

Published in: J.A. Bullinaria
& W. Lowe (Eds), *Connectionist Models of Cognition and Perception*. Singapore: World Scientific, pp. 40-50.

THE ROLE OF PERCEPTION AND ACTION IN OBJECT CATEGORISATION

A. BORGHI

*Department of Psychology, University of Bologna, 5 Viale Berti Pichat, 40127 Bologna
Italy
e-mail: borghi@psibo.unibo.it*

A. DI FERDINANDO

*University of Rome "La Sapienza" and Institute of Psychology, National Research Council,
15 Viale Marx, 00137 Rome
Italy
e-mail: andread@ip.rm.cnr.it*

D. PARISI

*Institute of Psychology, National Research Council, 15 Viale Marx, 00137 Rome
Italy
e-mail: parisi@ip.rm.cnr.it*

This work moves from the general hypothesis that action influences knowledge formation, and that the way we organise our knowledge reflects action patterns [7]. The traditional assumption in the categorisation literature is that categories are organised on the basis of perceptual similarity among their members. But much evidence shows that, when we need to perform an action, we can group objects which are perceptually dissimilar. Many studies have shown that we are able to flexibly organise and create new categories of objects on the basis of more or less contingent goals [2,3].

We present some simulations in which neural networks are trained using a genetic algorithm to move a 2-segment arm and press one of two buttons in response to each of 4 stimuli. The neural networks are required to group the stimuli, by pressing the same button, in 2 categories which, depending on the particular task (which is encoded in a set of additional input units), may be formed by perceptually very similar, moderately similar, or different objects.

We find that task information overrides perceptual information, that is, the internal representations of neural networks tend to reflect the current task and not the perceptual similarity between the objects. However, neural networks tend to form action-based categories more easily (e.g. in fewer generations) when perception and action are congruent (perceptually similar objects must be responded to by pressing the same button) than when they are not congruent (perceptually similar objects must be responded to by pressing different buttons). We also find that at hidden layers nearer the sensory input, where task information still has not arrived, internal representations continue to reflect perceptual information.

1 Introduction

One traditional assumption in the categorisation literature is that categories are organised on the basis of perceptual similarity among their members. Perceptually

similar objects are put in the same category, and perceptual dissimilar objects are put in different categories. However, the role of similarity has been criticised and it has been argued that similarity is just a post-hoc notion, as things can be similar for many different aspects [12]. Thus it is necessary to have a principle, or at least a point of view, in order to categorise objects. Objects become similar only after choosing some particular principle or point of view.

The present work is more radical in criticising the notion that categories are based on perceptual similarity. It moves from the general hypothesis that it is action, not perceptual similarity, which determines categorisation and more generally knowledge formation. Ultimately, the way we organise our knowledge reflects the actions with which we respond to perceived objects [7], not the intrinsic perceptual characteristics of objects.

The evidence shows that, when we need to perform an action, we may group together objects which are perceptually dissimilar. Many studies have demonstrated that we are able to flexibly organise and create new categories of objects on the basis of more or less contingent goals [2,3]. Consider for example the category 'things to take to the camping place': it can include tents, shoes, and toothpaste, i.e., objects which are physically dissimilar. Moreover, this category is not a stable one and can be recreated on line: for example, its members can vary depending on the particular place where the camping place is located.

Assigning a relevant role to action does not necessarily imply a reduction of the importance of perceptual similarity for category formation [15,16,17,8]. But actions and goals remain the originary basis for category formation. The reason why perceptual aspects are important for categorisation is that generally there is a correlation between perception and action: similar objects tend to elicit common motor responses [20]. It is plausible, then, that we have evolved the capacity, in absence of a specific task indication, to react in the same way to perceptually similar objects [7,19].

In this paper we address these questions using neural network simulations. Our simulations are aimed at providing a better understanding of the role played by perception and action in categorisation. An artificial organism performs 3 different tasks which require to group 4 objects into 2 categories respectively formed (a) of perceptually similar objects, (b) of moderately similar objects, (c) of objects which are not similar at all. Thus there can be congruence, partial congruence, and total incongruence between perceptual information and the task to perform. We predict that the task is acquired earlier when perceptual and task information are congruent and that it is acquired later when the task requires to group objects which are perceptually dissimilar.

