Objects and Affordances: An Artificial Life Simulation

Giorgio Tsiotas (tsiotas@libero.it)
Department of Computer Science, 7 Mura Anteo Zamboni
Bologna, 40127 Italy

Anna M. Borghi (annamaria.borghi@unibo.it)
Department of Psychology, 5 Viale Berti Pichat
Bologna, 40127 Italy

Domenico Parisi (domenico.parisi@istc.cnr.it)
Institute of Sciences and Technologies of Cognition (ISTC), 15 Viale C. Marx,
Rome, 00137 Italy

Abstract

We simulated organisms with an arm terminating with a hand composed by two fingers, a thumb and an index, each composed by two segments, whose behavior was guided by a nervous system simulated through an artificial network. The organisms, which evolved through a genetic algorithm, lived in a bidimensional environment containing four objects, either large or small, either grey or black. In a baseline simulation the organisms had to learn to grasp small objects with a precision grip and large objects with a power grip. In Simulation 1 the organisms learned to perform two tasks: in Task 1 they continued to grasp objects according to their size, in Task 2 they had to decide the objects' color by using a precision or a power grip. Learning occured earlier when the grip required to respond to the object and to decide the color was the same than when it was not, even if object size was irrelevant to the task. The simulation replicates the result of an experiment by Tucker & Ellis (2001) suggesting that seeing objects automatically activates motor information on how to grasp them.

Introduction

In opposition to theories that posit perception and action as separate (Pylyshyn, 1999), it has recently been suggested that perception and action are strictly interwoven and that perception is guided and filtered by action (Berthoz, 1997; Ward, 2002). In a related way, recent theories emphasize the interconnections between sensorimotor and cognitive processes. In particular, it has been proposed that cognition is embodied, i.e., that it depends on the experiences that result from possessing a body with given physical characteristics and a particular sensorimotor system (Barsalou, 1999; Glenberg, 1997; Goldstone & Barsalou, 1998). This view of cognition is clearly in opposition to the classical cognitivist view according to which mind is a device for manipulating arbitrary symbols.

More specifically, some theories suggest that object concepts re-enhance sensorimotor experiences with objects (Barsalou, 1999; Glenberg, 1997; Pecher & Zwaan, 2005). Studies on the relationships between the visual system and the motor system claim that seeing an object tends to evoke its affordances, re-activating previous experiences and interactions with the object. An affordance refers to a property on an object that influences how the object can be used (Gibson, 1979). For example, the properties of the handle of a door determine how one opens the door - by pulling, or pushing, or twisting, and so on. Accordingly, seeing an object such as a cup may re-activate the affordances linked to reaching and grasping the cup's handle, even if the position of the handle is not relevant to the task at hand (Tucker & Ellis, 1998).

Most evidence regarding the strict interrelations between perception and action concerns simple interactions with objects, rather than complex actions probably mediated by the actor's goals. So, for example, protruding object parts may activate reaching motor behaviors, whereas objects of a specific size may activate specific grasping behaviors. In order to refer to these "low-level" affordances, Ellis & Tucker (2000) created the expression of "micro-affordances". Micro-affordances facilitate simple and specific kinds of interactions with objects but they do not pertain complex, goal-mediated actions. Most importantly, micro-affordances probably imply access to conceptual knowledge, as they are rather specific and suitable for a given object. So, for example, seeing an object does not elicit simply a grasping behavior, but a specific type of grasping behavior which is suitable for that particular object.

Ellis & Tucker (2000) and Tucker & Ellis (2001) have demonstrated the role played by affordances in eliciting motor behavior by presenting participants with real objects of different sizes. Participants had to categorize the objects as natural or artefact using either a power grip or a precision grip. They found a compatibility effect between the kind of grasp and a task-irrelevant dimension, the object`s size. The effect was also observed when the object was located outside the reaching space, which suggests that seeing the object activates the simulation of a specific type of grasping. A similar compatibility effect was found between the direction of the wrist's rotation and the kind of grasp required by the object. For example, objects such as bottles facilitated responses with a clockwise wrist rotation, while objects such as toothbrushes facilitated a counter-clockwise wrist rotation.

