Objects, words, and actions.

Some reasons why embodied models are badly needed in cognitive psychology.

Anna M. Borghi [○Δ], Daniele Caligiore [○Δ], Claudia Scorolli[*].

[○] Department of Psychology, University of Bologna

[*] Department of Communication Disciplines, University of Bologna

[Δ] Institute of Science and Technology of Cognition, CNR, Rome

To appear in MATHEMATICS AND SOCIETY, Springer.

Objects, words, and actions.

Some reasons why embodied models are badly needed in cognitive psychology.

Abstract

In the present chapter we report experiments on the relationships between visual objects and action and between words and actions. Results show that seeing an object activates motor information, and that also language is grounded in perceptual and motor systems. They are discussed within the framework of embodied cognitive science. We argue that models able to reproduce the experiments should be embodied organisms, whose brain is simulated with neural networks and whose body is as similar as possible to humans' body. We also claim that embodied models are badly needed in cognitive psychology, as they could help to solve some open issues. Finally, we discuss potential implications of the use of embodied models for embodied theories of cognition.

1. Introduction

This chapter begins and ends with the conviction that cognitive neuroscience need to broaden their objectives making a more extensive use of models that adequately reproduce the brain and bodily characteristics of human beings. In the introduction we will briefly outline the theoretical framework of our work describing the main assumptions and claims of embodied theories of cognition. Then we will focus on some experimental studies, in which the dependent variables consist either in response times or in kinematics measures. First we will describe a study on object categorization, then three studies on language grounding: all the studies provide evidence of the strict relationship between concepts, words, and the motor system.

While describing these studies we will show that the experimental results obtained cannot be modelled without using embodied models, i.e. models that reproduce, at least in part, both the neural structure of the human brain and some of the crucial sensorimotor characteristics of human arms/hands. In addition, we will illustrate some studies performed within embodied cognition that leads to ambiguous results. In particular, we will focus on the issue of whether reading sentences related to a given effector (e.g., hand, mouth, foot) leads to a facilitation or to an interference of responses performed with the same effector. Referring to this example, we will show that embodied models might be crucial in order to disentangle some crucial issue for the field, and to formulate more precise predictions.

In the conclusions we will discuss potential implications of the use of embodied models for embodied theories of cognition. On one side, embodied models can represent a strong test for embodied theories of cognition. On the other side, however, assuming a very strong embodied view could render it impossible to propose a real comparison between the performance of real humans and the performance obtained by simulated organisms.

2. Theoretical framework: embodied theories of cognition

Within the field of cognitive science, traditional views intend concepts and words as abstract, amodal, and arbitrarily related to their referents. In the recent years embodied views have provided

a lot of evidence showing that concepts and words are not abstract, amodal, and arbitrary (Fodor, 1975). Rather, they are grounded in sensorimotor processes, therefore in our bodily states. Among others, Barsalou (1999) has proposed that concepts should be conceived of as simulators. In other words, concepts are seen as re-enhancement of the pattern of neural activation recorded during perception and interaction with objects and entities. Thus, for example, the concept of "dog" would consist of the re-enhancement of the neural pattern that is active while observing a dog, caressing it, listening to its barking.

Claiming that concepts imply the re-enactment of sensorimotor experiences that pertain different modalities (vision, touch, audition etc.) entails that concepts are modal rather than amodal. The term "amodality", which is generally used within traditional theories of cognition, entails assuming the existence of a translation process from the experience (which is modal) to the mind (which would be amodal). Bodily experience would be translated into an abstract representation mode, sensory modalities would be transducted into propositional (simil-linguistic) symbols that are stored in our mind in an arbitrary way.

According to embodied theories of cognition, modality-specific systems play a central role in representing knowledge (Barsalou,1999; Damasio, 1989; Gallese & Lakoff, 2005; Martin, Wiggs, Ungerleider, & Haxby, 1996; Martin, 2001; Martin & Chao, 2001; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004). This does not entail that information is not distributed across different brain areas. In fact each concept property (e.g., visual, tactile, motor property) produces significant activation in the specific neural system (for example, the visual system becomes active while processing visual properties); furthermore it also produces multi-modal activation. It means that, rather than being restricted to the modality of the specific property, neural activity is multi-modal. For example, on verifying that 'keys can be jingling' (auditory property) not just neural auditory areas are active, but also the visual and the motor areas (Simmons, Pecher, Hamann, Zeelenberg, & Barsalou, 2003). Therefore the dominant brain activation for each concept property does not reside only in its respective modality. Thus, the nature of conceptual knowledge seems to be modal. But, even if we assume a not-amodal view, it is essential to distinguish between supramodality and multimodality (Gallese & Lakoff, 2005). Namely, the claim that concepts are modal is compatible with two different perspectives: the view according to which concepts are supramodal and that according to which they are multimodal.