We also predict that the role played by perceptual information and by the task varies in the different layers of the neural networks of organisms. More specifically, using neural networks with three layers of hidden units, we predict that in the layer which is closer to sensory input the internal representations of the objects reflect the perceptual properties of the objects, in the layer which is closer to the output the

internal representations of the objects reflect the task to perform, and in the intermediate layer there is conflict or compromise between perceptual and task information. From the point of view of the nervous system, we might say that information concerning objects independently from action will be located in the sensory areas, while information on objects and actions to perform on them should be located closer to the motor areas [10,11,9].

2 The simulation set-up

A population of 100 artificial organisms live and behave in an environment. The environment contains four different objects but at any given time each organism sees only one of the four objects. The organism has a single arm composed by two segments. The arm sends proprioceptive information to the organism specifying the arm's current position. The behaviour of the organism is controlled by a nervous system, which is simulated with a neural network (see Figure 1).

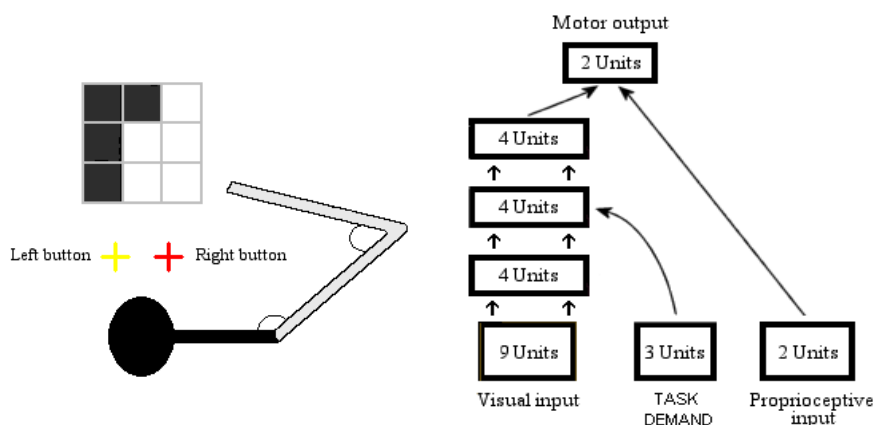


Figure 1. The organism and the network architecture.

The network architecture consists of 5 layers: one input layer with 3 different groups of units, three layers of 4 hidden units each, and one output layer of 2 units. In the input layer 9 units encode the perceptual properties of the objects, 3 units encode the task, and 2 proprioceptive input units encode the current angles between the shoulder and the arm and between the arm and the forearm. The 2 output units encode the actions performed by the organism: namely, they encode the variation of the previously described angles.

As shown in Figure 1, the visual input units are connected with the first of the 3 layers of hidden units, the task input units are connected with the second hidden layer, while the proprioceptive input units are directly connected with the output

units. We use this architecture in order to analyse the role played by perceptual similarity and task demand in the categorisation of the four objects.

Graphically the objects are represented by a 3x3 matrix in which 4 of the 9 slots are filled. Objects A-B and C-D are perceptually more similar as they have 3 slots in common; objects A-C and B-D have some physical similarity as they have 1 common slot; objects A-D and B-C are perceptually dissimilar as they have no common slots (see Figure 2).

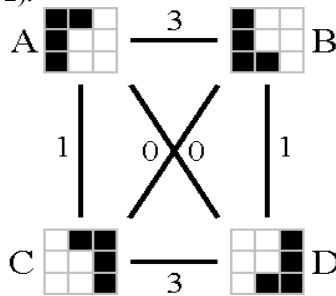


Figure 2. The four objects. The numbers indicate the common slots between pairs of objects.

