be able to decode superposed representations unambiguously. Patterns of activation in the brain can be construed as multi-dimensional vectors. Elementary linear algebra constrains the number of linearly independent vectors to the number of dimensions represented in the system. Any additional vector can be expressed as a linear combination of other vectors.Consequently, there can be at most N independent patterns unambiguously coded across N neurons, on the assumption that each neuron encodes a separate dimension. Beyond that limit, additional representations can be misconstrued as a superposition of one or more other representations *even for fully-distributed representations*. If the network's task is to encode multiple patterns that can be superposed in the same neural substrate, the combinatorial advantage of distributed representations is compromised.

The case of continuous encoding

Activation levels in neurons do not need to be restricted to a binary coding scheme. For example, simple rate coding models make use of a range of activation levels to encode different stimuli. More complex, and more realistic models of neurons make use of the rich dynamics of neuronal firing. A strict and complex mathematical analysis of the coding and decoding capacities in both local and distributed representations for continuous coding schemes is beyond the scope of the present article. Nevertheless, it is worth commenting some implications of this approach.

A first observation is that coding is further enriched by the increased range of values any neuron can take. In fact, a single neuron could encode as many different patterns as needed, as long as the decoder has a resolution that is fine enough. The combinatorial advantage of distributed coding is still present for a decoder with a fixed resolution (e.g., the ability to detect subtle differences between relevant, and different, neural activation levels). However, as single neurons can encode many more patterns with continuous encoding schemes than with binary coding, fewer neurons are required to encode the same number of patterns. For example, 10 binary coding neurons are needed to represent 1000 patterns ($2^{10} = 1024$), but if continuous activation levels can be detected with greater accuracy so that each neuron could take 10 distinct values each, only 3 neurons would then be required ($10^3 = 1000$).

A second observation is that decoding subtle differences between different activation levels of neurons is a non-trivial problem. Presence of noise would limit the decoder's resolution and the more neurons needed to encode a representation, the more difficult it will be to decode that information and the more neurons required to act as decoders (e.g., see Földiák (2003) for a discussion of the advantage of representations that span fewer neurons (sparse representations) than fully-distributed representations for decoding).

The limited resolution of the decoder effectively reduces the case of continuous encoding to a simple extension of binary coding, where each neuron can have two distinct activation levels, to a case of N-coding, in which each neuron can take N distinct values. Small values for N magnify the problem of superposition of representations for distributed representations while reducing the problem of combinatorial limitations for localised representations. Large values for N would furthermore undermine the claim that localised (or even localist) representations cannot encode a sufficiently large number of different representations, while increasing the complexity and vulnerability of a decoder network that requires an increasingly large number of neurons.

The metabolic argument

Let us turn now onto a consideration that can be made independently from the nature of the neural coding itself. It is often claimed that the resource needed to encode N patterns is less using fully-distributed representations than localist codes, thereby minimising metabolic expenses, because fewer neurons are required for distributed representations.

However, consider a brain structure with a given number of neurons required to represent a number of different patterns. The metabolic expense of the brain structure is, to a first approximation, proportional to the number of neurons that participate in the representation of the pattern(s) present in a scene (or maintained/sustained in that brain structure). As a rule of thumb, if less neurons are required to participate in the representation of a pattern, the less the energy required for that task.

Fig. 2 depicts the energy consumption (as indexed by the number of neurons that participate in the representation) as a function of the number of patterns (or objects) that need being represented in a network of 20 neurons. Different curves correspond to different localisation levels: in a purely localist coding, each pattern is represented by a single neuron. In this case, the energy consumption is proportional to the number of objects represented simultaneously. As the number of neurons involved in the representation of a given pattern increases, the energy expenditure increase for any number of patterns (or objects) being represented at a given time. Ultimately, fully-distributed representations require all neurons to participate in the representation of even a single pattern. Unless the system is working at full capacity at all times, energy consumption is minimised for localist representations but maximised for fully-distributed representations. Localised representations consume intermediate levels of energy.

