TRoPICALS: A Computational Embodied Neuroscience Model of Experiments on Compatibility Effects

Authors:

Daniele Caligiore*+^, Anna M. Borghi+*, Domenico Parisi*, Gianluca Baldassarre*

Affiliations:

* Laboratory of Computational Embodied Neuroscience,
Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR)
Via San Martino della Battaglia 44, I-00185 Roma, Italy
Email: {daniele.caligiore, domenico.parisi, gianluca.baldassarre}@istc.cnr.it

\+ Embodied Cognition Laboratory (EMCO-Lab)
Dipartimento di Psicologia, Università di Bologna
Viale Berti Pichat 5, I-40127 Bologna, Italy
Email: annamaria.borghi@unibo.it

^ Biomedical Robotics and Biomicrosystem Lab
Università Campus Biomedico
Via Álvaro del Portillo 21, I-00128 Roma, Italy

Keywords: embodied cognition; affordances; human and non-human primate cortex; dorsal and ventral cortical pathways; prefrontal cortex; dynamic field neural networks; language and simulation theory.

Running head: TRoPICALS: A model of compatibility effects

Abstract

Recent evidence shows that the perception of objects automatically activates the representation of their affordances. For example, some experiments found compatibility effects between the size of objects (small/large) and the kind of grip (precision/power) required to categorise them as natural or as artefacts, and between common location of object parts (top or bottom) and the kind of movement (up and down) required to indicate whether or not these parts belong to a whole object. This article presents a neural-network model that reproduces these results and also provides a general framework to account for several other types of compatibility effects. This model is based on four general principles: (a) visual perception and action are organised along a dorsal neural pathway (encoding affordances) and a ventral pathway; (b) within the ventral pathway, the prefrontal cortex biases action selection based on context and goals; (c) action selection results from neural dynamic competitions that cause variable reaction times; (d) words trigger "internal simulations" of their referents. The model was designed within a methodological approach that aims at developing it cumulatively so as to furnish increasingly general and comprehensive accounts of compatibility effects. The approach imposed four types of constraints on the model: (a) neuroscientific constraints on their architecture and functioning; (b) reproduction of specific psychological experiments; (c) functioning within an embodied system; (d) reproduction of the learning processes that result in the target behaviours. The claim on the generality of the model is supported by a critical comparison with other models that are related to the above four principles and by an analysis of how the model could be developed to account for other compatibility effects. The heuristic power of the model is also shown by presenting two testable predictions.

TRoPICALS: A Computational Embodied Neuroscience Model of Experiments on Compatibility Effects

Title level 1: Introduction

Title level 2: The role of actions in the internal representation of objects

According to traditional views of cognition, perception precedes action and is not influenced by it. Sensory input determines how the world is represented in an organism's nervous system whereas processes underlying action only play a role in how the organism modifies the environment. This passive, "contemplative", view of knowledge, that has a long philosophical tradition, is increasingly challenged by recent studies which suggest that visual perception and action are closely interdependent and that the way organisms represent the world depends not only on sensory stimuli per se but also on the actions with which they respond to them (Gallese & Metzinger, 2003; Barsalou, 2003).

Behavioural and brain imaging evidence indicates that visually perceived objects activate motor information. Seeing objects or pictures of objects elicits the actions that tend to be performed on or with objects. In this perspective, the notion of *affordance* (Gibson, 1979) has been given new relevance. An affordance can be described as a set of properties of an object which suggest possible actions and uses to an organism. An affordance is not an intrinsic property of an object, but rather a relational property: an object may provide different affordances depending on the features of the organism body and of the context.

Jeannerod (1994) and Arbib (1997) proposed that objects tend to evoke actions which are appropriate to them and suggest that this process involves motor representations and their distal goals. Many behavioural and brain imaging studies on humans and animals have provided evidence showing that observation of objects activates their associated affordances. Rizzolatti, Fogassi, and Gallese (1997) conducted a study with monkeys illustrating that the sight of objects tends to automatically evoke the activation of suitable neural representations in premotor cortex, a brain district believed to underlie action preparation (for example in relation to precision or power grip actions). However, the extent to which this activation occurs in an automatic bottom-up way or is modulated by the task itself is subject of debate (e.g., see Borghi et al., 2007; Castiello, 1999; Creem & Proffitt, 2001; Creem & Lee, 2005; de' Sperati & Stucchi, 1997).