The organisms are required to group the objects into 2 categories by pressing two different buttons: the left button indicates the first category, the right button indicates the second category. The first task is performed correctly when the 2 categories are composed by perceptually similar objects, i.e., by objects with 3 common slots (A-B; C-D), the second when the 2 categories are composed by objects with 1 common slot (A-C; B-D), and the third one when the 2 categories are composed by perceptually dissimilar objects, i.e., objects with no common slots (A-D; B-C) (see Figure 3).

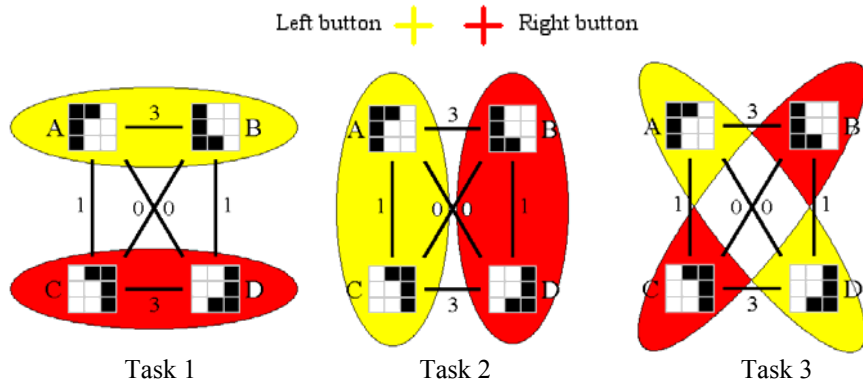


Figure 3. The four objects grouped differentially depending on the task.

To find the connection weights which allow the organisms to perform correctly the three tasks we used a genetic algorithm. The use of a genetic algorithm is dictated by

the difficulty to find a good learning algorithm for ecological simulations but of course there is no implication that the ability to classify perceived objects is genetically inherited rather than learned during life. We first assigned random connection weights to the neural networks of an initial population of 100 individuals. Then we tested each of these 100 individuals on 20 randomly selected trials. In each trial an individual starts with a randomly chosen arm's position and sees one of the four objects. At the end of the 20 trials we assign each individual a fitness value reflecting the individual's ability in performing the task (pressing the correct button). Each individual has a genotype encoding the individual's connection weights. We used a direct one-to-one mapping: each gene encodes a different connection weight as a real number. The 20 best individuals are selected for (nonsexual) reproduction and each of them generates 5 offspring inheriting their parent's genotype with the addition of some random mutations (mutation rate = 10%). The $20 \times 5 = 100$ individuals thus obtained represent the new generation. The process is repeated for a sufficient number of generations so that at the end the organisms are able to execute the three tasks.

3 Data analysis and results

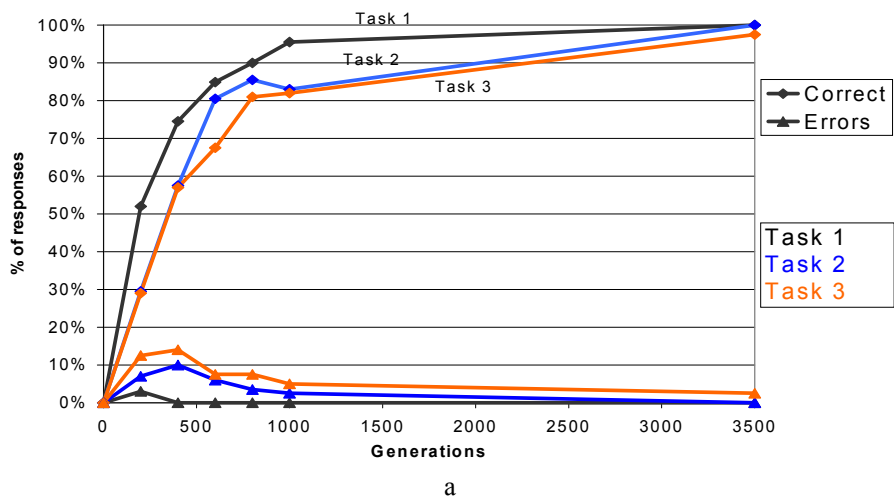
We repeated the simulation 10 times, starting with different sets of initial connection weights. All the results presented are the average of the 10 replications, each of which lasted 3500 generations.

