

The Role of Action in Object Categorization

Andrea Di Ferdinando*^o

Anna M. Borghi[^]

Domenico Parisi*^{*}

*Institute of Cognitive Sciences and Technologies
National Research Council, Rome
Viale Carlo Marx, 15
00137, Roma, Italy

^oUniversity of Rome "La Sapienza"
Piazzale Aldo Moro, 5
00185 Roma, Italy

[^]Department of Psychology- University of Bologna
Viale Berti Pichat, 5
40127 Bologna Italy

andread@ip.rm.cnr.it borghi@psibo.unibo.it parisi@ip.rm.cnr.it

Abstract

The hypothesis underlying the present work is that knowledge organization in the mind/brain reflects the actions with which an organism responds to the input and not the perceptual properties of the input and it can be flexibly adapted to new goals. We present the results of two simulations using artificial neural networks in which an organism must generate various movements of its 2-segment arm in response to visually perceived objects. In the first simulation we find that the internal representations of the organism's neural network are organized in terms of macro-actions, that is, sequences of movements (micro-actions) that allow the organism to correctly respond to the objects. In the second simulation the organism's neural network is able to flexibly organize itself in order to adapt to different tasks. The network's internal representations of objects reflect the current task and not the perceptual similarities among the objects. In absence of task information, however, perceptual similarity is the best predictor of categorization. An account of categorization in terms of action makes it possible to explain the so-called 'shape-bias' in object categorization. Shape is a better predictor of categorization than other object properties such as color and size because we tend to respond in the same way to objects with the same shape rather than to objects with the same color or size.

Introduction

Categorization is the process, or the result of the process, of putting things together as members of the same category. The assumption of many classical categorization models is that categories of objects are formed on the basis of similarity among their members. However, the question is: what do we mean by similarity? The classical answer is that objects that are put together in the same category are perceptually similar. This answer is consistent with a general view of knowledge as based on the perceptual input we receive from the external environment. Things that send us inputs which are perceptually similar tend to

be put in the same category whereas things that are perceptually different are put into different categories.

However, many recent studies suggest that perceptual similarity may be a post-hoc notion unable to fully explain category membership (Murphy and Medin 1985; Rips 1989; Smith and Sloman 1994). In particular, many criticisms of perceptual similarity as a basis of category membership have been advanced by authors who have proposed a 'theory-based' view of categories. This view assumes that we need a principled way to put together objects in the same category. Similarity can be heuristically powerful only if we have a criterion to judge similarity and if we know how to apply the criterion. These authors emphasize the role of causal and explanatory relations among object features in providing such a criterion.

A further problem of the perceptual similarity view of categorization is that it seems to imply that knowledge organization is stable and non-adaptive (for a discussion of the different views of similarity, cf. Sloman and Rips 1998). However, the empirical evidence shows that knowledge is not stable but can be flexibly tuned and adapted to different needs (Barsalou 1987; Smith 1995; Smith and Samuelson 1997). For instance, when we have to perform an action to pursue a specific goal, we can form categories whose members are not necessarily perceptually similar such as "things to carry with us for a picnic". We are able to flexibly organize and create new categories of objects on the basis of more or less contingent goals (Barsalou 1991; 1999), as well as to select different object features which are relevant for the task at hand (Schyns, Goldstone, and Thibaut 1998). Accordingly, even though perceptual similarity may generally represent the basis for categorization (Goldstone and Barsalou 1998), it can be overridden for specific needs.

A different conception of categorization views categorization as based on action rather than on perception (Glenberg 1997). Objects are put in the same category if the organism tends to execute the same (types of) action(s) with respect to them. Notice that underlying the importance

of action for categorization does not mean that perception doesn't play a relevant role for categorization. Rather, according to this view perception and action are strongly interconnected, and perception is influenced by action (Viviani and Stucchi 1992). In fact, different perceptual features may become salient to fit the needs of the actions. In addition, since it is frequently the case that objects with respect to which the organism tends to execute the same action(s) are perceptually similar, perceptual similarity can become the basis on which objects may be classified together. However, the real basis of categorization is action, not perceptual similarity. Similarity of action is at the origin of categorization and when the two criteria are in conflict it is action that prevails, not perceptual similarity.

Simulations

In this paper we present some results of neural network simulations that try to make more concrete and specific this view of categorization as based on action. Our neural networks control the behavior of organisms that live in a bidimensional world and respond to visually perceived objects by moving their (single) arm (see Figure 1).