Aim of the paper

Aim of the present paper is to reproduce with an Artificial Life connectionist simulation the situation explored by T&E. A typical feature of Artificial Life connectionist simulations is that the experimenter can control how organisms learn to perform a given task but he or she doesn't pre-define the single learning steps and the organisms find their own way to solve the problem. In our case, we first taught the organism to grasp small and large objects using two different types of grip, trying to reproduce in this way real life experiences in interacting with objects. Then we reproduced Tucker & Ellis's (2001) experiment with some minor variations.

Our aim was to assess whether previous grasping experiences with objects of different size influenced the organism's performance when the object's size was irrelevant to the task at hand, i.e., in a task in which the organisms had to perform a different kind of grip depending on the color of the object they were seeing.

Connectionist simulations not only make it possible to replicate behavioral tasks but they also allow us to analyze the activation patterns of the neural network's hidden units, i.e., the neural organization that the network acquires to solve the problem. This may seem odd, as the artificial networks used in the simulations are enormously simplified compared to real brains, and the analysis of hidden units' activations may seem an impoverished replication of brain scanning techniques. However, connectionist simulations allow us to analyze the hidden units organization at different learning stages, which brain scanning and other techniques can't easily do.

The model

In our study we simulated artificial organisms endowed with a visual system and a motor system. The visual system allows the organisms to see different objects, one at a time. The motor system consists of a single arm composed by two segments; by moving the arm different points of space can be reached with the arm's endpoint (the hand) (see Figure 1). The arm sends proprioceptive information to the organisms that specifies the arm's current position. The arm terminates with a hand composed by only two fingers, each in turn composed by two segments. One of the finger, the thumb, is shorter than the other one, the index finger. With their hand the organisms can perform all kinds of grips. In our simulation we were interested in two different kinds of grips. The precision grip was defined by the opposition between the thumb and the index which are spatially close to each other. The power grip was defined by the fact that it is a kind of grip suitable for larger objects, i.e., the distance between the finger and the thumb is much larger than for the precision grip. Consider that the index finger can be conceived of as a "virtual finger". Namely, when we grasp objects, we use all the fingers except the thumb as a single functional unit, applying force to the object (Arbib, 2002). Thus our index finger does not model only the index movements but the movements of the whole hand (except the thumb). The fingers also send proprioceptive information to the organisms, which therefore are "aware" both of the arm and the finger positions. Consider that our organisms receives both proprioceptive and visual information, but are unable to see their hand moving in space and to detect visually how they are grasping the object.

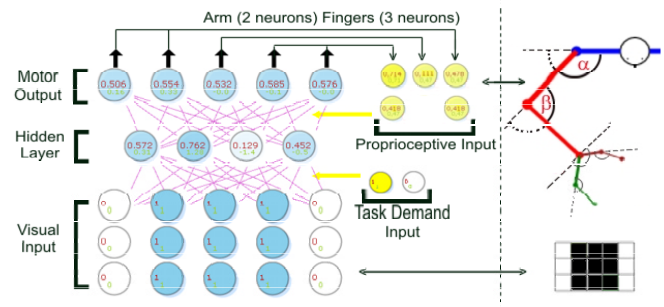

Figure 1: The arm and the hand of the organism and how sensory-motor information is encoded in the organism's neural network.

The behavior of the organisms is controlled by a nervous system, which is simulated with a neural network (see Figure 1 and Figure 2). The network architecture consists of 3 layers: one input layer with 3 different groups of units, one intermediate layer of 4 hidden units, and one output layer of 5 units.

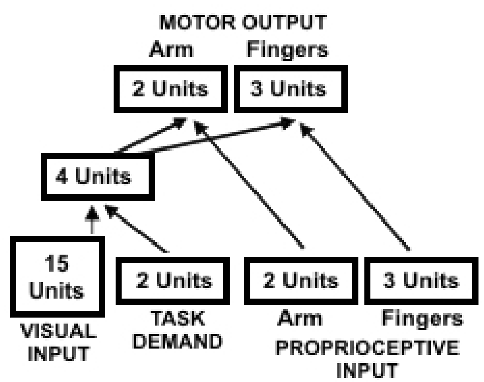

Figure 2: The neural network architecture.

