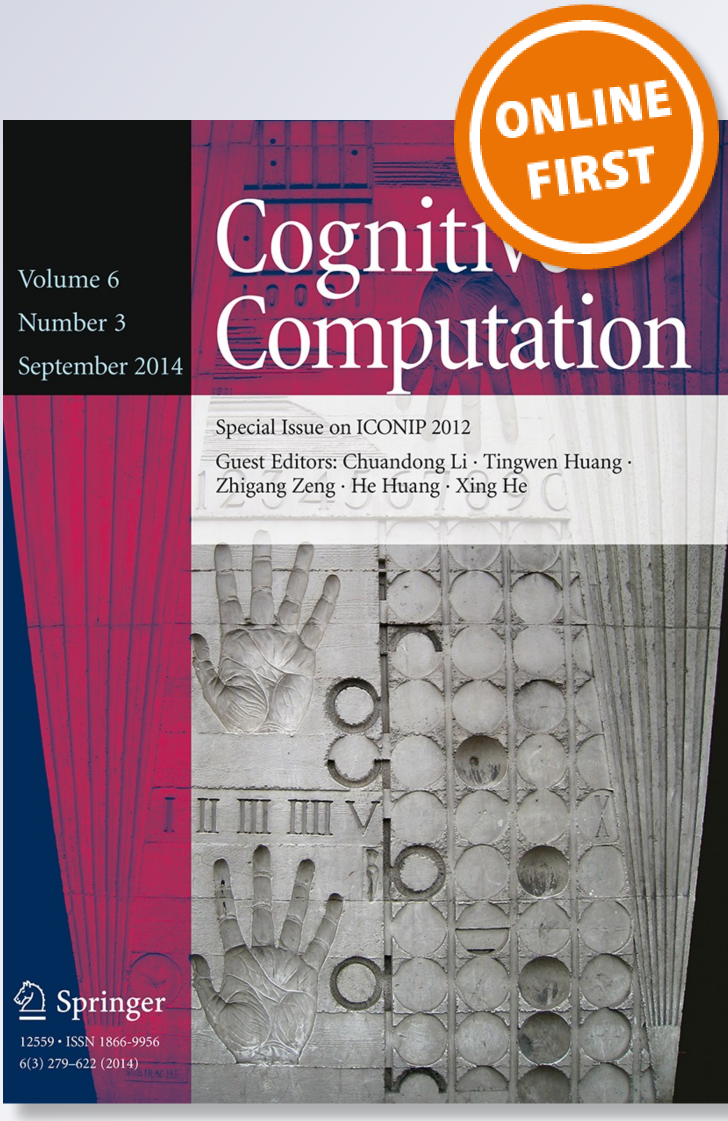
On Findings of Category and Other Concept Cells in the Brain: Some Theoretical Perspectives on Mental Representation

Asim Roy

Cognitive Computation

ISSN 1866-9956

Cogn Comput
DOI 10.1007/s12559-014-9307-7



Springer

Cogn Comput
DOI 10.1007/s12559-014-9307-7

On Findings of Category and Other Concept Cells in the Brain: Some Theoretical Perspectives on Mental Representation

Asim Roy

Received: 8 September 2014 / Accepted: 1 October 2014

Abstract There is substantial neurophysiological evidence from decades of single-cell studies that category and other concept cells exist in both human and animal brains. This indicates that the brain can generalize and create abstract concepts and encode and represent such abstractions using single cells. These single-cell findings cannot be accounted for and explained by the theory of distributed representation and population coding, the dominant theory in the brain sciences. In light of these findings, this paper reexamines the two contending mental representational schemes, localist and distributed, on the basis of computational efficiency, the ability to simultaneously process and activate many different concepts, and the structure for semantic cognition. The evidence for category and concept cells favors localist representation in the brain.