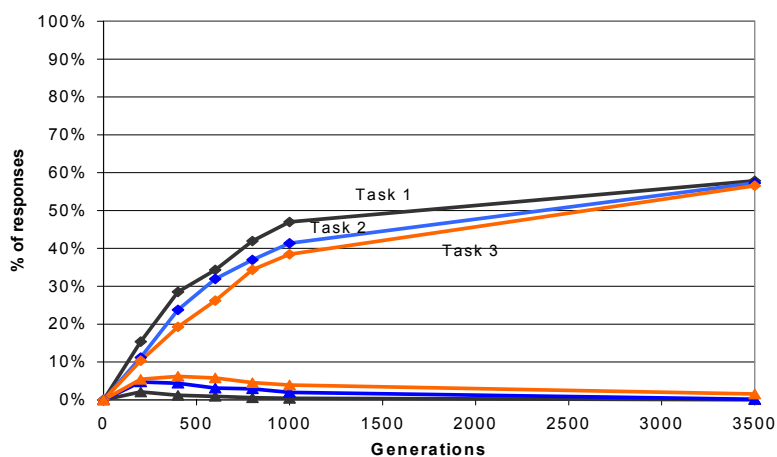
In order to study the role of perception and action in object categorisation, we performed two different analyses. In the first analysis, we calculated the percentage of correct responses and of errors (trials in which the organism presses the wrong button) in performing the three tasks. At the end of the simulation (generation 3500) the performance was about the same in all the three tasks. The best individual responded correctly almost 100% of the times, while the population average was around 60%.

However, comparing the performance of the best individual and of the population average in the three tasks during the first 1000 generations, we found that performance in Task 1 was better than performance in Task 2, and performance in Task 2 was better than performance in Task 3. In other words, Task 1 is learned somewhat earlier (in terms of generations) than Task 2, and Task 2 is learned earlier than Task 3 (see Figure 4). This indicates that when there is congruence between perceptual similarity and categorisation task, the performance is facilitated. In other words, the ability to perform the task is influenced by the perceptual similarity between the objects.

The second analysis we performed concerns the way objects are represented inside the neural network of the best organisms. Each object causes some particular activation pattern in the 4 hidden units of each hidden layer, and these internal representations of the objects can be considered as points in a 4-dimensional space.

We calculated the Euclidean distance between the points which represent the objects in each layer of hidden units. The results are shown in Figures 5-9.





b

Figure 4. Percentage of correct responses and errors for the three tasks across 3500 generations (a: best individual; b: population average).

Figure 5 shows that in the first hidden layer the points which represent perceptually similar objects (A-B; C-D) are the closest points, while the points which represent perceptually dissimilar objects (A-D; B-C) are the most distant ones. This result can be explained by the fact that, in this layer, the task information still has not arrived, so the activation patterns of these hidden units reflect the visual input and the distance between them is determined by the perceptual similarity between the objects.

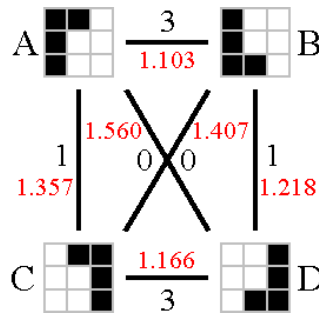


Figure 5. First hidden layer: Euclidean distances between the four points representing the activation patterns of the four objects.

In the second hidden layer, however, when the task information has arrived, the smallest distances are generally those between the points which represent objects which require the same response (Figure 6). In other words, at this level task information overrides perceptual similarity. However, perceptual similarity also has an effect since we observe some differences among the tasks.