In the input layer there are 3 groups of neurons. The first group, composed by 15 units, encodes the perceptual properties of the objects that the organism "sees". The input value can vary within a range from 0 to 1. As shown in Figure 1, a large object corresponds to 9 filled cells in a matrix of 15 cells, whereas a small object corresponds to 1 filled cell in the same matrix of 15 cells. Grey objects are encoded with a 0.5 activation value; black objects with a 1.0 value. A second group of 2 input units encodes information specifying the task the organism is required to perform, and a third group of 5 units encodes proprioceptive information. Two proprioceptive input units encode the current angles between the shoulder and the arm and between the arm and the forearm, while the remaining 3 units encode the 3 angles of the fingers (2 angles between the fingers' falangi and another one between the 2 fingers).

The 5 output units encode the movements of the organism's hand and fingers, by specifying the variation of the previously described angles. As shown in Figure 2, the visual input units and the task units are connected with the hidden units, while the proprioceptive input units are directly connected with the output units.

On each trial the organism sees a single object. There are 4 different objects: 2 of them are small and 2 are large. Both the small and the large objects can be of two different colors: either grey or black.

Each organism is a member of a population of 100 organisms. To find the connection weights which allow the organisms to perform correctly the task we used a genetic algorithm, the evolution strategy described by Rechenberg (1973). We first assigned random connection weights to the neural networks of an initial population of 100 organisms. Each organism had a genotype encoding the organism's connection weights. We used a direct one-to-one mapping: each gene encoded a different connection weight as a real number. Then we tested each of these 100 organisms on 16 randomly selected trials in Simulation 1 and on 32 randomly selected trials in Simulation 2. In each trial each organism started with a randomly chosen position of the arm and fingers and saw one of the four objects. At the end of the trials we assigned each organism a fitness value reflecting the organism's ability to perform the task.

The fitness value was calculated when the fingers stopped to move. A positive fitness value was given to the organisms if the fingers stopped while touching the borders of the object. Also the distance of the fingers from the object and the distance between the index finger and the thumb were evaluated and influenced fitness. The 20 best organisms were selected for (nonsexual) reproduction and each of them generated 5 offspring inheriting their parent's genotype with the addition of some random mutations. The 20x5=100 organisms thus obtained represented the new generation. The process was repeated for 2000 generations.

Predictions

In Simulation 1 the task consisted in reaching and grasping appropriately the objects, i.e., the organisms had to learn to reach the objects and to grasp small objects with a precision grip and large objects with a power grip. Thus, Simulation 1 was simply aimed to replicate what we typically do in real life, so we only expect that the organisms learn the task. In Simulation 2 the organisms had two different tasks: either they had to grasp the object with the appropriate grip or they had to respond to the object's color by using either a precision or a power grip. Our critical predictions concern Simulation 2. We expect that responses are faster (in terms of number of generations necessary to learn the task) in case of compatibility between the object size and the kind of grip used to classify the object, thus suggesting that object's size automatically evokes, or "affords", a specific response. We also predict that this advantage of the Compatible condition is reflected in the hidden units organization.

Simulation 1

In Simulation 1 we reproduced the real-life experience of grasping objects of different sizes in an appropriate way. In this simulation the organisms had to learn to react appropriately to the object's affordances: they had to learn that, when they saw an object, they had to reach the object and grasp it in the appropriate way, i.e., using a power grip for large objects and a precision grip for small ones. This reflects exactly what we typically do in real life: we learn to react appropriately to objects' affordances and to adapt our grip to object size.

The fitness of the organisms depended on both their capability to reach the visually perceived object, i.e., to bring the hand to the right region of space, and to grasp the object appropriately.

In order to obtain reliable results the simulation was repeated 10 times, starting with different sets of initial connection weights.