Keywords Localist representation · Distributed representation · Semantic cognition · Category cells · Concept cells

Introduction

Single-cell recordings of recent years, in both animals and humans, have revealed the existence of abstract category and concept cells in the brain. The existence of abstract place cells has been known for decades. More recent findings reveal the existence of different types of face cells (e.g., monkey faces and human faces), cells that reflect emotions, cells for categories of objects (e.g., nests, animals, and houses), multimodal invariant concept cells for persons (e.g., Jennifer Aniston and Saddam Hussein), and objects (e.g., Sydney Opera House). However, the most widely accepted theory of the brain, the theory of population coding or distributed representation, cannot explain the existence of these types of abstract cells—why are they there, what is their purpose, and so on. Their existence is obviously related to mental representational issues. This paper offers some theoretical perspectives on the existence of these types of cells in the brain. Representational issues obviously are the focus, and the two primary theories—local and distributed representation (population coding)—are compared. The first perspective is from the point of view of processing and representational efficiency and efficiency of access to information for decision making and problem-solving. The second perspective is from the point of view of parallel processing and simultaneous activation of many different concepts. The third perspective is from the point of view of semantic cognition and the role of these category and concept cells.

The paper is laid out as follows. Section 2 cites some of the findings on category and concept cells. Section 3 presents a localist semantic cognition model that is widely referenced and its corresponding distributed representation form. Section 4 deals with computational and processing efficiencies of the two representational schemes. Section 5 is about the ability to simultaneously activate many different concepts. Section 6 deals with semantic cognition and the implementation of a localist semantic cognition model using Hebbian cell assemblies and abstract categories and concepts. Section 7 has the conclusions.

A. Roy (✉)
Department of Information Systems, Arizona State University, Tempe, AZ 85287, USA
e-mail: asim.roy@asu.edu

The Evidence for Abstract Category and Concept Cells in the Brain

In general, there is significant evidence that the brain can create abstract multimodal invariant representations of objects and persons and encode them in single cells. Quian Quiroga et al. [1] report finding single medial temporal lobe (MTL) neurons in humans that encode object-related concepts irrespective of how the objects are presented—visual, textual, or sound. They checked the modality invariance properties of a neuron by showing the subjects different pictures of the particular individual or object that a neuron responds to and their spoken and written names. For example, Quian Quiroga et al. [1, p. 1308] found a neuron in the entorhinal cortex of a subject that responded "*selectively to pictures of Saddam Hussein as well as to the text 'Saddam Hussein' and his name pronounced by the computer… There were no responses to other pictures, texts, or sounds.*" Quian Quiroga [2, p. 588) found a hippocampal neuron that responded to pictures of Halle Berry, an actress, even when she was masked as a "*Catwoman.*" The neuron also responded to the letter string "HALLE BERRY" (and not to other names) and her name pronounced by a synthesized voice. Suthana and Fried [3, p. 428] found a neuron that responded to pictures of the Sydney Opera House, but not to 50 other landmarks. It also responded to "*many permutations and physically different representations of the Sydney Opera House, seen in color, in black and white, or from different angles*" and to the written words "*Sydney Opera.*"

Quian Quiroga et al. [1] found that a large proportion of MTL neurons responded to both pictures and written names of particular individuals or objects and speculates that "*MTL neurons encode an abstract representation of the concept triggered by the stimulus.*" These abstract cells are often called concept cells and are further discussed in [4, 5].

The Evidence for Category Cells

Cells that represent abstract categories have been found in both humans and animals. Fried et al. [6] and Kreiman et al. [7] led the discovery of "category-specific" neurons in the MTL in humans. Fried et al. [6] found single neurons in the MTL that discriminated faces from inanimate objects and also found ones that responded selectively to specific emotional expressions or jointly to particular facial expression and gender. Kreiman et al. [7] found single MTL neurons that responded selectively to "*visual stimuli from different categories, including faces, natural scenes and houses, famous people and animals.*" Kawasaki et al. [8] found neurons in the left and right orbital and anterior cingulate cortices of humans, which were selective for only one emotion class, most often aversive.