The rest of the paper is organised as follows. The next two sections introduce the target experiments investigated with the model, and explain the goals of the work and the method used to build the model ("computational embodied neuroscience"). The succeeding group of sections illustrates the four general principles at the core of the model, the details of the three target experiments, and the body of the simulated robot used to reproduce them; moreover, they present in detail the architecture and functioning of the model, the neuroscientific evidence used to constrain them, and the learning processes used to train some components of the model. The article then presents the results of the tests directed to reproduce the three target experiments, their interpretation, and partial conclusions; these sections also report two specific predictions of the model which might be tested in real experiments. A further group of sections extensively reviews other models which are related to the principles incorporated by the model, and shows how the model, if suitable developed, has the potential to account for many other compatibility effects (these sections can be skipped by the reader if not interested in these issues as they are rather independent of the rest of the paper). The article is closed by summarising the main achievements of the model.

Title level 2: Affordances and compatibility effects investigated with the model

One way of studying how internal representations of objects are related to action representations is to devise experimental tasks in which participants are exposed to objects and are asked to produce actions which are either in agreement (“compatible” or “congruent” trials) or in contrast (“incompatible” or “incongruent” trials) with the actions typically associated with the objects (e.g., grasping an object with the appropriate grip). For example, in compatible trials the participant is asked to execute a grip appropriate to the object size (e.g., a precision grip with a small object), whereas in incompatible trials the participant is asked to execute a grip which is not appropriate to the object size (e.g., a precision grip with a large object). If the participants employ longer reaction times (RTs) and higher error rates in incompatible trials than in compatible ones, one can infer that seeing objects automatically elicits the representations of their affordances. With the term “automatically” we refer to the fact that the representation of affordances of objects is elicited independently of the performance of the experimental task.

Tucker and Ellis (2004) and Borghi, Glenberg, and Kaschak (2004) performed various experiments using a compatibility paradigm to investigate whether vision of objects or of words referring to them automatically activates action representations. In “Experiment 3, picture-mode” of Tucker and Ellis (2004), participants were asked to classify large and small image objects either into the category of artefacts or the category of natural objects. To perform the task they had to mimic either a precision or a power grip by acting on a customised joystick. Importantly, object size was not relevant to the task which simply requested categorizing the objects into “natural object” or “artefact”. The authors found a compatibility effect between the size of the object (large and small) and the motor response (power and precision grip). In particular, they observed shorter RTs in congruent trials (large objects and power grip; small objects and precision grip) than in incongruent ones (large objects and precision grip; small objects and power grip).

In “Experiment 3, word-mode” of Tucker and Ellis (2004), the authors investigated the role of motor information during the processing of written names of objects. In these experiments the authors observed again an interaction between object size (large, small) and motor response (power, precision), even if object size was not relevant to the categorization task. Congruent trials caused shorter RTs, whereas incongruent trials produced longer RTs. Overall, the results of Tucker and Ellis (2004) show that knowledge of objects, independently of the fact that it is activated by images or words, relies not only on their perceptual features but also on the actions that can be performed on them, so the latter affects the behaviour of participants irrespective of their current goals.

The “experiment 3” by Borghi, Glenberg, and Kaschak (2004) aimed at studying whether reading sentences referring to objects automatically activates representations of the spatial features of objects and evokes the corresponding motor actions. Participants read a sentence referring to an object, for example “There is a doll on the table in front of you”, followed by a noun which could either refer to a part of the object in the sentence (e.g., “head”) or to something completely unrelated (e.g., “kindergarten”). Participants were asked to decide whether the noun referred to a part of the object or not. In the “yes-is-up” condition, the participants were asked to click a button positioned at the top of a row of two buttons to replay “yes”, and a button positioned at the bottom of the row to reply “no”. In the “yes-is-down” condition the position of the buttons to be pressed was reversed. A compatibility effect was found when the object part was in the upper part of the whole object (e.g., “head”) RTs were faster in the “yes-is-up” than in the “yes-is-down” condition, as if upper parts evoked an upward movement. Similarly, object parts located in the lower part of the whole object (e.g., “feet”) evoked a downward movement producing the opposite effects. The experiment demonstrated that object names evoke information on the spatial location of object parts, and this affects motor responses, even if such information is not relevant for the experimental task.