Results

All the results presented are the average of the 10 replications. We calculated the average fitness of the organisms based on the percentage of correct responses and of errors (trials in which the organism did not reach or did not grasp appropriately the object) in performing the task. At the end of the simulation, i.e., after 2000 generations, the best organisms were able to respond correctly to all patterns, as shown in Figure 3.

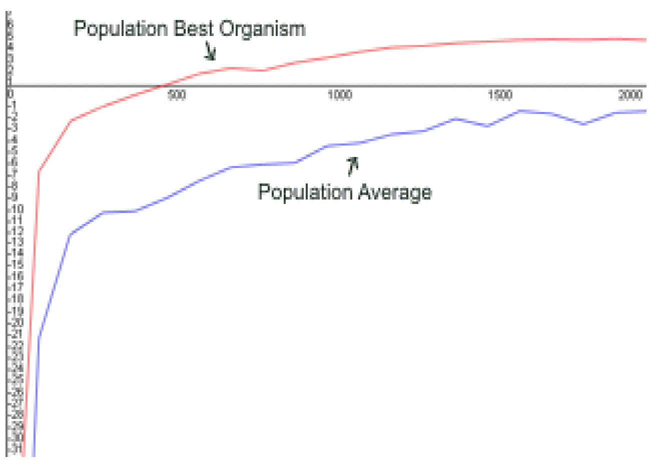

Figure 3: Simulation 1. Fitness of the best organisms and average population fitness across 2000 generations.

Simulation 2

After being trained for 2000 generations in Simulation 1, in Simulation 2 the organisms were trained for further 2000 generations with 2 different tasks. The 2 tasks were encoded in the 2 input task units as 01 and 10, respectively.

As in Simulation 1, the organisms saw four objects, one at a time, which could be small or large, black or grey. In order to make sure that the organisms didn't "forget" what they had previously learned, with Task 1, encoded as 01, the organisms had to respond to the size of the object by grasping the object with an appropriate grip, ignoring the object's color. With Task 2, encoded as 10, they had to respond to the object's color, ignoring its size. They had to respond using a precision grip when the object was grey and a power grip when the object was black. Thus in Task 2 the object's size became irrelevant to the task.

Task 2 reproduced the laboratory situation devised by T&E, with some small variations. T&E asked participants to classify the objects by pressing a device using either a power or a precision grip independent of whether natural objects (i.e., an apple or a cherry) or artifacts (i.e., a bottle or a needle) were small or large objects. In our simulation, teaching the network to distinguish between natural objects and artifacts would have been rather implausible. For this reason we decided to train the network to respond to objects of different color. Notice that the crucial aspect of T&E's experiment was maintained: namely, the objects' size was not relevant to the task (Task 2), independent of the fact that the organisms had to decide what kind of object (natural/artefact) they were seeing, or what color (grey/black) they were seeing. If the objects' size affords a specific kind of action, then it should influence the organisms performance not only in Task 1, but also in Task 2.

Accordingly, our crucial prediction is that in the Compatible Condition - i.e., when object size and the kind of grip used to classify the object correspond - learning occurs earlier (in terms of number of generations) than in Incompatible Condition. If this is true, it would suggest that object's size automatically evokes a specific kind of motor response, even if object's size is not relevant to the task at hand.

Results

Simulation 2 was also repeated 10 times and the results presented are the average of the 10 replications.

In order to control whether learning occurred earlier (in terms of number of generations) in the Compatible than in the Incompatible condition, we calculated the fitness for the best organism of each generation in Task 1 and in Task 2, in both the Compatible and in the Incompatible conditions; we also calculated the population average.

As shown in Figure 4, the fitness of the best organisms and of the population average was higher in Task 1 than in Task 2. However, the organisms learned both tasks. More importantly, in Task 2 Compatible Condition the average fitness for all the generations tested was superior than the average fitness in the Incompatible Condition. The advantage of the Compatible over the Incompatible condition was maintained also with respect to the performance of the best organisms of the generations we tested (see Figure 4).