Single neurons in the monkey visual temporal cortex have been found to respond selectively to certain categories of stimuli such as faces or objects [9–11]. Gothard et al. [12] found single neurons in the amygdala of monkeys that responded selectively to images of monkey faces, human faces, and objects. Their general observation is (p. 1674): "*These examples illustrate the remarkable selectivity of some neurons in the amygdala for broad categories of stimuli.*" Lin et al. [13] found "nest cells" in the mouse hippocampus that fire selectively when the mouse observes a nest or a bed, regardless of the location or environment. Yoshida et al. [14] found many neurons in the dorsoposterior part of rat anterior piriform cortex that are tuned selectively to either a single category or a specific combination of distinct categories of a panel of eight food-related categories of odorants. Sugase et al. [15] studied single neurons in the temporal cortex of macaque monkeys and found that they conveyed different types of information at different latencies. Initial response of the neurons categorized the stimulus as monkey faces, human faces, or other shapes. A later response of the same neurons contained more detailed information about identity and expression.

Freedman et al. [11] summarized these findings in the following way (p. 312): "*These studies have revealed that the activity of single neurons, particularly those in the prefrontal and posterior parietal cortices (PPCs), can encode the category membership, or meaning, of visual stimuli that the monkeys had learned to group into arbitrary categories.*"

Overall, the neurophysiological evidence is substantial that category cells exist in the brain and that the brain can abstract and generalize and use single cells to represent these abstractions.

Semantic Cognition Based on a Hierarchy of Abstract Categories and Concepts

Semantic cognition implies our ability to infer properties of objects and concepts when we encounter them in some form. For example, when we see a bird on the ground, we generally infer in our mind that it has wings, it can fly, it is a living thing, and so on. How such semantic knowledge is created, stored, and accessed in our brains has been of immense interest for decades. Figure 1 shows a possible way of storing semantic knowledge, as proposed in [16], where semantics are based on a hierarchy of abstract category concepts and their properties.

In this tree structure, nodes represent abstract categories and arrows reflect properties of a category. For example, the node *bird* has arrows for the properties *feathers*, *fly,* and *wings*. The semantic tree shows the hierarchical relationship of these abstract categories. For example, *plant* and *animal* are subcategories of *living thing*. The hierarchical tree of Fig. 1 produces propositions such as *living things grow*; *a plant is a living thing*; *a tree is a plant*; and *an oak is a tree*. It therefore follows that *an oak can grow*.

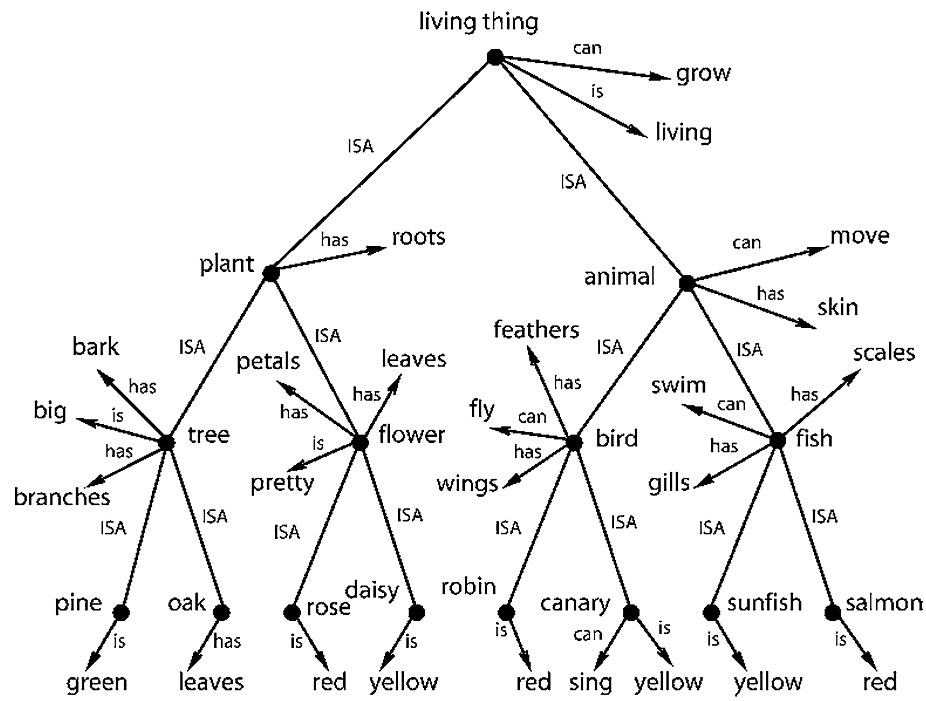


Fig. 1 A taxonomic hierarchy of the type used by Collins and Quillian [16]. Adapted from Fig. 2 in "Précis of Semantic cognition: a parallel distributed processing approach," by Rogers and McClelland [19]. Copyright Cambridge University Press