The argument in favour of supramodality is based on the fact that the different modality-specific parts of our brain can be brought together via "association areas" (convergence zones according to Damasio, 1989). These association areas integrate information that pertain distinct modalities. In terms of neural network, supramodality could be represented by a feed-forward

network endowed with distinct modality specific layers (e.g., audition, vision, etc.) as well as by a layer that precedes the output layer in which information from the different modality-specific areas would be integrated. To claim that information is multimodal (Gallese & Lakoff, 2005; Fogassi & Gallese, 2004) implies that information is not represented in areas that differ from the sensorimotor ones. Therefore, no integration would occur in associative areas. Accordingly, multimodality could be represented by a feed-forward neural network in which no integration layer exists, and in which the layers that concern the different sensorimotor modalities are linked by bidirectional connections. Therefore, concepts would be directly grounded in the sensorimotor system, and different modality-specific areas would be activated during conceptual processing. The same is true for words, as words evoke their referents and re-enhance experiences with them. For example, the word "ball" would refer to a ball, and evoke its bouncing sound, its visual and tactile properties, the experience of throwing it.

3. Embodied models are necessary to reproduce experimental results on objects vision and action

Imagine sitting in front of a computer screen, in a quiet laboratory room. On the computer screen, after a briefly presented fixation cross, the photograph of a hand appears. The hand can display two different postures, a power grip, i.e. a grip adequate for grasping larger objects, such as bottles and apples, and a precision grip, adequate for grasping small objects, like pens and cherries. The image of the hand is followed and substituted by a photo representing an object graspable with a power grip (e.g., a hammer, an orange) or a precision grip (e.g., a pencil, a strawberry). The presented objects can be either artefacts (e.g., hammer, pencil) or natural objects (e.g., apple, strawberry). The task participants are instructed to perform is to press a different key on the keyboard to decide whether the object is an artefact or a natural object. Thus, the object size and its "compatibility" with the hand posture are not relevant to the task at hand. Interestingly, however, the "compatibility" between the hand and the object impacts response times. Namely, participants are significantly faster to respond when the hand and the object are compatible than when they are not, provided that, before the experiments, they are trained to reproduce with their own hands the hand posture they see.

The results found in this experiment (Borghi, Bonfiglioli, Lugli, Ricciardelli, Rubichi & Nicoletti, 2007) are in keeping with an embodied theory of concepts. Namely, they suggest that vision is strictly interwoven with action, as it evokes motor information. More specifically, the

results suggest that, upon viewing the image of a hand performing an action, a sort of motor resonance process is activated. Similarly, upon observing the image of an object, action-related information is activated. The results suggest that, when we see somebody acting with an object we "simulate" an action, and we take into account the possibility of a motor action even if we are not required to consider the congruency between the two visual stimuli we see, the hand and the object (for the notion of "simulation" see Gallese & Goldman, 1998; Jeannerod, 2007).

How could we model such a result? Would it be possible to reproduce such a situation using a model that is not embodied, i.e. with a simulated organism that is not endowed with a visual and a motor system? We believe it is not possible. Namely, modelling such an experiment with organisms that are not endowed with a neural system and a sensorimotor system roughly similar to the human ones, will impede to capture important features of the human cognitive and sensorimotor systems. Therefore when we talk about embodied models we don't refer to generic mathematical models in which the output is obtained on the basis of equations between input and output variables. Rather, we refer to *neural networks models*, as these models could offer more possibilities than the traditional mathematical models to reproduce the functional and physiological aspects of the brain. Some well known characteristics of the real nervous system as robustness, flexibility, generalization, recovery based on the content, learning, parallel processing, can be reproduced by neural networks models and sometimes also by some mathematical models. However, using neural networks models it is possible to obtain an emergent behaviour that leads the artificial brain to auto-organize its different parts, exactly as it happens in the real brain. Finally, the elaboration of information in the real nervous system is distributed, as there are many neurons involved in the same operation and a single neuron can be involved in different operations at the same time or at different times. Neural networks models well reproduce this distributed elaboration of information.