Title level 2: Aim of the study, constraints on designing the model, and methodological approach adopted

The goal of this paper is to present a bio-constrained neural-network model which aims at providing a comprehensive computational account of experiments looking into compatibility effects. As a first step towards the achievement of this overall goal, the proposed model is used here to interpret the three experiments on compatibility effects and affordances illustrated in the previous section

(Tucker & Ellis, 2004; Borghi, Glenberg, & Kaschak, 2004).

The model has been designed on the basis of an approach which might be termed *Computational Embodied Neuroscience* as it is based on principles advocated by “embodied cognitive science” and “computational neuroscience” (other information on the approach can be found in Mannella, Mirolli, & Baldassarre, 2010; cf. Prescott, Montes-Gonzalez, Gurney, Humphries, & Redgrave, 2006, for a similar but less principled approach). In particular, the model was built by trying to comply with four classes of constraints: (a) constraints deriving from the available knowledge on the *anatomy and physiology of the brain structures* underlying the behaviours of interest (e.g., based on Rizzolatti, Luppino, & Matelli, 1998, Culham & Kanwisher, 2001, Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002, and Grèzes, Tucker, Armony, Ellis, & Passingham, 2003); (b) constraints deriving from the requirement to accurately reproduce the behaviours observed in specific psychological experiments, for example in this case those from Tucker and Ellis (2004) and Borghi, Glenberg, and Kaschak (2004); (c) constraints related to the requirement to reproduce the target behaviours *within an embodied system* (with “embodied system” we refer to a “simulated participant” which interacts with the environment via realistic sensors and actuators as it happens for the human participants of the target experiments; for example, the model presented here perceives objects in terms of realistic RGB images, similar to eye vision, and performs actions on the basis of a simulated 3D human-like realistic arm and hand); (d) constraints related to the requirement that the model should reproduce and explain the *learning processes* leading the human participants to acquire the behaviours of the target experiments.

Several methodological reasons motivated the use of these four classes of constraints. According to the *computational neuroscience* approach (Churchland & Sejnowski, 1992; Arbib, 2002), it is important to translate theories of brain and behaviour into detailed computational models as this forces the experimenter to *specify and make explicit* vague or implicit assumptions of theories and also allows deriving specific *falsifiable* and often *quantitative predictions* from these theories. In addition, the use of the first two types of constraints is essential to allow selecting and developing models in a cumulative fashion. In particular, the first type of constraints (anatomical and physiological constraints) are important as one can always produce a number of alternative models that reproduce a given behaviour (behavioural constraints), so one is left without criteria to rank and select the models. On the contrary, if one tries to design a model by deriving its architecture and functioning from detailed anatomical and physiological data, the range of possibilities for reproducing the target behaviours and cognitive processes drastically decreases. The support of theories of cognition and behaviour from neuroscientific evidence (particularly in the area of localisation of cognitive functions in the brain) is also a fundamental methodological principle endorsed by *cognitive neuroscience* (Posner, Pea, & Volpe, 1982; Posner & Di Girolamo, 2000). In line with cognitive neuroscience, one of the strong values of the model presented here, and of computational models in general, is that it accounts for cognitive and behavioural targeted phenomena in terms of low-level neural mechanisms. Explanations based on low-level mechanisms usually have the advantage of being more general and of offering more predictions than the explanations provided at a higher functional level.

The third type of constraints (“embodiment” constraints) are in line with *embodied cognitive science* and the *simulation of adaptive behaviour* approaches (Meyer & Wilson, 1991; Varela, Thompson, & Rosch, 1991; Webb, 1995; Webb, 2009). These state that a real understanding of cognitive processes can only come from computational models which view organisms as entities possessing an entire body (Brooks, 1989) and engaging in interactions with a realistic environment through realistic sensors and actuators (Clark, 1997). As we shall see, for example, these type of constraints allow the model presented here (a) to have realistic differentiated internal representations of objects in the dorsal and ventral pathways and (b) to control actions on the basis of specific sensorimotor neural transformations. This contributes to explaining some of the results of the target experiments.

In line with embodied cognitive science, and also with the *Artificial Life* approach*,* the fourth source of constraints (“learning” constraints) derive from the idea that a full understanding of behaviour requires not only explaining its final organisation but also the ontogenetic processes, and possibly also the phylogenetic processes, that lead to its development and evolution (Parisi & Schlesinger, 2002; Webb, 2009).