A Distributed Representation/Population Coding View of Semantic Cognition

PDP connectionists [17–19] and population coding theory deny the existence of abstract concept cells in the brain. That is, they deny that single cells could encode and represent such abstract concepts as *living thing*, *plant*, *animal*, and so on. Hence, in their models [17–19], they restructure this type of semantic tree by replacing the complex abstract units with units of the distributed form. Therefore, concept nodes—such as *robin, living thing*, *plant*, and *animal*—are replaced with string of nodes for distributed representation.

Figure 2 is the Rumelhart-type network model used by Rogers and McClelland [18, 19] to learn the semantic knowledge of Fig. 1. The network of Fig. 2 learns three-element propositions of Fig. 1 that are true of objects such as *pine* and *canary*. This particular model uses localist representation for objects and concepts, but [18, 19] later use perceptual features of objects to explain the concept of distributed representation. The *Item* and *Relation* layer nodes of Fig. 2 are the input nodes, and the *Attribute* layer nodes are the output nodes. The hidden layer nodes have nonlinear processing functions. The *Item* layer nodes represent objects, whereas *Relation* layer nodes represent contexts. For example, the input pair *canary* and *can* is analogous to showing the network a picture of a *canary* and asking what it *can* do.

In the following quote, Rogers and McClelland [18, p. 147, 148] explain how a distributed representation model is created from the semantic model of Fig. 2, which is a localist model.

"The distributed version of the model uses the same corpus of training items and a similar model architecture. However, in place of twenty-one *Item* units employed in the localist version (with one unit standing for each item), we instead construe the input units as representing the subset of each object's attributes that are apparent from its visual appearance—for example, features such as *red*, *big*, and *legs*…The model might be shown a picture of a robin, for instance, by activating input units corresponding to *red*, *legs*, and *wings*….To this end, we employed as 'perceptual' input attributes seven of the eight *is* properties from the training corpus—excluding *living*, but including *big, pretty, green, red, yellow, white*, and *twirly*—as well as a subset of the *has* properties: specifically, *wings, legs, gills, petals*, and *branches*. Note that these properties are also included as output attributes, along with *is living*, the remaining *has* properties, and the *can* properties for all twenty-one instances."

Note, however, that the distributed representation model proposed above is still based on localist units (i.e., it uses localist units such as *big, pretty, green,* and *red*) and is essentially used to explain the concept of distributed representation in a simple way. Their ultimate distributed model, however, is made of units that have no meaning and labels at all, neither at the input layer nor at the output layer.