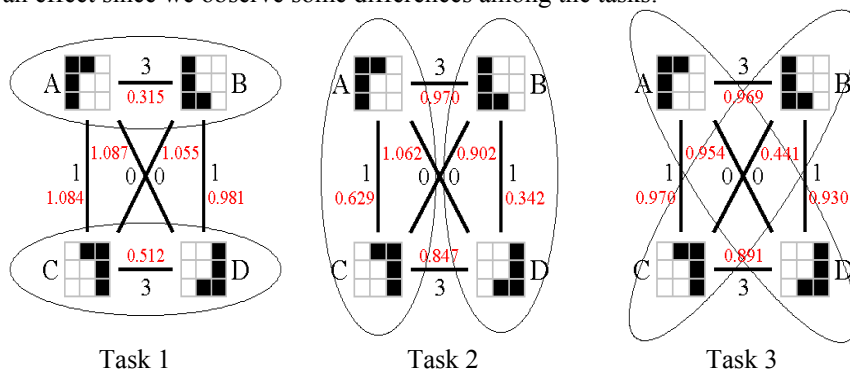


Figure 6. Second hidden layer: Euclidean distances between the points representing the activation pattern of the four objects in the three tasks.

In order to better understand the respective roles played by perception and task demand in the second hidden layer, we calculated the distances between the activation patterns of objects belonging to the same vs different categories in the three tasks. Figure 7 shows that task information is more important than perceptual similarity, as the distance between the activation patterns of the objects of the same category is lower than the distance between the activation patterns of objects of different categories. However, perceptual similarity has an effect on categorisation, in that the difference between same-category distances and different-category distances is highest for Task 1 (congruence between perceptual similarity and task), while it is lowest for Task 3 (no congruence between perceptual similarity and task).

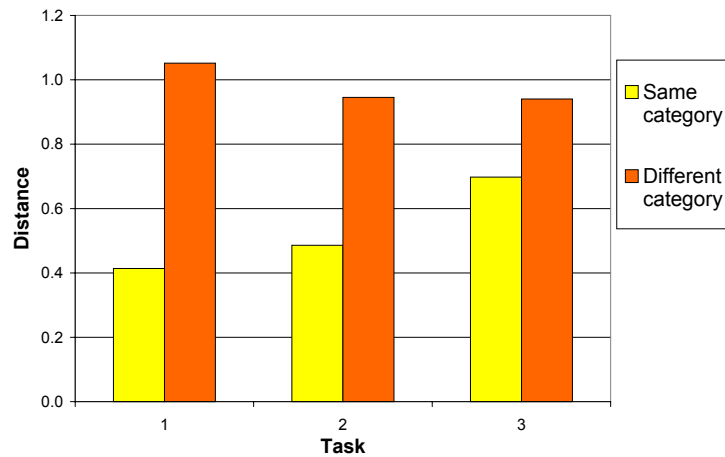


Figure 7. Second hidden layer: Euclidean distances between the points representing the activation patterns of objects belonging to the same vs different categories in the three tasks.

We now turn to the third hidden layer. As it can be seen in Figure 8, in the third hidden layer the smallest distances are always those between points which represent objects which require the same response. At this later level of neural processing, task information is clearly more important than perceptual similarity; furthermore, unlike what happens in the second hidden layer, in the third hidden layer perceptual similarity has no effect. In fact, the difference between same-category distances and different-category distances increases consistently in comparison with the second hidden layer, and there is no sensitivity to the task (see Figure 9).

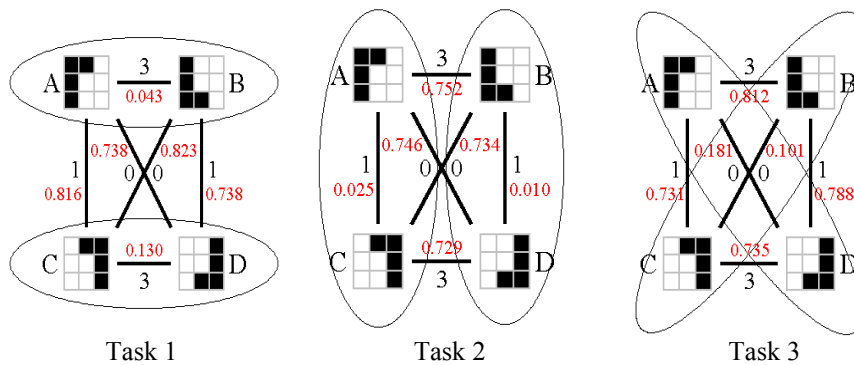


Figure 8. Third hidden layer: Euclidean distances between the points representing the activation pattern of the four objects in the three tasks.