The approach used here is also guided by the fundamental "meta-constraint" of *theoretical cumulativity*. This meta-constraint consists to avoid producing ad-hoc models which account for only specific single experiments and to aim at producing general models that account for an *increasing number of experiments* related to a certain class of target cognitive, behavioural, and neural phenomena so as to incorporate a progressively larger number of constraints. This effort is expected to lead to *isolate the general principles* underlying the class of studied phenomena and hence to produce theoretical cumulativity. The importance of theoretical cumulativity resides in the fact that psychology and neuroscience are accumulating a huge body of evidence on detailed phenomena related to cognition, brain, and behaviour, but they often fail to produce coherent and unified theoretical accounts of these phenomena. We consider theoretical cumulativity the most significant added value that the computational models can produce with respect to our understanding of cognition, behaviour, and brain (the need to produce general theories is in line with what stated by other researchers, for example Newell, 1973). The model described in the paper reproduces only some specific experimental results but our emphasis on cumulativity and generality with which it has been designed implies that the same model can also be applied, with the necessary modifications, to other experimental and ecological behavioural evidence (see last sections on this).

A last important observation with respect to the method used here regards the neuroscientific detail of the model presented here. The model is a *system-level* model. This implies that, from a neuroscientific perspective, the main goal of it is to provide an operational hypothesis of the cerebral "network of networks" which underlies compatibility effects (i.e., the system formed by a network of brain areas each in turn formed by a network of neurons). This sides with the perspective adopted here which postulates that the different classes of behaviours are generated by the interplay of different subsets of components of the brain rather than by specific components in isolation. For this reason, the single components of the brain, and hence of the model, can be more accurately characterised in terms of computational *mechanisms* rather than *functions*. For example, the premotor cortex components of the model can be characterised in terms of the "neural competition" they perform. Whereas the function of "action-selection" they implement, ascribed to them in various parts of the paper for ease of reference and to aid understanding, depends on the information they get from afferent areas and on the effects they cause on downstream areas and motor systems. In line with this, the model takes into consideration ten brain areas whereas neural network models usually focus on the specific functions of two-three areas (although there are notable exceptions, e.g., Fagg & Arbib, 1998, and Oztop & Arbib, 2002). This allowed us to outline an integrated hypothesis on the system-level architectural and functioning brain mechanisms which might underlie compatibility effects. Indeed, all four of the principles which specify such a hypothesis capture rather high-level mechanisms.

A caveat, however, is that this approach led us to represent the single components of the model with neural networks (dynamic-field neural networks and Kohonen neural networks) which are rather abstract with respect to the brain micro-circuits and the functioning of neurons. In this respect, however, due to the general biological plausibility of the neural networks used to design the model (cf. Kohonen, 1997; Erlhagen & Schöner, 2002), it should be possible to improve the realism of the micro-architecture and functioning of the *components* of the model without changing much of its *overall architecture* and core functioning principles.

Title level 1: TRoPICALS: principles, architecture, functioning, and learning

Title level 2: The four core principles at the basis of the model

Figure 1 illustrates the general functions implemented by the model and the putative brain areas involved in such functions (these will be further explained below). The model is based on *four general principles*: (a) the *two-route* organisation of the visual system and, more generally, of an important portion of the brain, relying upon a ventral and a dorsal neural pathway; (b) the guidance of action selection on the basis of *prefrontal cortex "instructions"*; (c) the selection of actions on the basis of a *competition* between different *affordances* based on such prefrontal bias; (d) the capability of *language* to trigger *internal simulations* of the referents of words. For this reason, the model is called "TRoPICALS", an acronym which stands for the four principles: *Two Route, Prefrontal Instruction, Competition of Affordances, Language Simulation* model (the acronym also hints at the potential of the model to be developed in a cumulative fashion so as to account for an increasing number of compatibility effects, similar to "tropicals", the tropical plants which tend to grow in a layered fashion).