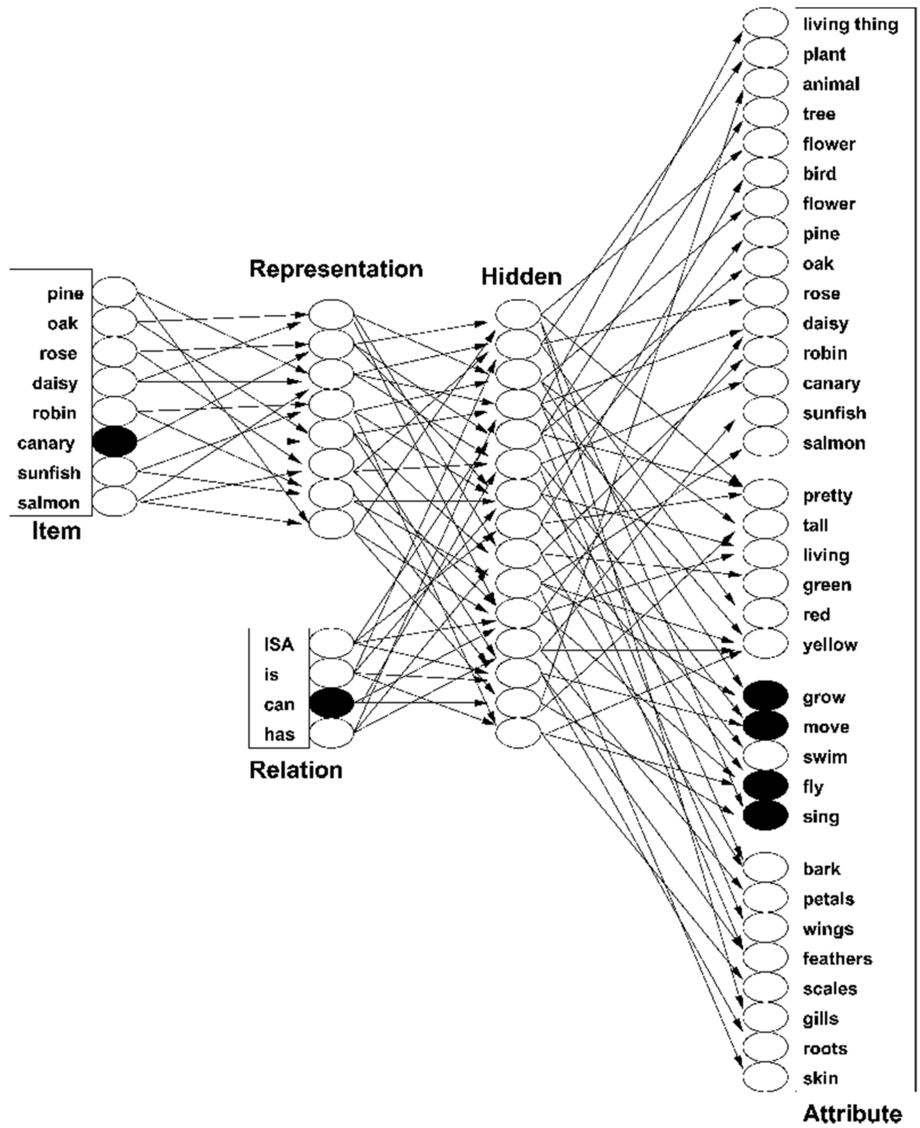


Fig. 2 Rumelhart-type network model used by Rogers and McClelland [18, 19] to learn propositions of Fig. 1

A Theoretical Perspective: On Efficient Access to Information by a Supervisory System

Abstract category and concept cells correspond to localist representation and here's one explanation why such cells exist in the brain.

Given that a supervisory system is the ultimate recipient of category and concept information, any type of system, biological or artificial, would try to provide that information in the most compact and readily usable form. Imagine a supervisory system being informed by a single cell that the visual stimulus corresponds to a monkey face versus being presented with a distributed pattern over a few thousand cells. In the case of a distributed pattern, the supervisory system would have to interpret the pattern first and that adds a layer of computation irrespective of the way it is performed. Such a pattern interpretation step would also take time and slow down processing. On the other hand, an abstract cell for a monkey face relieves the supervisory system of such a computational burden and speeds up processing. There is, therefore, significant computational efficiency with localist representation especially for frequent and repetitive tasks.

Extracting, abstracting, and summarizing information in additional cells, such as concept and category cells, streamlines computations, and interpretations, avoids redundant computations, and creates speed and efficiency in the system. Hence, computational efficiency of the overall system is one of the main reasons for the existence of category and concept cells and the constant abstraction of information in various stages of processing in the brain.

A Theoretical Perspective: On the Need for Simultaneous Activation of Many Different Concepts

Suppose that the human brain has a distributed representation system that is trained to recognize different categories of animals. Now, imagine a person looking at an advertisement for a zoo at a distant billboard that shows a tiger, a lion, a giraffe, a rhinoceros, and an elephant, all in close proximity to each other and within the person's span of attention [20–22]. A distributed representation system would have to process each animal separately, one after the other, in order to recognize them. Such a system, therefore, is not capable of recognizing all of the animals simultaneously, even when all of them are in focus. In this situation, the distributed representation system essentially becomes a sequential (serial) processor, processing one animal at a time. And parts of the visual stimulus would have to be in a queue and wait for processing. A localist representation system, on the other hand, is not bound by the strict limits of such a single-structure recognition system (i.e., the distributed representation system) and can create separate parallel recognition systems for each category of animal, very similar to parallel processing in the brain for color and motion detection. It can segment the stimulus (for each animal) and process the segments in parallel through the multiple recognition systems and activate the corresponding category cells simultaneously. One can, therefore, build efficient parallel processing systems by using localist representation. In general, simultaneous activation of many different concepts, a phenomenon observed in the brain, is possible only with localist representation and that is one of the reasons we find category and concept cells in the brain, which are localist cells.