In order to compare the data obtained in the two conditions, we performed four within subjects Anovas. We compared the average fitness of the best organisms of each of the 10 replications in the Compatible and in the Incompatible conditions at generations 500, 1000, 1500, and 2000. The results were straightforward and confirmed our prediction. At generation 500 and 1000 the performance of the best organisms in the Compatible condition was significantly better than in the Incompatible condition, as indicated by the significant difference between the Fitness value in the Compatible and in the Incompatible Condition [F (1,9) = 6.42, Mse = 2.6, p >.03;F (1,9) = 5.81, Mse = 2.87, p >.04]. The advantage of the Compatible over the Incompatible condition dramatically decreased at generations 1500 and 2000, suggesting that the best organisms had learned to respond to all patterns. This clearly indicates that the Compatible patterns were learned earlier, in terms of number of generations, than the Incompatible ones.

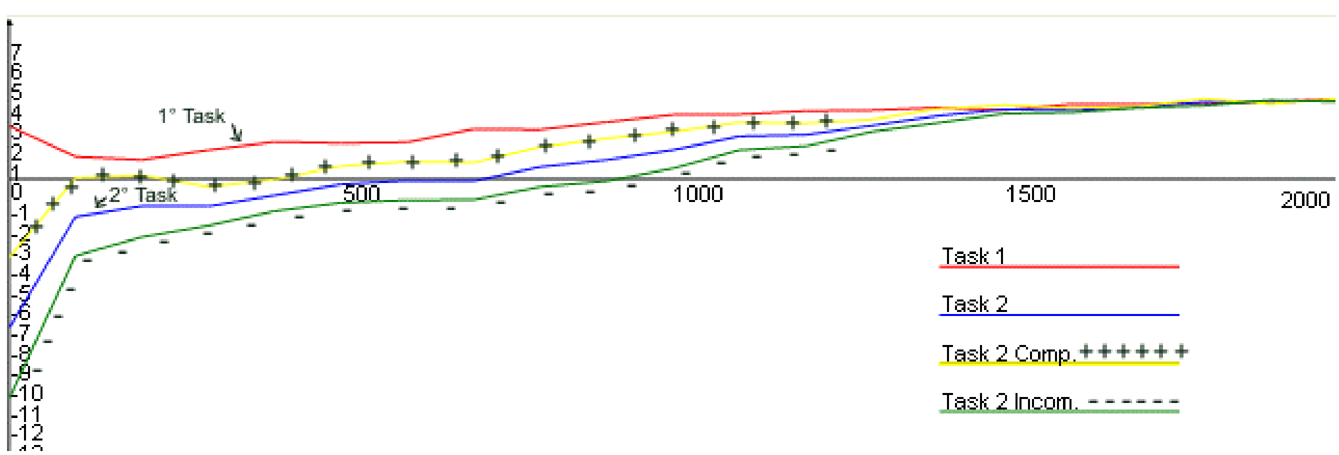

Figure 4: Simulation 2. Fitness of the Best Organisms in Task 1 and in Task 2, and comparison between the Compatible and the Incompatible Conditions in Task 2.

Hidden units analysis

The results we found support our initial predictions: seeing objects automatically activates information on how to grasp them, also when this information is not relevant to the task. The hidden units analysis we performed was aimed to detect what kind of neural organization developed to perform the tasks. More specifically, we were interested in the neural organization used in Task 2 in both the Compatible and Incompatible conditions. We selected the networks of the 10 best organisms of the last generation (generation 2000) and analyzed their hidden units' activation patterns. The hypotheses we wanted to test are the following. If the motor information related to object size is automatically activated in Task 2, then: (1) the network should tend to change only the activation pattern of the Incompatible condition; (2) in the activation pattern of the units used for Task 2, the network should keep some "trace" of what it learned in Task 1. To find such a trace could help explain the advantage of the Compatible over the Incompatible condition.

Figure 5 shows the activation patterns of the hidden units for two replications of the simulation. In the first row of each replication one can see the activation patterns of the 4 hidden units for each of the 4 patterns of Task 1, while the second row shows the activation patterns of the 4 units for each of the 4 patterns of Task 2.