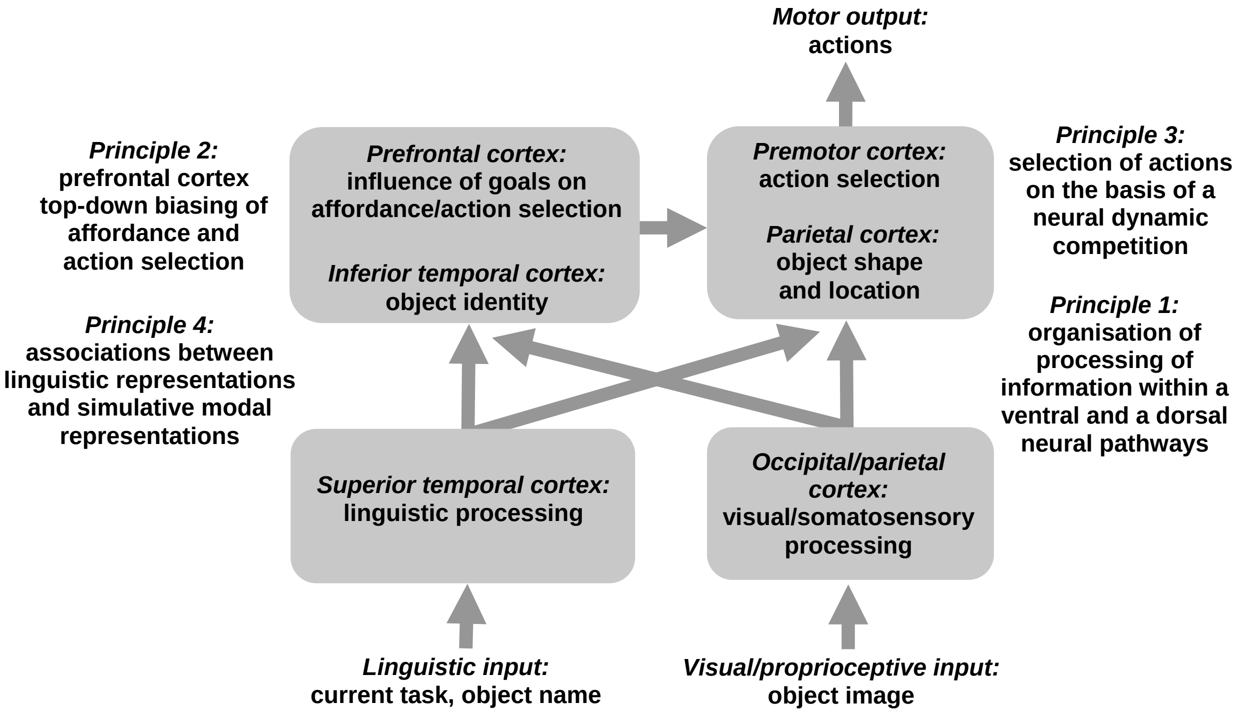

Figure 1: Abstract schema of the main functions implemented by TRoPICALS and the putative anatomical brain areas where they might be executed: (a) the occipital cortex and anterior parietal cortex which performs respectively visual pre-processing (e.g., edge detection) and somatosensory processing; (b) the inferior temporal cortex and the prefrontal cortex which respectively extracts information on identity of objects and biases action selection processes taking place in premotor cortex based on the current context and goal; (c) the parietal cortex and the premotor cortex which respectively extracts affordance information from objects and selects the desired final posture of the arm fingers and (sent to the muscle models, not shown in the figure); (d) the cortical areas involved in language processing, such as the superior temporal cortex involved in auditory processing and language comprehension. The graph also indicates the four main principles incorporated by the model

Emphasised heading of paragraph: First principle: two neural pathway organisation

This principle concerns the broad organisation of brain cortex responsible for visual processing into the *dorsal* and *ventral neural pathways.* This theory was initially proposed by Ungerleider and Mishkin (1982). In the original proposal, the *ventral stream* is a neural pathway which runs from early visual cortex areas (VC) to inferotemporal cortex (ITC) and carries information about the identity of objects ("what" pathway). The *dorsal stream* is a neural pathway which runs from VC to the parietal cortex (PC) and processes spatial information, concerning the location of objects in the visual field ("where"). The scope of this theory was later extended (Goodale & Milner, 1992; Milner & Goodale, 1995; Milner & Goodale, 2008) by proposing that the ventral stream communicates visual information to support higher cognitive processing taking place in downstream cortical areas (e.g., not only object recognition but also decision making on actions to be executed, and higher-level reasoning). In contrast, the dorsal stream transfers visual information to support on-line performance of actions in downstream motor cortex areas (e.g., not only location of objects, but also identification of their shape and implementation of the sensorimotor transformations needed for visually guided actions). Note that although the ventral and dorsal components are rather segregated their segregation is not complete. In fact the information processed by the two pathways is partially integrated not only in the late frontal stages of sensorimotor information flow, but also through connections existing between early visual stages (Rolls & Deco, 2002) and in the PC (Gallese, Craighero, Fadiga, & Fogassi, 1999).