3.1 An example of an experiment with a possible model

Consider a situation which is easier to model. Participants observe objects (artefacts or natural objects) that can be graspable either with a power or a precision grip. The task consists in deciding whether the objects are artefacts or natural objects by mimicking a precision or a power grip with a sort of joystick. The authors find a compatibility effect between the object size and the kind of grip used to respond (Tucker & Ellis, 2001; Ellis & Tucker, 2000). In other words, participants' response times are faster with a precision than with a power grip when they see cherries, strawberries, pens etc., and they are faster with a power than with a precision grip when they see apples, hammers, bottles etc. The results indicate that seeing an object activate information about how to grasp it, even if this information is not relevant to the task, which is simply a categorization

one. They confirm that vision and action are not separated, but that vision incorporates a sort of simulated action.

How can we model experimental results like the described ones? Could we use simple feed-forward neural networks? Feed-forward models probably do not provide an adequate formalization for embodied theories. Namely, a great lesson of embodied theories concerns the reciprocal and circular relationship between perception, action, and cognition. The problem with feed-forward model is that they risk reproducing the traditional "sandwich" of disembodied cognitive sciences. The metaphor of the sandwich expresses clearly how perception and action were considered in traditional theories: they were intended merely as peripheral parts, having a scarce influence on the most tasty part, the inside, that is cognition. Similarly to a sandwich, feed-forward neural networks are typically endowed with one or more input layers, one or more hidden unit layers and one or more output layers. However, they are not characterized by a kind of circularity among the layers which do not have recursive character.

What kind of models should be used, then? First of all, we shouldn't use dis-embodied models, i.e. model that do not take into account the fact that cognition emerges from bodily experience. In the recent years within robotics there is a trend towards humanoid robots, i.e. robots that share sensorimotor characteristics with human beings. At the same time, it is clear that for the foreseeable future there will still be substantial differences in physical embodiment between robots and humans. Therefore, it is important that models are endowed with at least some characteristics of human reaching and grasping system, and that their sensorimotor system is at least roughly similar to the human one.

In addition, it is important that models possess at least some characteristics of human neural systems. For example, studies on concepts and language grounding might have a neurophysiological basis in the recent discovery, first in monkey and then in humans, of two kinds of visuomotor neurons: canonical and mirror ones (see Gallese, Fadiga, Fogassi & Rizzolatti, 1996). Canonical neurons discharge for motor actions performed using three main types of prehension: precision grip, all finger prehension, and whole hand grip. Interestingly they fire also to the visual presentation of objects requiring these kinds of prehension, even when grasping movement is not required. Mirror neurons instead fire when the monkey makes or observes another monkey or an experimenter performing a goal-directed action. In addition, recent studies propose that Prefrontal Cortex (PFC) is as an important source of top-down biasing where different neural pathways, carrying different sources of information, compete for expression in behavior (Miller and Cohen, 2001).

Experimental results like those obtained by Tucker and Ellis (2001) should be modelled by taking into account the role played by canonical neurons as well as by the PF cortex. In synthesis: an appropriate model of experiments such as the reported ones should be an embodied model, endowed with at least some crucial characteristics of human neural structure (neural network), and it should be able to replicate the behavioural results found.

As an example, we will describe a neural network model that suggests an interpretation of the results by Tucker and Ellis (2001) in light of the general theory on the functions of Prefrontal Cortex (PFC) proposed by Miller and Cohen (2001) (Caligiore, Baldassarre, Parisi & Borghi, *in preparation*). It has been simulated an artificial organism endowed with a human-like 3-Segments/ 4-Degree-Of-Freedom arm and a 21-S/19-DOF hand, and with a visual system composed of an "eye" (an RGB camera) mounted above the arm and looking down to the arm working plane. The organism's brain is a neural network (*see* Fig. 1) formed by maps of dynamical leaky neurons (Erlhagen and Schöner, 2002) that represent input and output signals on the basis of population codes (Pouget & Latham, 2003). The neural controller of the hand acts only on the thumb and on a "virtual finger" corresponding to the four other fingers.