The first and the last pattern in the second row represent the Compatible Condition: given a large, black object, the organisms have to respond with a power grip; given a small, grey object, the organisms have to respond with a precision grip. The two central patterns represent the Incompatible Condition, i.e., the cases in which the organisms had to respond with a power grip to a small, black object, and with a precision grip to a large, grey object. In interpreting the figures, our analysis will focus on the units whose value varies, as they may be active or not active depending on the pattern and the task.

Figure 5 shows that in order to perform Task 1 the network typically uses a single hidden unit, the value of which switches from 0 to 1. For example, in Seed 250 (see Figure 5) only the fourth neuron varies, and it has value 0 for large objects and value 1 for small objects.

The situation is more complex for Task 2. First of all, consider that across all seeds the network changes the 2 values of both Incompatible patterns. In the Compatible condition in 4 out of 10 replications the activation patterns do not change at all, while in the remaining 6 cases, only one of the two patterns changes. This confirms our first hypothesis, at least for the majority of the replications: in the Compatible condition the network maintains the same activation patterns used in Task 1 and a new unit is simply added to encode the new task. On the contrary, in the Incompatible condition the network has to be re-organized, and different units are used to encode the 2 tasks.

Consider now the strategies used by the network in Task 2. It uses two different strategies, which we could call a "modular" strategy and a "distributed" strategy. In 6 out of 10 replications the network, in order to determine the object's color (Task 2), does not use the same neuron used to determine the object's size (Task 1). The replication based on random seed 114 is a good example (see Figure 5): while the network uses the fourth neuron for Task 1, it uses the second neuron for Task 2. This suggests that while doing Task 2 the network "keeps track" of what it has learned in Task 1, and in order to determine color it does not use the same module "dedicated" to size. In the remaining 4 cases, the network distributes the information concerning color, i.e., the information useful to perform the motor task, over two neurons, one of which was the one used in Task 1 to determine the object's size. An example is represented by the replication based on seed 250 (see Figure 5). This strategy has the advantage to be quite economical, as it allows the network to change the values of only one of the neurons used for Task 1.

Both the modular and the distributed strategies suggest that the network did not "forget" information used for Task 1. Consider, however, that the network could use a different strategy: it could keep the neuron "dedicated" to size with the same values and use 2 other neurons to determine color. This did not happen. The network worked in a very economical way, using only two neurons for both tasks. Thus in Task 2 the network "kept track" of Task 1, but only for the Compatible condition. Otherwise it used primarily an action-based strategy, i.e. a strategy which was oriented to the output. Thus our second hypothesis is only partially confirmed.

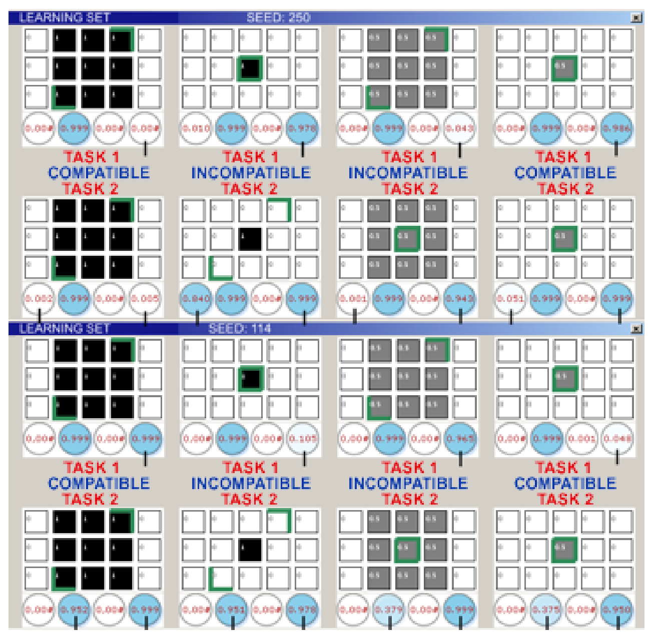

Figure 5: Simulation 2. The activation pattern of the hidden units in Task 1 (first row) and Task 2 (second row).