Based on the above ideas, the architecture of the model includes a first component that corresponds to the sensorimotor dorsal pathway. This pathway is formed by a first group of three "maps" of artificial neurons abstracting the visual pre-processing performed by VC. Then it is formed by two other maps which abstract some of the functions implemented by the PC, here responsible for extracting the shape and location of objects. Finally, it is formed by two further maps which abstract some of the functions implemented by the premotor cortex (PMC), in particular they contribute to the selection and preparation of grasping and reaching actions. Overall, the dorsal pathway is assumed to encode the "affordances" of objects, i.e. to implement information processes related to actions which can be successfully executed on objects.

In primates, information on affordances can be learned during the first months of life, and refined later in life, based on processes such as "motor babbling" (von Hofsten, 1982; Ognibene et al., 2006; Caligiore et al., 2008), also called "direct inverse modelling" (Kuperstein, 1988). This process consists in the production of various motor acts which allow the acquisition of basic useful associations between representations in sensorial and motor areas (Piaget, 1952). In the model presented here, motor babbling processes are used to mimic what might happen in early stages of development. In these stages, for example, children learn to form associations between the perceived position in space of an object held with an hand and the corresponding posture assumed by the arm and the hand holding it. The knowledge acquired in this way allows the performance of goal-directed reaching movements in later stages of life (e.g., in psychological experiments as those simulated here). In fact the selection of a target with eyes allows retrieving the internal representation of the correspondent posture of the arm and hand and this can then guide movements towards such posture.

The skills implemented by the dorsal pathway have the advantage of allowing highly familiar motor behaviours to be executed quickly and automatically (for example, a simple automatic manipulative behaviour such as the tendency to grasp a large object with a power grip and a small object with a precision grip) and can develop gradually as learning mechanisms form suitable cortical patterns. These skills are rather rigid and stereotyped, for example they tend to always lead to the automatic execution of actions corresponding to the most active affordances. In this respect, Borghi and Riggio (2009) have proposed that humans activate object motor prototypes corresponding to the most frequently used affordances (e.g., the size and key parts of objects, such as the handles).

Emphasised heading of paragraph: Second principle: prefrontal cortex instructions

This principle concerns the role of the prefrontal cortex (PFC) as *a source of top-down biasing* which instructs and informs the neural competitions between potential actions that compete for expression in behaviour (Miller & Cohen, 2001; Cisek, 2007). The sensory system of primates provides detailed information about the external world and as a result of this their motor system can acquire a large repertoire of actions. This introduces a great potential for flexibility but also for interference. To effectively cope with the multitude of possible actions to perform, the brain has acquired mechanisms that coordinate low-level sensory and motor processes on the basis of internal motivations, goals, and external context (Fuster, 1997; Fuster, 2001). PFC plays a key role in these processes especially when "top-down" control based on motivations and goals is needed (Fuster, 2001; Miller & Cohen, 2001; Wallis, Anderson, & Miller, 2001).

PFC also exhibits a high degree of plasticity. At the neural level, the mechanisms that underlie this plasticity could involve the modification of synapses, possibly with the assistance of rapid learning processes taking place in the hippocampus (Rolls & Treves, 1998). PFC can perform these functions in a more flexible way than lower sensorimotor processes, for example it can learn behavioural sequences "on the fly" (in humans it can do this on the basis of reciprocal interconnections with language areas, Pulvermüller, 2005) and it can store context in working memory so that it can suitably switch the "rules" of the produced behaviour in correspondence to relevant events (Hazy, Frank, & O'Reilly, 2007; Wallis, Anderson, & Miller, 2001).