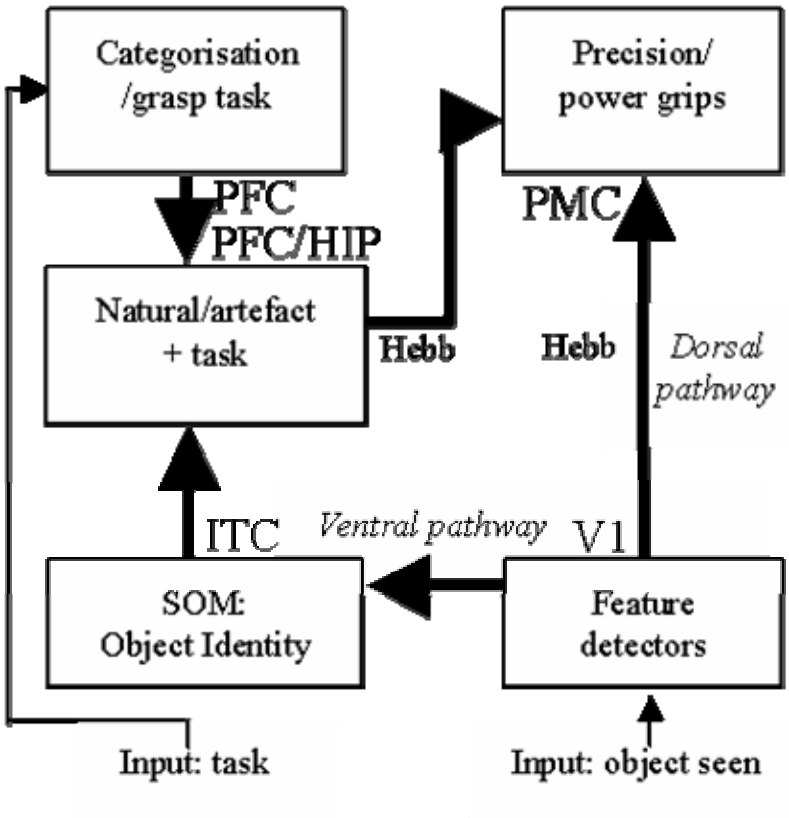

Figure 1. The architecture of the neural controller.
The *Dorsal Pathway* is formed by two neural maps, the Primary Visual Cortex map (V1) that encodes the shape of the foveated objects through edge detection filters mimicking simple cells functions, and the Pre-Motor Cortex map (PMC) that encodes the finger postures corresponding to precision/ power grip macro-actions.
The *Ventral Pathway* includes three neural maps: the Inferior Temporal Cortex map (ITC) encodes objects' identity, the Pre-Frontal Cortex map (PFC) encodes the current task (grasp/categorise), and the Pre-Frontal Cortex/Hippocampus map (PFC/HIP) encodes the representations of the task-dependent reactions to be associated to the various objects.

During the learning phase the V1-PMC connection weights are developed using an Hebb learning rule while the organism performs randomly-selected grasping actions ("motor babbling"), which mimics the acquisition of common experience, in particular the associations between big/small objects (V1) and power/precision grips (PMC), respectively. Within the ventral pathway, the V1-ITC connection weights develop the capacity to categorise objects on the basis of a learning rule that produces a self-organised map (SOM). Through an Hebb rule the PFC/HIP map learns object/task categories on the basis of input patterns from ITC and PFC. The PFC/HIP-PMC connection weights are formed using a Hebb rule which mimics the acquisition of knowledge on the task to be accomplished (e.g., during psychological experiments).

After the learning phase when the organism sees an object (e.g. a hammer) this activates the Primary Visual Cortex (V1). V1 transmits information to two primary pathways: the Dorsal stream and the Ventral stream. Firmed up that the processing of visual information is task-dependent, Dorsal Neural representation is mainly endowed with "action/perception" functions, Ventral Neural representation with "semantic" functions (Milner & Goodale, 1995). Therefore, the Dorsal Neural Pathway tends to trigger an action on the basis of the affordances elicited by the object (e.g. a power grip) in PMC. If the task is categorising objects, the Ventral Pathway might instead evoke a different action (e.g. a precision grip in the "incompatible" condition of the target experiments) through a bias coming from the PFC, so causing a competition between the two different tendencies within PMC. As the model assumes that the Ventral Pathway/PFC signal is stronger, this will cause this signal to finally prevail (so triggering a precision grip), but the reaction times will be slower with respect to the compatible condition.