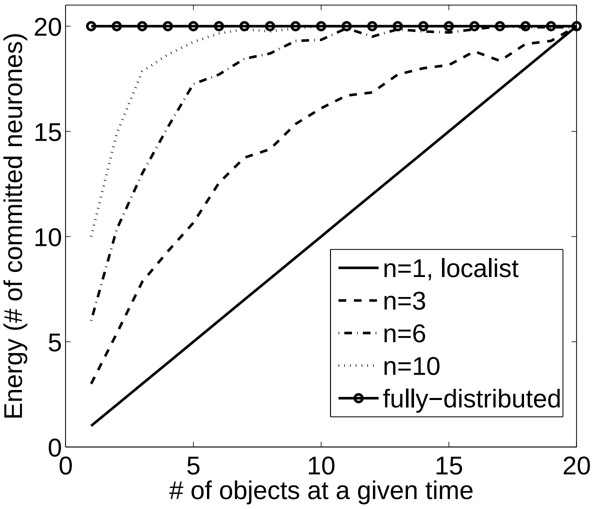

Figure 2: Energy as indexed by the number of neurons required to represent a number of objects simultaneously. The different curves correspond respectively to a purely localist encoding, different levels of localisation and a fully-distributed encoding. Localised representations use less energy than fully-distributed representations for a given architecture (in the present simulation, total number of neurons=20).

The synaptic pruning argument

Neural resources involve not only the cell bodies of neurons but also the connections between them. Associations between representations require appropriate connections or synapses. Synaptic maintenance is also a contributor of energy consumption. For example, neural mappings between an object representation and its corresponding label require appropriate cross-modal synapses between visual and auditory areas. The number of cross-modal synapses required to form the mapping between the different brain structures depends on the degree of localisation of each representation in both structures. Figure 3 depicts the number of cross-modal synapses needed to maintain an appropriate mapping between representations in different neural structures, as a function of the degree of localisation of the representations in each structure (which have been chosen to be identical for the sake of simplicity). The number of synapses needed increases with the number of objects that are encoded in each modality. Note that fully-distributed representations (where the degree of localisation equals the number of neurons, 20) require the full set of cross-modal synapses ($20^2 = 400$) even when each structure is only required to encode a single object. A lower number of synapses can only be achieved for reduced levels of localisation for the neural representations.

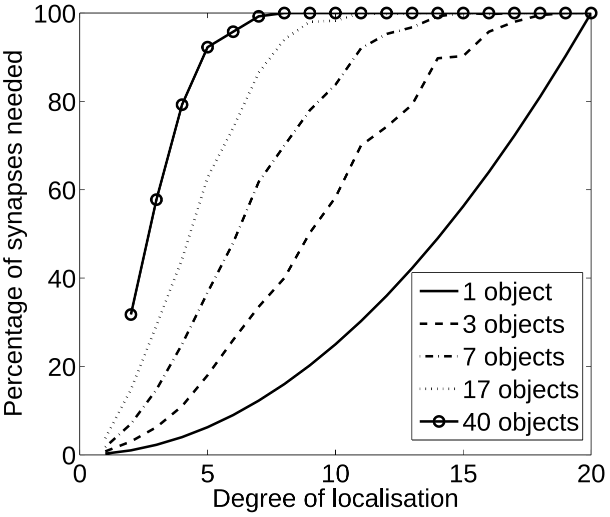

Figure 3: Percentage of cross-modal synapses required to maintain the mappings between uni-modal representations without degradation, as a function of the degree of localisation. The different curves correspond to different network loads in terms of the number of objects that need to be represented in each modality. Note that pruning can only be achieved with reduced levels of localisation.

It is important to note that the number of synapses is not constant during brain development. After an early proliferation of synapses, their number remains approximately constant, before environmentally induced synaptic pruning reduces the total number of synapses (see Huttenlocher, 2002). The observed synaptic pruning mechanism is usually associated with either an improvement in cognitive skills (Miller, Keller, & Stryker, 1989; Chechik, 1998) and/or an optimisation in metabolic resources (Roland, 1993; Feinberg, Thode Jr, Chugani, & March, 1990), potentionally leading to a internally-driven reorganisation of neural representations. So that synaptic pruning can operate without being detrimental to the task, representations benefit from a reduced degree of localisation.

Discussion: Solutions to the trade-offs

Mental representations can have different levels of localisation, as defined by the number of neurons that are required to take part in the representation of an individual pattern. Many constraints may impact this degree of localisation, such as the number of different patterns a neural structure can code, the capacity to decode a superposition of patterns, robustness arguments and metabolic constraints at the level of neurons and synapses. The exact solution to the trade-off between these constraints remains elusive. While there are many advantages in having more than one neuron involved in the representation of a given pattern (robustness, combinatorics), there are at least as many in restraining the number of neurons taking part in a representation; metabolic minimisation, synaptic pruning, simpler decoding. We suggest that computational approaches to neuroscience and psychology may need to adopt the perspective that patterns are represented in a localised fashion; not localist (only one neuron per pattern) nor fully-distributed. The degree of localisation can, of course, be modulated according to the structure under consideration or, if the function is highly abstract, according to the task.