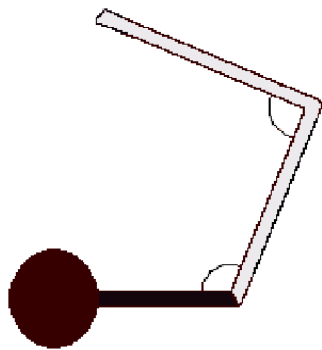


Figure 1. The organism with its two-segment arm.

A neural network functions in discrete cycles. In each cycle the network's input units encode the content of the organism's retina and the proprioceptive feedback from the arm; in fact the arm sends proprioceptive information to the network specifying its current position, i.e. the current angles between shoulder and forearm and between forearm and arm. The network's output units encode the movement of the arm in that cycle (micro-actions), i.e., the variation of the angles between shoulder and forearm and between forearm and arm. To correctly respond to perceived objects, the organisms must be able to generate an appropriate sequence of such movements. We will call macro-action a sequence of micro-actions at the end of which the arm stops. Notice that each macro-action can be realized by different sequences of micro-actions. For any given object, the arm's starting position can vary and, as a

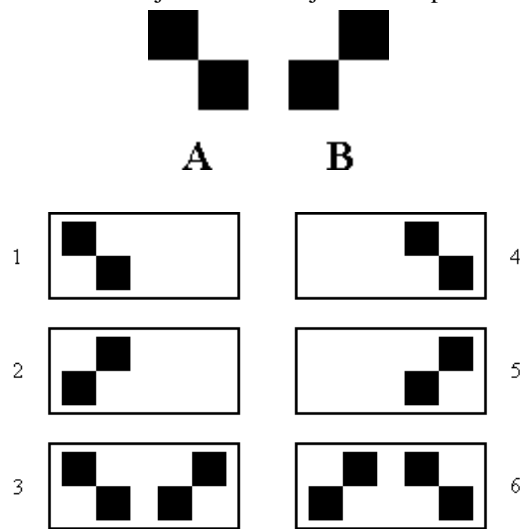
consequence, the arm's trajectory to reach the object will be different.

Simulation 1

In a first simulation we have shown that visually perceived objects are internally categorized in terms of the macro-action with which the organism responds to them rather than in terms of their perceptual properties (Di Ferdinando and Parisi 2001).

In this simulation the objects can be of two different shapes (A and B; cfr. Figure 2) and they can appear either in the left portion or in the right portion of the organism's retina. The organisms must respond according to the following two rules:

- 1) When a single object is presented, the organism has to reach for it, whatever the shape and the position of the object.
- 2) When both object A and object B are presented, the



organism has to reach for object A, ignoring object B.

There are six possible visual inputs (Figure 2).

Figure 2. The two objects and the six visual inputs.

In the first three cases, the organism has to reach the left portion of the visual space. In the last three cases the organism has to reach the right portion of the visual space. In other words, the organisms has to perform two different macro-actions: "moving the arm toward the left portion of the visual space" and "moving the arm toward the right portion of the visual space".

We used a genetic algorithm for evolving the network's connection weights allowing the organisms to generate the appropriate behavior. The network architecture is shown in Figure 3.

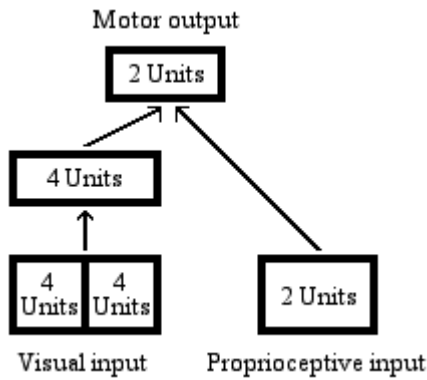


Figure 3. The network architecture.

At the end of evolution, we analyzed the internal representations of the best organisms in order to understand how the neural networks categorize the perceived objects. In neural terms categorization means that input patterns (vectors of activation states of the input units) that are categorized together tend to be transformed into internal patterns (vectors of activation states of the internal units) that are close together in the abstract hyperspace of all possible internal patterns, whereas input patterns that are categorized as belonging to different categories tend to be transformed into internal patterns that are more distant in the hyperspace of internal patterns.