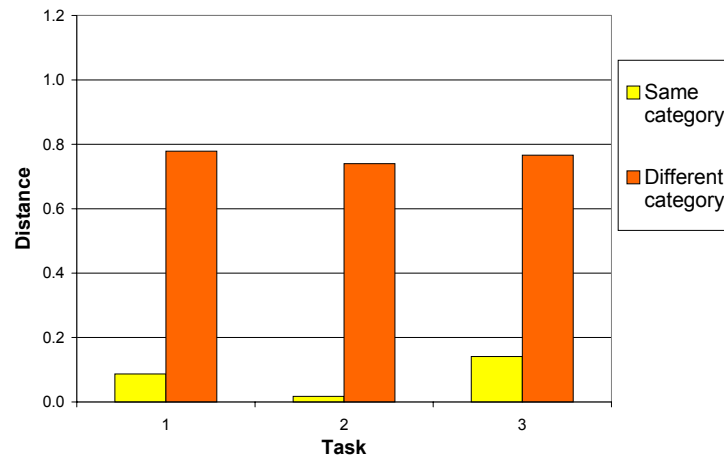


Figure 9. Third hidden layer: Euclidean distances between the points representing the activation patterns of objects of the same category vs. of different categories in the three tasks.

4 Discussion

The results of our simulations show that even in the very simple scenario of our simulations there is a complex interplay between perception and action in categorisation.

Consider first the role played by perception. As predicted, the task which is learned earlier (in terms of number of generations) and which therefore appears to be the simplest one is the task which requires to assign perceptually similar objects to the same category, while the most difficult task is the one which requires to assign objects which have no perceptual similarity to the same category. Furthermore, we found that, in the first hidden layer of our neural networks, where the task information still has not arrived and the units of this layer are processing visual information only, the activation patterns observed in the hidden units reflect the properties of the sensory input and the perceptual similarities among the objects.

These results show that when the neural networks prepare themselves for acting but haven't specific information on the action to perform, the perceptual characteristics of the objects on which they are expected to act represent the first cues they tend to use for categorisation. Consider that, differently from humans beings, neural networks do not have any previous experience of the world and of interactions between perception and action. The role played by perception in absence of specific action indication may be even more important in humans. In fact, in humans the tendency to group according to perceptual cues may get reinforced as generally in our life perception and action are linked. In fact, objects with similar perceptual characteristics, and especially objects with similar parts

[14], may elicit similar motor responses, as they have similar affordances [6,13,19]. If it is true, as argued by Glenberg [7], that categories are patterns of possible actions with objects, we may have evolved the tendency to put together perceptually similar objects in that perceptually similar objects tend to elicit similar actions.

But what happens when we have to perform a specific action and to reach a specific goal? In this case, if there is congruence between perceptual similarity and the task to perform, categorisation is facilitated (in our simulations this means that the task is learned in fewer generations). However, if in order to pursue a specific goal we need to put together perceptually dissimilar objects, task information overrides perceptual similarity. In both the second and the third hidden layers of our neural network model, categorisation is based on the task and therefore on the actions to perform, not on the perceptual characteristics of the objects. This does not mean that perceptual similarity doesn't play any role: in fact, in the second hidden layer categorisation is facilitated if there is congruence between perceptual similarity and task demand.

Of course, in the last (third) hidden layer the perceptual characteristics of the objects cease to play any role and at this level the nervous system is only preoccupied with encoding the specific action to perform. Notice, however, that actions are encoded at this level as macro-actions, that is, in terms of the overall goal of the action - pressing the right or left button -, not in terms of the specific sequence of micro-actions - specific micro-movements of the two segments of the arm - which can vary for the same macro-action depending on the starting position of the arm [5].