Distributed representations problem with simultaneous activation was pointed out earlier by Garagnani et al. [23, p. 161]: "*In this approach, the same set of hidden nodes is used to encode different items as different patterns of graded activation; this, however, makes it impossible to maintain separate different item representations when these are simultaneously active. In general, cognitive arguments (e.g. our proven ability to maintain multiple item representations distinct) favour localist representations, whereas neuroscience arguments weight in favour of distributedness.*"

A Theoretical Perspective: Category and Concept Cells and Semantic Cognition in the Brain

The existence of category and concept cells might also explain how semantic cognition works in the brain and vice versa. The abstraction-based semantic model of Fig. 1 appears realistic in many ways given the neurophysiological evidence. First, there is evidence for hierarchical categories in the brain. Second, a Hebbian cell assembly can easily implement such a hierarchical semantic cognition system.

The Evidence for Category Hierarchies

If one examines the variety of category cells found in the brain, one would observe that some of the cells reflect very broad categories, whereas others are not so broad. For example, there are face cells, but also cells for subcategories like human and monkey faces. There are also cells for specific types of emotions, for gender, and for different types of odors and their combinations. There are object-related cells, but also cells for specific types of objects like houses and nests. From these neurophysiological observations, one can infer that the semantic cognition system in the brain is hierarchical, very much in the sense of Fig. 1.

A Hebbian Cell Assembly Implementation of Semantic Knowledge Using Category and Concept Cells

Quian Quiroga [2] presents a model where abstract concept cells are part of Hebbian cell assemblies. That model could be extended to include category cells to build a semantic system that is somewhat similar to Fig. 1. Thus, there could be a cell assembly for each node of the hierarchical semantic tree of Fig. 1. For each such node, a cell assembly would consist of a cell for the category and separate cells for the properties of that category. For example, for the node *bird* in Fig. 1, the cell assembly would have a cell for the category *bird* and additional concept cells for the properties *fly*, *feathers,* and *wings*. Similarly, there could be a cell assembly for the category *fish* and concept cells in that assembly for the properties *scales*, *swim,* and *gills*. Since a category and its properties are related, the firing of a category cell, say the *bird* cell, would cause the property cells to fire in the Hebbian sense. In this case, the cells for *fly*, *feathers,* and *wings* would activate.

The cell assemblies could also be linked in a hierarchical fashion since categories at different levels of the tree are related. Thus, the *bird* cell in the *bird* assembly and the *animal* cell in the *animal* assembly are related, and the *bird* cell activation would activate the *animal* cell and its property cells. In this way, a hierarchy of cell assemblies can produce a semantic system similar to the semantic tree of Fig. 1. The concatenation of such cell assemblies and the appropriate activation of categories and their properties provide the supervisory system with relevant semantic information.

There is obviously an efficiency aspect to such a semantic cognition system: to the supervisory system, it provides easy and quick access to semantic information through just a few cells instead of through thousands or millions of cells in a population code-based system. Thus, simplification, automation, and computational efficiency are the key advantages of a semantic system based on category and concept cells. Plus, a Hebbian cell assembly system provides simultaneous activation of appropriate concepts, something that is not feasible with a distributed representation system. Thus, category and concept cells in the brain solve many theoretical problems and that explains their existence in the brain.

Conclusions

This paper tried to explain the existence of category and other concept cells in the brain by presenting possible theoretical justifications for them. The overarching considerations were that of computational efficiency and simultaneous processing and delivery of information to a supervisory system. A distributed representation (population coding) system, being a single-structure system built to represent many different things, has the drawback of becoming a serial processor when many different things have to be recognized at the same time.