This higher flexibility and context sensitivity of PFC allows it to strongly bias action selection when the multimodal richness of information and external/internal context requires to suitably modulate action instead of acting on the basis of habits and automatic tendencies (e.g., see also the literature on goal-driven behaviour, e.g. Balleine & Dickinson, 1998, and Daw, Niv, & Dayan, 2005). To this purpose, it is important that PFC can rely upon mechanisms that allow it to overcome the low-level automatic resonances when needed (e.g., the model proposed here uses strong "top-down" connections linking PFC to motor areas).

Emphasised heading of paragraph: Third principle: affordance competition

This principle states that decision making and action selection are processes which rely on brain cortex connectivity such that neurons taking part in common representations are linked through excitatory connections whereas neurons belonging to different representations are linked through inhibitory connections (Ts'o, Gilbert, & Wiesel, 1986; Georgopoulos, 1995; Sparks & Groh, 1995; Bastian, Riehle, Erlhagen, & Schöner, 1998). Due to this connectivity, neuron clusters encoding different alternatives compete against each other and only the cluster receiving the strongest signals from other brain areas within a certain time frame survives. These signals carry "evidence" (or "support") in favour of the different clusters whereas the competitive neural mechanisms work as a "clearing" mechanism which allows one cluster to prevail on others.

This view of the functioning of decision making and action selection is in agreement with the neurobiological findings of Cisek and Kalaska (2005) who show that competing actions elicit competing activations at the level of frontal cortex until the activation of the selected action is the only one to remain. Cisek (2007) proposes that such competition is based on a biasing effect produced by the prefrontal cortex (see second principle) and basal ganglia. In this respect, note that the actual competition based on reciprocal inhibition of competing clusters might actually take place within the frontal cortex (Cisek & Kalaska, 2005), within the basal ganglia (Redgrave, Prescott, & Gurney, 1999), or within whole systems formed by basal ganglia/cortical loops (including PFC, Kandel, Schwartz, & Jessel, 2000). For simplicity, we will refer to the competition as taking place within the cortex but this does not intend to exclude the other possibilities.

Brain processes underlying action selection and decision making can be reproduced with the *dynamic neural field models*. These models are usually implemented on the basis of neural maps (or "neural fields"; Amari, 1977; Erlhagen & Schöner, 2002). Such models are strongly related to the "population code models" (Zemel, Dayan, & Pouget, 1998; Tipper, Howard, & Houghton , 2000; Pouget & Latham, 2002), also usually based on neural maps, and to the "leaky competing accumulator models" (Usher & McClelland, 2001; Bogacz, Usher, Zhang, & McClelland, 2006), which implement the principle of the competition based on reciprocal inhibition between discrete neural representations instead of neural fields. An important outcome produced by the use of these types of models to mimic decision making and action selection is that they naturally exhibit RTs which can be compared with empirical data. This is an important advantage for modelling because in cognitive psychology the independent variables measured in experiments are often the reaction times of participants.

Emphasised heading of paragraph: Fourth principle: language-triggered simulations

This principle is related to the language associative mechanism with which the model starts to give an account of the compatibility effects found with object names. This mechanism is implemented through a Hebbian correlation learning rule which creates associations between arbitrary patterns of active neurons representing the phonological aspects of words and internal "simulations" (e.g., the representations of the categories of objects and the representations of the aspects of objects that guide action, such as their shape and location). Note that the notion of "simulation" has been defined in different ways (for a detailed analysis of this concept, see Borghi & Cimatti, 2010, and Pezzulo, 2008; for a review, see Decety and Grezes, 2006). Here we define simulation (as in Jeannerod, 2007) as the offline recruitment (for instance, during language processing) of the same neural networks involved in perception and action. In addition, we qualify it (as in Gallese, 2009) as an embodied and automatic mechanism which can also allow one to understand others' behaviours.