The model successfully reproduces the experimental results of Tucker and Ellis (2001) and also allows interpreting their results on the basis of the role of PFC as an important source of top-down biasing. In particular, the model shows how the PFC bias can cause organisms to perform actions different from those suggested by objects' affordances. However, affordances still exert their influence on behaviour as reflected by longer reaction times.

4. Embodied models are necessary to reproduce experimental results on language grounding

The evidence we reported so far concern vision and categorization of objects. In this chapter we will consider what happens when objects are referred to by words. According to embodied views, not only concepts but also words are grounded. Thus, for example, the word "dog" would index its referent, a dog, and re-enhance the sensorimotor experience with dogs. According to this view,

during language processing the same systems used for perceiving and acting would be activated. For example, while comprehending an action sentence we would "simulate" the situation it describes, activating the same neural substrate used in action.

Now imagine to complicate a bit the previously described experiment, and to introduce language. Imagine participants sitting in front of a computer screen, and reading sentences that appear on the screen. Sentences describe an object in a given position, as for example "There's a kangaroo in front of you". Once they have read the sentence, participants press a key and a noun appears; their task consist in deciding by pressing a key whether the noun refers to an object part (e.g., head, legs) or not (e.g., wood, jump). The selected parts are located either in the upper part of the object (e.g., head) or in its lower part (e.g., legs). In order to respond, in the first part of the experiment participants are required to move upwards to press a key to respond "yes, it's a part" and to perform a movement downwards to respond "no". In the second part of the experiment the mapping is the opposite (yes = downward movement, no = upward movement). The results show that there is a compatibility effect between the part location (upper parts, lower parts) and the direction of the movement to perform (upwards, downwards). In other words, it is faster to move upwards in order to respond to upper than to lower parts, and it is faster to move downwards to respond to lower than to upper parts (Borghi, Glenberg & Kaschak, 2004).

This suggests not only that object parts incorporate some sort of motor representation, but also that processing a part noun implies activating the motor system. Not only objects, but words are grounded. A number of studies in the last years have shown that visual and acoustic inputs activate motor information. In the very last years, an increasing body of evidence indicates that this is not the whole story: it has been shown that language comprehension makes use of the same neural systems used for perception, action and emotion.

Since the seminal paper by Rizzolatti and Arbib (1998) on the relationship between language and motor system, a number of studies on canonical and mirror neurons have shown that these neurons might provide the neural basis underlying the language comprehension mechanism. Even if both behavioural and neural demonstrations have been provided, some issues remains to be clarified. In particular, it is still unclear the role played by canonical and mirror neurons during language processing. A recently advanced proposal is that mirror neurons might be mostly involved during verb processing, while canonical neurons would be primarily activated during noun processing (Buccino, *personal communication*).

In the previous study the effect of language on motor system was investigated using reaction times and movement times. In the next study we *directly* address if the mere act of comprehending language affects the production of action, focusing on body kinematics parameters (Scorolli, Borghi