Self-Organising Maps (SOMs) offer an approach in which degree of localisation is discovered from exposure to the input structure (Kohonen, 1984). SOMs form topographically organised maps of neurons, such that neighbouring neurons respond to similar input. The resources on the map dedicated to a particular pattern or category is determined by many factors such as the number of different patterns that a SOM has been exposed to, the number of neurons on the SOM, the frequency with which a given category of patterns is presented, and the magnitude of the pattern variations in each category. After exposure to a structured environment, SOMs display a partitioned map from a representational perspective: each pattern creates a unique pattern of *localised* neural activity and each category of patterns would tend to solicit the same group of neighbouring neurons in order to represent different patterns that belong to the same category.

The organisation of a SOM after learning mimics cortical maps observed throughout many different cortical areas. SOMs have been very successful at modelling the architecture of the primary visual cortex (Miikkulainen, Bednar, Choe, & Sirosh, 2005) where neighbouring neurons are responsive to similar orientations of the visual scene (Hubel & Wiesel, 1962). Topologically-organised maps have also been found in the human auditory cortex (Romani, Williamson, & Kaufman, 1975; Pantev et al., 1995), in the human frontal and prefrontal cortex (Hagler & Sereno, 2006) and in parietal cortex (Sereno & Huang, 2006).

Beyond mimicking the neuro-anatomical organisation of cortical maps, SOMs sustain representations that possess interesting properties from a psychological perspective. For example, categories are formed in an unsupervised way, similarly to infant's capacities to form categories in the absence of supervision (Younger, 1985) and discrepancies between a pattern and its representation provide an accurate index of looking behaviour of young infants during categorisation tasks (Gliozzi, Mayor, Hu, & Plunkett, 2009). When input pattern possess a family resemblance structure (e.g., basic level categories of objects, Rosch & Mervis, 1975), representations on the SOM are warped in a manner that mimic categorical perception (Mayor & Plunkett, 2010). Since a single pattern activates a localised pattern of neural activity on the map, only a limited number of neurons contribute to the pattern representation. However, when the number of patterns represented exceeds the number of neurons on the map, a single neuron must participate in the representation of multiple patterns from the same category. Consequently, some neurons are maximally active when an average of a few patterns is presented to the map. This provides a representation advantage for central tendencies, thereby implementing at a representational level the advantage of prototypes over atypical members of a category (Rosch, 1973; Mervis, 1984). Interestingly, the fact that multiple neurons contribute to the representation of different members of the same category of patterns maintains a sensitivity to within category variations, as observed in speech perception (McMurray, Tanenhaus, & Aslin, 2002).

Mayor and Plunkett (2010) have also evaluated the impact of synaptic pruning in a model of early word learning, consisting of two SOMs connected by cross-modal Hebbian synapses. Synaptic pruning was shown to enhance the quality of word-object mappings, once stable representations of objects and labels were achieved on the maps. The localised representations of individual objects and labels permitted high levels of pruning so as to associate objects categories and their corresponding labels in a one-to-one mapping. Synaptic pruning of any one-to-one mapping between cortical representations (or thalamo-cortical projections) would also benefit from such localised representations. In contrast, high levels of pruning would be detrimental to highly distributed representations. The presence of high levels of synaptic pruning from mid-childhood would seem to favour the formation of these relatively localised mental representations.

It is noteworthy that any representations requiring a relatively small number of neurons also satisfy the conditions for metabolic constraints and synaptic pruning. However, these constraints do not require that neurons supporting the representation of a given pattern need to be neighbours. Examples of sparse coding (Quiroga, Kreiman, Koch, & Fried, 2008; Quiroga & Kreiman, 2010), in which only a small subset of neurons is active for a pattern have been shown offer decoding advantages (Földiák, 2003) as well as minimising metabolic demand. SOMs offer sparse coding in which the few neurons taking part in the representation of a pattern are proximate, thereby providing additional advantages in terms of mimicking cortical maps that are found across the human brain (Hubel & Wiesel, 1962; Romani et al., 1975; Pantev et al., 1995; Hagler & Sereno, 2006; Sereno & Huang, 2006) and constraining the need for long distance connections (Durbin & Mitchinson, 1990). SOMs may also provide a potential advantage in terms of decoding the information, as representations of different patterns that belong to the same category tend to be similar. As a consequence, SOMs, localised representations in general, should lead to enhanced robustness in the presence of noise.