What we find is that visual inputs that must be responded in the same way (with the same macro-action) are categorized as belonging to the same category whereas visual inputs that must be responded in different ways (with different macro-actions) are categorized as belonging to different categories (Figure 4).

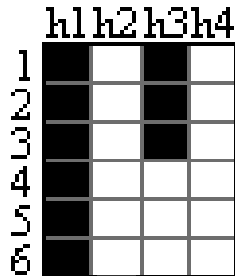


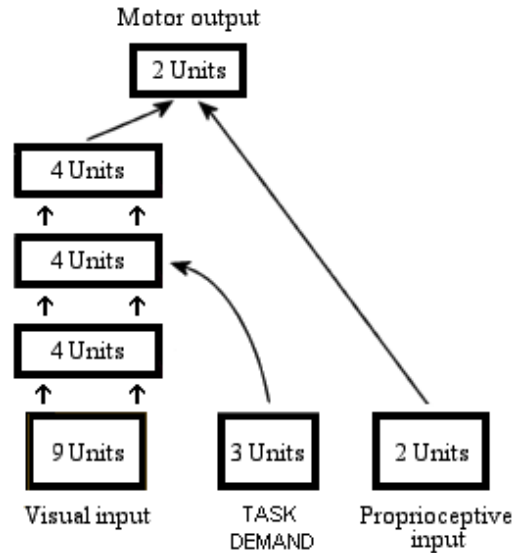
Figure 4. Activation level of the 4 hidden units (columns) in response to the six visual stimuli (rows) for one of the organisms analyzed. The activation level is mapped into a grey scale (0 = black and 1 = white).

Simulation 2

In the neural networks of the preceding simulation there was a single layer of internal units, and we have found that these internal units encode motor macro-actions rather than visual information. If we provide our neural networks with

a succession of internal layers we might be able to observe a more gradual mapping of visual into motor information.

In a second simulation (Borghi, Di Ferdinando, and Parisi 2002) the organism's neural network has three layers of internal units, instead than a single one, so that the visual input is progressively transformed by passing from



one layer to the next before reaching the output motor units (Figure 5).

Figure 5. The new network architecture.

Moreover, in the new simulation the same visually perceived object may be responded to with a variety of different actions, not just one as in the preceding simulation. The particular action with which the organism responds to the visual input depends on the particular task (goal) which currently controls the organism's behavior. In our simulations the current task is communicated to the neural network from outside, by encoding the task as a particular activation pattern in an additional set of input units (task units). In real organisms task information can also be internally generated inside the organism's body or even inside their neural network. The task input units send their connections not to the first layer but to the second layer of internal units which elaborate the visual input. Notice that until the organism's neural network is informed about the current task, i.e., prior to the second layer, the neural network cannot decide what is the appropriate action with which to respond to the visual input. Given this arrangement, we expect that in the early internal layers of the network architecture, i.e., those nearer to the visual input, the objects will tend to be represented and internally categorized in terms of their visual properties whereas as soon as the task information arrives to the subsequent layers of internal units, the objects will be represented and internally categorized in terms of the particular action with which, given the current task, the organism has to respond to the visual input.

In this second simulation there are four different objects, A, B, C, and D, which are graphically represented as a 3x3 matrix in which 4 of the 9 slots are filled, as shown in Figure 6.

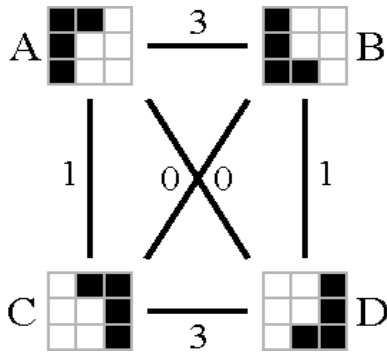


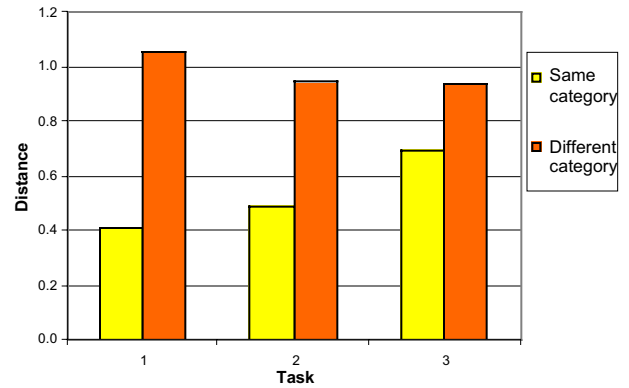
Figure 6. The four objects. Perceptual similarity of pairs of objects (number of shared filled slots) is also indicated.