& Glenberg, *in preparation*). In particular, we study a bimanual lifting action. In such an action, the motor pattern is modulated by perceptual visual cues, such as object's orientation, size and shape. Orientation is an extrinsic property of the object, as it depends on the observer, and/or on the observation conditions. Instead size and shape are intrinsic properties, i.e. invariant object features. Mass is another intrinsic object property that, unlike the other ones, cannot be visually detected, because it is intrinsically linked to the action, it originates from the interaction with the object. Thus mass is a suitable property to address with kinematics measures the characteristics of the simulation activated by language. In the study participants are standing in a quiet laboratory room, with their feet on a fixed point of the floor. They listen to sentences describing the lifting of different objects (e.g., "Move the pillow from the ground to the table"). Objects to which the words refer could be 'heavy' or 'light' (e.g. "tool chest" *vs.* "pillow"), but they do not vary in size and shape. After listening to the sentence, participants are required to lift a box placed in front of them, and to rest it on a pedestal. The box can be 'heavy' (mass of 12 Kg) or 'light' (mass of 3 Kg). The kinematics of the body movements is recorded using a motion capture system. The apparatus is made up of three cameras (acquisition frequency: 50-60Hz) catching sensors movements. Sensors are placed on the outside of the participants' body, mainly on joints position. After the box lifting, participants are asked a comprehension question (e.g. "Is the object on the table soft?"), in order to verify whether they have listened and comprehended the sentence. The production of action is addressed focusing on the first positive maximum of elbow angular velocity, detected immediately after having grasped the box. The results show that participants are faster in case of correspondence between the weight object suggested by the sentence and the *relative* weight of the actually lifted box. As previously described, an appropriate motor lifting pattern is shaped by objects visual features as size, shape and orientation, which in the present study are constant across the experiment. After grasping an object (box), the movement is adjusted depending also on object mass. If language did not have any effect on motor system, changes in biomechanical parameters should have been determined only by the actual object weight. Instead we found that listening to sentences about lifting light objects lead to higher velocity values (in extending) on the actual lifting of light boxes compared to the lifting of the same boxes preceded by heavy sentences. Symmetrical results are obtained for the heavy boxes. Thus, the lifting simulation activated by language modulates the applied force, indirectly detected by velocity parameters. Therefore the results show that the simulation activated by language is sensitive to an intrinsic object property such as mass.

4.1 Examples of possible models

How could we model and reproduce the results of the experiment we have illustrated? Most common models of language comprehension, developed within the cognitivist tradition, are based on association frequency. One of the most influential models, Latent Semantic Analysis (Landauer & Dumais, 1997), explains word meaning in terms of the associations between one word and other words in large corpora. The higher the index of co-occurrence of words in similar texts, the higher their similarity in meaning. Even if it represents a very useful tool, this model fails when it claims to represent conceptual meaning formation. Namely, it does not take into account that words are grounded in our sensorimotor system, as it only considers the network of verbal associations in which words are embedded. For these reason a model like this cannot capture and cannot predict the fact that an upper part is processed faster when moving upwards than when moving downwards. Namely, these results cannot be simulated with a model that does not possess at least some features of human sensorimotor system – some form of proprioception, some kind of sensitivity to sensory inputs. In addition, the simulated organism should possess at least some features of the human motor system (e.g. reaching and grasping "devices" like an arm, a hand with fingers, etc.). For the experiment in which kinematics measures were used, the simulated body should be quite complex, in order to reproduce a lifting movement. In addition, this model should be endowed with the capacity to comprehend language by referring words to its sensorimotor experience.

Only an embodied model can reproduce the experimental results we described. Namely, reproducing the results of an experiment does not simply mean to model just a behavior performed by a decontextualized brain simulated through a neural network, but to reproduce the behavior of an organism endowed not only with a brain but also with a body. In an ideal condition, this model should be able to reproduce learning though a mechanism of weight selection that avoids a-priori hardwiring any inhibitory or excitatory connections between or within modules. Using an embodied model has two further advantages. First, it allows one to observe rather than to infer the results. Finally, it allows reproducing the real structure of the experiment we aimed to simulate – for example, the actual button reaching behaviors could be reproduced.

A possible example of such a model is given by a self-organizing architecture developed within the Mirrorbot project and based on data on language grounding (Pulvermüller, 1999; 2003). In this project the model "takes as inputs language, vision and actions … [and] … is able to associate these so that it can produce or recognize the appropriate action. The architecture either takes a language instruction and produces the behaviour or receives the visual input and action at the particular time-step and produces the language representation" (Wermter, Weber, Elshaw, Panchev, Erwin, & Pulvermüller, 2004; Wermter, Weber, Elshaw, Gallese, & Pulvermüller, 2005). The model executes

predefined actions; while performing them, it learns associations between vision, action and language. Self-organizing artificial neural networks are used to associate sensor and actor data.

5. Embodied models can help to formulate clearer predictions

In the previous part of the chapter we have shown that, in order to model experimental results on concepts and language grounding, embodied models are necessary. First, models based on the assumption that the mind is a device for symbol manipulation are in contrast with embodied theories from a theoretical point of view. In addition, from a methodological point of view they are not able to capture the richness derived from the experimental setting and the experimental results.