Note that category and concept cells tie well into a semantic cognition theory, and such a theoretical connection has not been made before in cognitive or neurosciences.

One can infer from the analysis here that the most efficient, easily accessible, and compact form of semantic knowledge exists in the brain in a network of abstract concept and category cells, and that this semantic network is used by the supervisory system in the brain for decision making and problem-solving.

References

1. Quian Quiroga R, Kraskov A, Koch C, Fried I. Explicit encoding of multimodal percepts by single neurons in the human brain. Curr Biol. 2009;19:1308–13.
2. Quian Quiroga R. Concept cells: the building blocks of declarative memory functions. Nat Rev Neurosci. 2012;13(8):587–97.
3. Suthana N, Fried I. Percepts to recollections: insights from single neuron recordings in the human brain. Trends in cognitive sciences. 2012;16(8):427–36.
4. Roy A. A theory of the brain: localist representation is used widely in the brain. Front Psychol. 2012;3:551. doi:10.3389/fpsyg.2012.00551.
5. Roy A. An extension of the localist representation theory: grandmother cells are also widely used in the brain. Front Psychol. 2013;4:300. doi:10.3389/fpsyg.2013.00300.
6. Fried I, McDonald K, Wilson C. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. Neuron. 1997;18:753–65.
7. Kreiman G, Koch C, Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci. 2000;3:946–53.
8. Kawasaki H, Adolphs R, Oya H, Kovach C, Damasio H, Kaufman O, Howard Iii M. Analysis of single-unit responses to emotional scenes in human ventromedial prefrontal cortex. J Cogn Neurosci. 2005;17(10):1509–18.
9. Gross CG. How inferior temporal cortex became a visual area. Cereb Cortex. 1994;5:455–69.
10. Logothetis N, Sheinberg D. Visual object recognition. Annu Rev Neurosci. 1996;19:577–621.
11. Freedman D, Miller E. Neural mechanisms of visual categorization: insights from neurophysiology. Neurosci Biobehav Rev. 2008;32:311–29.
12. Gothard KM, Battaglia FP, Erickson CA, Spitler KM, Amaral DG. Neural responses to facial expression and face identity in the monkey amygdala. J Neurophysiol. 2007;97:1671–83.
13. Lin LN, Chen GF, Kuang H, Wang D, Tsien JZ. Neural encoding of the concept of nest in the mouse brain. Proc Natl Acad Sci USA. 2007;104:6066–71.
14. Yoshida I, Mori K. Odorant category profile selectivity of olfactory cortex neurons. J Neurosci. 2007;27(34):9105–14.
15. Sugase Y, Yamane S, Ueno S, Kawano K. Global and fine information coded by single neurons in the temporal visual cortex. Nature. 1999;400(6747):869–73.
16. Collins A, Quillian M. Retrieval time from semantic memory. J Verba Learn Verbal Behav. 1969;8:240–7.
17. Rumelhart D, Todd P. Learning and connectionist representations. In: Meyer Kornblum, editor. Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience. Massachusetts: MIT Press; 1993. p. 3–30.
18. Rogers T, McClelland J. A parallel distributed processing approach. Semantic cognition: MIT Press; 2004.
19. Rogers T, McClelland J. Precis of semantic cognition: a parallel distributed processing approach. Behav Brain Sci. 2008;31:689–749.
20. Kootstra G, de Boer B, Schomaker LR. Predicting eye fixations on complex visua stimuli using local symmetry. Cogn Comput. 2011;3(1):223–40.
21. Foulsham T, Underwood G. If visual saliency predicts search, then why? Evidence fro normal and gaze-contingent search tasks in natural scenes. Cogn Comput. 2011;3(1):48–63.
22. Wischnewski M, Belardinelli A, Schneider WX, Steil JJ. Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. Cogn Comput. 2010;2(4):326–43.
23. Garagnani M, Wennekers T, Pulvermüller F. Recruitment and consolidation of cell assemblies for words by way of Hebbian learning and competition in a multi-layer neural network. Cogn Comput. 2009;1(2):160–76.