Conclusion

The resources needed for mental representation are constrained by many different, and often opposing, pressures. A solution to the trade-off between robustness and combinatorial power, which favour representations with many neurons, and metabolic and synaptic pruning constraints, which favour fewer neurons, is to limit the number of neurons needed to represent a pattern. Sparse, localised representations provide an elegant alternative to purely localist representations and fully-distributed ones. Self-Organising Maps provide a natural, and unsupervised, approach for forming localised representations which mimic cortical maps found throughout the human cortex. The topographical structure of these SOMs also permit efficient pruning mechanisms to operate, maximising metabolic efficiency and providing accurate models of human cognitive performance and development.

Acknowledgments

This work is supported by the Swiss National Science Foundation grant 131700 awarded to Julien Mayor and by the Economic and Social Research Council Grant RES-062-23-0194 awarded to Kim Plunkett.

References

Bowers, J. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, *116*(1), 220–251.

Chechik, G. (1998). Synaptic Pruning In Development: A Computational Account. *Neural Computation*, *10*(7), 1759–1777.

Durbin, R., & Mitchinson, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, *343*, 644-647.

Feinberg, I., Thode Jr, H., Chugani, H., & March, J. (1990). Gamma distribution model describes maturational curves for delta wave amplitude, cortical metabolic rate and synaptic density. *Journal of Theoretical Biology*, *142*(2), 149–61.

Földiák, P. (2003). Sparse coding in the primate cortex. In M. Arbib (Ed.), *The handbook of brain theory and neural networks.* MIT Press, Cambridge, MA.

Gliozzi, V., Mayor, J., Hu, J.-F., & Plunkett, K. (2009). Labels as features (not names) for infant categorisation: A neuro-computational approach. *Cognitive Science*, *33*(4), 709–738.

Hagler, D., & Sereno, M. (2006). Spatial maps in frontal and prefrontal cortex. *Neuroimage*, *29*(2), 567–577.

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*(1), 106–154.

Huttenlocher, P. (2002). *Neural Plasticity: The Effects of Environment on the Development of the Cerebral Cortex.* Harvard University Press.

Kohonen, T. (1984). *Self-organization and associative memory.* Berlin: Springer.

Mayor, J., & Plunkett, K. (2010). A neuro-computational model of taxonomic responding and fast mapping in early word learning. *Psychological Review*, *117*(1), 1–31.

McClelland, J. L., & Rogers, T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310–322.

McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), 33–42.

Mervis, C. (1984). Early lexical development: The contributions of mother and child. In C. Sophian (Ed.), *Origins of cognitive skills.* Hillsdale, N.J.: Lawrence Erlbaum.

Miikkulainen, R., Bednar, J., Choe, Y., & Sirosh, J. (2005). *Computational Maps In The Visual Cortex.* Springer.

Miller, K., Keller, J., & Stryker, M. P. (1989). Ocular dominance and column development: analysis and simulation. *Science*, *245*, 605–615.

Page, M. (2001). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *23*(04), 443–467.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., et al. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and clinical Neurophysiology*, *94*(1), 26–40.

Pareti, G., & De Palma, A. (2004). Does the brain oscillate? The dispute on neuronal synchronization. *Neurological Sciences*, *25*(2), 41–47.

Plaut, D., & McClelland, J. (2010). *Psychological Review*, *117*(1), 284-288.

Quiroga, R., & Kreiman, G. (2010). Measuring sparseness in the brain. *Psychological Review*, *117*(1), 291-297.

Quiroga, R., Kreiman, G., Koch, C., & Fried, I. (2008). *Trends in Cognitive Sciences*, *12*(3), 87–91.

Roland, P. (1993). *Brain activation.* Wiley-Liss New York.

Romani, G., Williamson, S., & Kaufman, L. (1975). Tonotopic organization of the human auditory cortex. *Psychiatry*, *132*, 650.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.

Rosch, E., & Mervis, C. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1: Foundations). Cambridge, Massachusetts: The MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Sereno, M., & Huang, R. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. *Nature neuroscience*, *9*(10), 1337.

Younger, B. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*, 1574-1583.