In this part of the chapter we intend to show that embodied models might be very useful in order to formulate more detailed predictions that might help to disentangle unsolved issues within cognitive science. In the last years a hotly debated issue concerns the kind of relationship existing between language and the motor system. We will label it the "Interference or Facilitation" (IF) issue. As before, we will make an example in order to clarify what it's all about.

Consider the following study (Scorolli & Borghi, 2007). Participants are sitting in a quiet laboratory in front of a computer screen. After a fixation cross, a verb appears on the screen (e.g., to kick), and then it is substituted by a noun. Verbs refer to actions typically performed with the foot and with the hand (e.g., to kick *vs.* to throw *the ball*), or with the mouth and with the hand (e.g., to suck *vs.* to unwrap *the sweet*). The task consists in deciding whether the combination between the verb and the noun makes sense or not. For example, the combination "to lick the ice-cream" makes sense, while the combination "to kill the pot" doesn't. In order to respond that the combination makes sense, in one condition participant are required to press a pedal with a foot, in another condition they are instructed to respond "yes" on the microphone. If the combination doesn't make sense, they have to refrain from responding. The results show that responses are faster in case of correspondence between the effector used to respond and the verb-noun combination to process. Thus, responding with the foot to 'foot sentences' is faster that responding by pressing a pedal to 'mouth sentences'; the opposite is true with responses with the microphone.

The results suggest that, when we comprehend an action sentence (or a verb noun combination), we "simulate" the described situation, and our motor system is activated. This simulation is sensitive to the kind of effector implied by the action sentence; namely, responses are faster when the effector implied by the sentence and the effector used to respond are the same than when they are not. This is in keeping with embodied theories, as it argues for a strong link between

language processing and the motor system. However, why should we predict that, say, reading a foot sentence leads to facilitation in foot responses rather than to interference? If during language processing and action executing the same neural structures are activated, then this might slow down responses. In the literature studies performed with tasks that slightly differ from the one we described report an interference effect (Buccino, Riggio, Melli, Binkofski, Gallese, & Rizzolatti, 2005).

The IF issue consists in the fact that it is difficult predicting whether an interference or a facilitation effect will occurr. The problem is further complicated by the fact that both interference and facilitation are compatible with an embodied account. Namely, they both indicate that language is grounded, and that reading sentences leads to an activation of the motor system. However, so far it is difficult to make accurate predictions about the direction of the interaction between language and motor system. Models can help to further detail the predictions, for example they can contribute to solve the IF issue.

6. Tentative conclusions and open issues

In this chapter we have tried to show that embodied theories of cognitive science badly need models, and that these models need to be embodied. In the first part we illustrated some behavioral experiments and showed that they can only be reproduced with embodied models. First, we reported experimental results that show that seeing an object activates its affordances, as it evokes potential actions to perform with it. Then, we described experimental evidence indicating that reading words and sentences activates the motor system – for example, reading the word "head" activates an upwards movement –. In both cases we showed that the behavior examined and the results found cannot be modelled with computational models that assume that the mind is a mechanism for symbols manipulation. Rather, adequate models can only be given by embodied artificial organisms that are endowed with neural and sensorimotor structures that at least roughly reproduce human ones.

In the second part we illustrated behavioural experiments leading to ambiguous results. As an example, we illustrated the controversial results that concern the relationship between language and the motor system – in some cases processing action-related words facilitates movements that are compatible with them, in other cases it renders them more difficult (interference vs. facilitation issue).We showed that embodied models can be a powerful means that helps to disentangle ambiguous issues and to formulate clearer predictions. The embodied cognition field has greatly

expanded in the last years. In the last ten years much experimental evidence has been collected, but now it is crucial to formulate more detailed and precise predictions. Embodied cognition field badly needs for well specified theories, and models can help to formulate these theories.

One last issue is worth of notice. At a very general level, embodied models might provide a powerful way to test embodied theories of cognition. Namely, comparing models whose physical (neural, sensorial, motor) structure is more or less similar to the human one, will allow understanding to what extent possessing the same kind of "body" is necessary in order to understand the world and to comprehend and to use language.