In any given episode the organism sees only one of the 4 objects and it is required to categorize the object in one of two categories by pressing two different buttons with the endpoint of its arm. There are three tasks which are encoded in three task units. Task 1 is performed correctly if the two categories are composed by perceptually similar objects, i.e., objects with 3 common slots (A-B and C-D, Figure 6), Task 2 if the two categories are composed by objects with 1 common slot (A-C and B-D), and Task 3 if the two categories are composed by perceptually dissimilar objects, i.e., objects with no common slot (A-D and B-C).

In this case too we used a genetic algorithm for evolving good organisms. At the end of evolution, after 3.500 generations, the performance of the organisms does not differ in the three tasks. However, we find that Task 1 requires a smaller number of generations to be learned than Task 2 and Task 3. Thus, categorization performance is influenced by perceptual similarity between the objects.

Let us now turn to the analysis of the way objects are internally represented in the neural networks of the best organisms. We calculated the distance between pairs of points (internal activation patterns) which represent the objects in the hyperspace of each layer of internal units. In the first layer, where task information hasn't arrived yet, the closest points are those which represent perceptually similar objects while the more distant points are those which represent perceptually dissimilar objects. Thus, in absence of task information categorization is predicted by perceptual similarity between the objects.

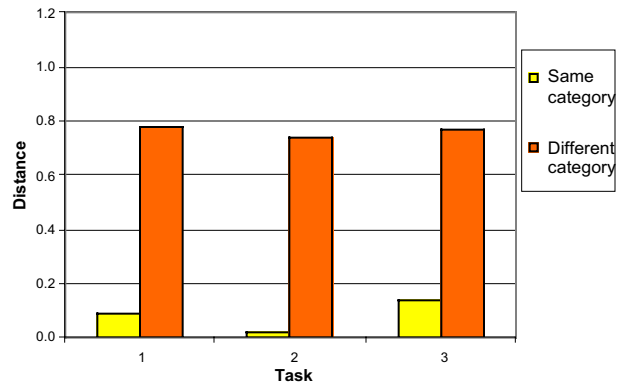
But what happens in the second and third internal layers, when task information has arrived? In the second layer, the critical one, the closest points are those representing objects which are responded to with the same macro-action (see Figure 7). As expected, already at this level task information overrides perceptual similarity. However, perceptual similarity still plays a role in this second layer. In fact, the result is stronger for Task 1 (highly perceptually similar objects) than for Task 2 (less



perceptually similar objects) and for Task 3 (perceptually dissimilar objects).

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

In the third internal layer action clearly overrides perceptual similarity. As in the second layer, the smallest distances are those between points representing objects which require the same response. However, in the third layer the result is stronger than in the second layer and there are no differences among the three tasks (see Figure 8). In other words, at this level of neural processing perceptual input is represented in terms of the macro-actions with which the organism must respond to the input



and not in terms of the perceptual properties of the input.

Figure 8. Third hidden layer: distances between the points representing the activation patterns of objects belonging to the same vs. different categories in the three tasks.

Discussion

Our simulations can be interpreted as showing that categorization is action-based. Assigning a relevant role to action in categorization does not mean that perceptual similarity between objects is unimportant (Hampton 1998; Goldstone and Barsalou 1998). In everyday life perceptually similar objects generally elicit similar actions and therefore we probably have evolved the ability to respond in the same way to perceptually similar objects

(Gibson 1979). Empirical evidence shows that the affordances of objects direct our attention and activate particular kinds of actions. Preparation to act on an object produces faster processing of stimuli congruent with that object (Craighero et al. 1999), thus we are faster to respond in the same way to perceptually similar objects. However, as recently shown by Tucker and Ellis (1998), the way visual objects are represented may include information on how to act upon them.