Conclusion

Our results confirm the findings obtained by T&E in their experiments: seeing objects automatically elicits motor information concerning the way we interact with the objects. Even if the object's size is irrelevant to the task at hand, seeing the object activates the kind of prehension appropriate for its size. When the kind of grip and the object's size are compatible, responding to the object's color is quicker than when the kind of grip and the object's size are not compatible.

Our simulations leave the question open of whether compatibility results as those that we have described imply access to conceptual knowledge. According to an influential account, two different routes to action exist: a direct vision-to-action route, mediated by on-line dorsal system processes, and a mediated vision-to-semantics-to-action ventral route (Rumiati & Humphreys, 1998). However, recent evidence suggests that the direct route to action may be limited to novel objects. With a dual task paradigm, Creem and Proffitt (2001) have shown that the ability to grasp common objects such as a hammer or a toothbrush appropriately by, for example, reaching for the object's handle even if it is not oriented towards us, decreased with a semantic interference task, but not with a spatial interference task. This suggests that to perform gestures appropriate to objects it is necessary to combine conceptual knowledge with affordances derived from objects (Buxbaum, Schwartz & Carew, 1997). In addition, there is recent evidence of action-based compatibility effects, as those we obtained in the present simulation, with words rather than pictures (Borghi, Glenberg & Kaschak, 2004; Tucker & Ellis, 2004). The presence of these effects also with words argues either for a representation of the object encoded also in the dorsal stream (Gentilucci, 2003) or for the involvement not only of the dorsal but also of the ventral system and of long-term knowledge in generating affordances (Tucker & Ellis, 2004). These effects would be accounted for by long-term visuomotor associations between objects and actions executed on them. Simulating compatibility effects also with words referring to objects is one of the possible extensions of our work.

Acknowledgments

Thanks to Andrea Di Ferdinando and Marco Mirolli for discussion and useful suggestions.

References

Arbib, M.A. (2002). Grounding the mirror system hypothesis for the evolution of the language-ready brain. In A. Cangelosi, D. Parisi (eds.), *Simulating the evolution of language*. London: Springer.

Barsalou, L.W. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences,* 22, 577-609.

Berthoz, A. (1997). *Le sens du movement*. Paris: Odile Jacob.

Borghi, A.M. (2005). Object concepts and action. In D. Pecher, R. Zwaan (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking.* Cambridge: Cambridge University Press.

Borghi, Glenberg, Kaschak (2004). Putting words in perspective. *Memory and cognition,32,* 863-873.

Buxbaum, L. J., Schwartz, M. F., & Carew, T. G. (1997). The role of semantic memory in object use. *Cognitive Neuropsychology, 14*, 219-254.

Creem, S.H. & Proffitt, D.R. (2001). Grasping objects by their handles: a necessary interaction between perception and action. *Journal of Experimental Psychology: Human Perception and Performance, 27,*1, 218-228.

Ellis, R. & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology, 91,* 451-471.

Gentilucci, M. (2003). Object motor representation and language. *Experimental Brain Research, 153*, 260-265.

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., Barsalou, L. W. (1998). Reuniting cognition and perception. The perceptual bases of rules and similarity. *Cognition,* 65, 231-262.

Parisi, D., Borghi, A.M., Di Ferdinando, E. & Tsiotas, G. (in press). Meaning and motor actions: Artificial Life and behavioral evidence. Commentary to Arbib, M.A. (in press). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences.*

Pecher, D. & Zwaan, R. (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking.* Cambridge: Cambridge University Press.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341-423.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution.* Stuttgart, Germany: Frommann-Holzboog Verlag.

Rumiati, R.I. & Humphreys, G.W. (1998). Recognition by action: Dissociating visual and semantic routes to action in normal observer. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 631-647.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance, 24*, 3, 830-846.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*, 769-800.

Tucker & Ellis (2004). Action priming by briefly presented objects. *Acta Psychologica, 116,* 185-203.

Vogt, S., Taylor, P., & Hopkins, B. (2003). Visuomotor priming by pictures of hand pictures: perspective matters. *Neuropsychologia, 41,* 941-951.

Ward, R. (2002). Independence and integration of perception and action: An introduction. *Visual Cognition, 9,*385-391.