From a theoretical point of view, the assumption of a very strong embodied view would lead to the avoidance of any kind of direct comparison between experiments and computer simulations. Namely, a strong embodied view could predict that only models that share the same bodily characteristics with the entities they have to reproduce (human beings) can be adequate to explain them. From this claim might derive the choice to consider the artificial world as a parallel word, with its own laws, that should (and could) not be compared with the world of human beings.

Our position is that this claim can be made less strong, and a more mild embodied view can be adopted. Namely, it is possible that a certain degree of similarity between humans and their embodied models can allow capturing important aspects of human cognition and behaviour. One of the fascinating questions the research of the next years has to face is the following: to what extent do we need to be similar in body in order to share a common view of the world, and to communicate with others? And again: to what extent (and for what aspects) do models need to resemble to humans in order to be considered good models?

References

- Barsalou L. W. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences, 22***,** 577-609
- Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., & Martin, A. (2004). Unraveling Multisensory Integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7*, 1190-1192.
- Borghi, A., Bonfiglioli, C., Lugli, L., Ricciardelli, P., Rubichi, S., & Nicoletti, R. (2007). Are visual stimuli sufficient to evoke motor information? Studies with hand primes. *Neuroscience Letters, 411,* 17-21.
- Borghi, A.M., Glenberg, A., & Kaschak, M. (2004). Putting words in perspective. *Memory and cognition,32,* 863-873.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research, 24*, 355-63.
- Caligiore, D., Baldassarre, G., Parisi, D. & Borghi, A.M. (in preparation). Affordances and compatibility effect: a computational model.
- Damasio, A. R. (1989).Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*, 25–62.
- Ellis, R. & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology, 91,* 451-471.
- Erlhagen, W., Schöner, G. (2002). Dynamic field theory of motor preparation. *Psychological Review*, *109*, 545-572.
- Fodor J. (1975). *The language of Thought*. Harvard University Press, Cambridge (MA).
- Fogassi, L., & Gallese, V. (2004). Action as a binding key to multisensory integration. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes*. Cambridge: MIT Press.
- Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Science, 2,* 493-501.
- Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensorimotor system in conceptual knowledge. *Cognitive Neuropsychology, 21*, 455-479.

- Jeannerod, M. (2007). *Motor cognition. What actions tell to the self.* Oxford: Oxford University Press.
- Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review,* 104, 211-240.
- Martin, A., Wiggs, C.L, Ungerleider, L.G., & Haxby, G.V. (1996). Neural correlates of category specific knowledge. *Nature, 379*, 649-652.
- Martin, A. (2001). Functional neuroimaging of semantic memory. In Cabeza R & Kingstone A (Eds), *Handbook of Functional NeuroImaging of Cognition,* 153-186. Cambridge: MIT Press.
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology, 11,* 194-201.
- Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*,167-202.
- Milner, A.D., & Goodale, M.A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Pouget, A. & Latham, P. E. (2003). Population codes. In M.A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, 2nd ed. Cambridge, MA, USA: The MIT Press.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioural and Brain Sciences, 22*, 253–336.
- Pulvermüller, F. (2003). *The neuroscience of language: On brain circuits of words and serial order.* Cambridge: Cambridge University Press.
- Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188-194.
- Scorolli, C., & Borghi, A. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain research*, 1130, 119-124.
- Scorolli, C., Borghi, A. & Glenberg, A.(in preparation). Effects of language on bimana object lifting.
- Simmons, W.K., Pecher, D., Hamann, S.B., Zeelenberg, R., & Barsalou, L.W. (2003) fMR evidence for modality-specific processing of conceptual knowledge on six modalities. *Meeting of the Society for Cognitive Neuroscience*, New York.
- Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8,* 769-800.

- Wermter, S., Weber, C., Elshaw, M., Panchev, C., Erwin, H., & Pulvermüller, F. (2004) Towards Multimodal Neural Robot Learning. *Robotics and Autonomous Systems Journal, 47, 171-175*.
- Wermter, S., Weber, C., Elshaw, M., Gallese, V., & Pulvermüller, F. (2005).Grounding Neural Robot Language in Action. In Wermter S., Palm G., Elshaw M. *Biomimetic Neural Learning for Intelligent Robots*, 162-181.