Our results can be interpreted as supporting the distinction, advanced by Barsalou [1], between primary and secondary categorisation. Primary categorisation has its basis in perceptual features; secondary categorisation in functional and action requirements. However, in our neural network model primary and secondary categorisation cannot be conceived of as separate processes. Rather, as soon as the task information arrives (in the second hidden layer), i.e., as soon as the goal and the macro-action the organism is going to perform become explicit, they influence and lead to a reorganisation of the sensory input. Thus perception and action cannot be considered as separate, but as meshed processes [18] which interact in categorisation. Altogether, the results of our simulations argue for an embodied view of cognition: categories reflect action patterns, and generally action patterns reflect perceptually salient parts of objects [4].

References

1. Barsalou L. W., Ad hoc categories. *Memory and Cognition* **11** (1983) pp. 211-217.
2. Barsalou L. W., Deriving categories to achieve goals. In *The psychology of learning and motivation: Advances in research and theory*, ed. by G. H. Bower (Academic Press, New York, 1991) Vol. 27, pp. 1-64.

3. Barsalou L. W., Perceptual Symbol Systems. *Behavioral and Brain Sciences* **22** (1999) pp. 577-609.
4. Berthoz A., *Le sens du mouvement* (Odile Jacob, 1997).
5. Di Ferdinando A. and Parisi D., *Micro-actions and macro-actions in neural networks* (in prep.).
6. Gibson J. J., *The ecological approach to visual perception* (Houghton Mifflin, Boston, 1979).
7. Glenberg A. M., What memory is for. *Behavioral and Brain Sciences* **20** (1997) pp. 1-55.
8. Goldstone R. and Barsalou L. W., Reuniting cognition and perception. The perceptual bases of rules and similarity. *Cognition* **65** (1998) pp. 231-262.
9. Magniè M. N., Ferreira C. T., Giusiano B. and Poncet M., Category specificity in object agnosia: Preservation of sensorimotor experiences related to objects. *Neuropsychologia* **37** (1999) pp. 67-74.
10. Martin A., Haxby J. V., Lalonde F. M., Wiggs C. L. and Ungerleider L. G., Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* **270** (1995) pp. 102-105.
11. Martin A., Wiggs C. L., Ungerleider L. G. and Haxby J. V., Neural correlates of highly specific knowledge. *Nature* **379** (1996) pp. 649-652.
12. Murphy G. L. and Medin D. L., The role of theories in conceptual coherence. *Psychological Review* **92** (1985) pp. 289-316.
13. Rosch E., Mervis C. B., Gray W. D., Johnson D. M. and Boyes-Braem P., Basic objects in natural categories. *Cognitive Psychology* **8** (1976) pp. 382-439.
14. Schyns P. G. and Murphy G. L., The ontogeny of part representation in object concepts. In *The psychology of learning and motivation*, ed. by D. Medin (Academic Press, New York, 1994) Vol. 31, pp. 305-349.
15. Smith L. B., Stability and Variability: the geometry of children's novel-word interpretations. In *Chaos theory in psychology*, ed. by F. D. Abraham and A.R. Gilgen (Praeger, London, 1995).
16. Smith L. B. and Heise D., Perceptual similarity and conceptual structure. In *Percepts, concepts and categories* ed. by B. Burns (Elsevier, Amsterdam, 1992).
17. Smith L. B. and Samuelson, L. L., Perceiving and Remembering: Category Stability, Variability and Development. In *Knowledge, Concepts, and Categories* ed. by K. Lamberts and D. Shanks (Psychology Press, Hove, 1997).
18. Thelen E., Schöner G., Scheier C. and Smith L. B., The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching. *Behavioral and Brain Sciences* **24** (2001) pp. 1-34.
19. Tucker M. and Ellis R., On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance* **24** (1998) pp. 830-846.
20. Tversky B. and Hemenway K., Objects, parts, and categories. *Journal of Experimental Psychology: General* **113** (1984) pp. 169-193.