An account of categorization in terms of actions can explain why shape is more important than other properties of objects such as color or size in determining how objects are categorized and why nouns tend to refer to objects that have the same shape (Biederman 1987; Landau, Smith, and Jones 1998; Shyi et al. 2001; Smith 1995). If categories of objects depend on perceptual similarity, it is not clear why perceptual similarity in shape is more important than perceptual similarity in color or size in forming object categories. The answer becomes clear if we recognize that action, not perceptual similarity, is the real basis of object categorization. For organisms such as us, objects that tend to be responded to with the same action(s) are more likely to be objects that have the same shape but can have different colors or sizes, whereas objects that tend to be responded to with different actions are more likely to be objects with different shapes and, perhaps, the same color or size. In some experiments texture emerges as another property of objects which plays an important role in object categorization. Texture tends to distinguish animals from nonanimals, and we behave in very different ways in response to animals than nonanimals. Hence, in this case too objects which vary in texture tend to be placed in different categories based on the actions with which we respond to them.

References

- Barsalou, L. W. 1987. The instability of graded structure: implications for the nature of concepts. In U. Neisser ed., *Concepts and conceptual development: Ecological and intellectual factors in categorization*, 101-140. Cambridge, Mass.: Cambridge University Press.
- Barsalou, L. W. 1991. Deriving categories to achieve goals. In G. H. Bower ed., *The psychology of learning and motivation: Advances in research and theory*, Vol. 27, 1-64. New York: Academic Press.
- Barsalou, L. W. 1999. Perceptual Symbol Systems. *Behavioral and Brain Sciences* 22: 577-609.
- Biederman, I. 1987. Recognition by components: A theory of human image understanding. *Psychological Review* 94: 115-147.
- Borghi, A. M., Di Ferdinando, A., and Parisi, D. 2002. The role of perception and action in object categorization. In J. A. Bullinaria and W. Lowe eds., *Proceedings of the seventh neural computation and psychology workshop: connectionist models of cognition and perception*. Singapore: World Scientific.
- Craighero, L., Fadiga, L., Rizzolatti, G., and Umiltà, C. 1999. Action for perception: a motor-visual attentional effect. *Journal of Experimental Psychology: Human Perception and Performance* 25: 1673-1692.
- Di Ferdinando, A., and Parisi, D. 2001. Micro-actions and macro-actions in neural networks, Technical Report, NSAL-01-03, Institute of Psychology, CNR, Rome.
- Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Glenberg, A. M. 1997. What memory is for. *Behavioral and Brain Sciences* 20: 1-55.
- Goldstone, R. and Barsalou, L. W. 1998. Reuniting cognition and perception. The perceptual bases of rules and similarity. *Cognition* 65: 231-262.
- Hampton, J.A. 1998. Similarity-based categorization and fuzziness of natural categories. *Cognition* 65: 137-165.
- Landau, B., Smith, L. and Jones, S. 1998. Object Shape, Object Function, and Object Name. *Journal of Memory and Language* 38: 1-27.
- Murphy, G. L. and Medin, D. L. 1985. The role of theories in conceptual coherence. *Psychological Review* 92: 289-316.
- Rips, L. J. 1989. Similarity, typicality, and categorization. In S. Vosniadou and A. Ortony eds., *Similarity and analogical learning*. New York: Cambridge University Press.
- Schyns, P. G., Goldstone, R. L., and Thibaut, J. P. 1998. Development of features in object concepts. *Behavioral and Brain Sciences* 21: 1-54.
- Shyi, G. S. W., Goldstone, R. L., Hummel, J. E., and Lin, C. 2001. Computing representations for bound and unbound object matching. Forthcoming.
- Sloman, S. A., and Rips, L. J. 1998. Similarity as an explanatory construct. *Cognition* 65: 87-101.
- Smith, E. E., and Sloman, S. A., 1994. Similarity versus rule-based categorization. *Memory and Cognition* 22: 377-386.
- Smith, L. B. 1995. Stability and Variability: the geometry of children's novel-word interpretations. In F. D. Abraham and A. R. Gilgen eds, *Chaos theory in psychology*. London: Praeger.
- Smith, L. B. and Samuelson, L. L. 1997. Perceiving and Remembering: Category Stability, variability and Development. In K. Lamberts and D. Shanks eds, *Knowledge, Concepts, and Categories*, 161-195. Hove: Psychology Press.
- Tucker, M., and Ellis, R. 1998. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance* 24: 830-846.
- Viviani, P., and Stucchi, N. 1992. Biological movements look uniform, evidence for motor-perceptual interactions. *Journal of Experimental Psychology, Human Perception and Performance* 18: 603-623.