Although very simple, this way of representing the effects of language is in line with other more articulated theoretical proposals such as the Indexical Theory (Glenberg & Robertson, 2000) and, more recently, the Language And Situated Simulation (LASS) theory (Barsalou, Santos, Simmons, & Wilson, 2008; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008). The indexical hypothesis by Glenberg and Robertson (2000) explains how words are linked to perceptual and motor information related to their referents. The idea is that a word like "dog" partially re-activates the multi-modal representations that are acquired during experience with dogs. Thus, according to this theory, given the tight relationship between perception and action, words evoke object affordances (Gibson, 1979) and affordances do not only influence the understanding of single words but also more complex linguistic structures such as sentences. Similarly, according to the LASS theory (Barsalou, Santos, Simmons, & Wilson, 2008; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008), when a word is perceived the brain activates the internal representations corresponding to its auditory or visual form and this causes the activation of other similar representations associated to other words (e.g., because such words have been often heard or seen in temporal conjunction). Such direct associations allow the system to tackle some simple tasks such as some word/nonword lexical decision problems which require a "superficial processing" (e.g., to solve word-nonword tasks where nonwords are clearly distinguishable from words). At a later stage (which can nevertheless take place very fast, i.e. within 200 ms as shown in Pulvermüller, 2005, or even faster, as shown by Boulenger et al., 2008), such neural representations of the sound or visual form of words act as pointers to "semantic" knowledge stored as "situated simulations". This means that they activate perceptual, motor, emotional and other brain-state patterns which are also activated by the actual interactions with the referents of the words in the world. Note that this theory is closely related to the Dual Code Theory (Paivio, 1971; Paivio, 1986) and is very similar to the revised version of such theory presented by Glaser (1992), for example in terms of associations between early representations of words, timing of the activation of such representations, and brain patterns underlying simulations. Of course language is a very complex human ability and these models only capture some aspects of this ability. For example, words do not simply evoke the neural representation of objects but they can modify its nature (see Mirolli & Parisi, 2005 for a model on this). Moreover, models based on internal simulations might have problems in explaining how abstract words, which do not refer to perceivable objects, can be represented (Barsalou, 2003; Borghi & Cimatti, 2009; Boroditsky & Ramscar, 2002; Glenberg et al., 2008).

Title level 2: The three target psychological experiments and their simulation

Emphasised heading of paragraph: Experiment using object images

In one of the experiments reported by Tucker and Ellis (2004) where the stimuli were object images (for ease of reference this experiment will be henceforth called "experiment with images"), participants were requested to perform a categorization task, namely to distinguish between pictures of natural and artefact objects, by using different grips, namely precision vs. power grip. To indicate whether the seen object was an artefact or a natural object, participants mimicked either a precision or a power grip with one hand by acting upon a device similar to a joystick. In particular, one group of participants was instructed to use a power grip to classify objects as natural and a precision grip to classify objects as artificial; a second group of participants was instructed to use the opposite associations ("power-artificial" and "precision-natural"). During the experiment, the RTs of the participants were recorded.

To clarify the goal of the experiment, suppose that seeing an object automatically evokes the actions to interact with it, such as a grasp action suitable for the object size and shape. For example, seeing an apple likely activates the representation of its size and shape and prepares to the production of a power grasp. This preparation should facilitate a categorization response (e.g., deciding that the object is a natural one) performed with a power grip and interfere with a categorization response performed with a precision grip.

The simulations aimed at reproducing this procedure but simplified secondary aspects of it. The "simulated participant" could see eight different objects drawn from the original experimental set: four natural objects (apple, potato, grape, raisin), and four artificial objects (cup, mallet, eraser, needle). These eight objects differed in their colour, shape, and size (four were relatively large, i.e. graspable with a power grip, and four were relatively small, i.e. graspable with a precision grip). In the simulated experiments, the nervous system of 20 participants was simulated by using 20 different neural networks having different randomly-drawn initial connection weights. Half of the participants (first group) were trained to respond with a power grip when they saw a natural object and with a precision grip when they saw an artefact, whereas the other half of participants (second group) were trained for the opposite associations. After this training, the RTs of the participants were recorded after blocking all learning processes.

In contrast with the laboratory conditions, before undergoing the experiment the simulated participant first learned to associate a suitable kind of grip (e.g., a precision one) to each object (e.g., an eraser). This learning procedure was used to mimic what happens in everyday life, in particular in early years, when individuals learn to suitably respond to affordances of objects. Note that this is an essential element of the explanation of the compatibility effects presented here. Such an explanation relies on the hypothesis that the automatic reactivation of internal representations of affordances is acquired *before* the psychological experiment is carried out.

Emphasised heading of paragraph: Experiment using words

In a further experiment Tucker and Ellis (2004) asked participants to undertake the object categorization task mentioned above by reading names of objects on a computer monitor instead of seeing their images (for ease of reference, this will be called "experiment with words"). The goal of the experiment was to evaluate if compatibility effects can be activated not only by the sight of objects but also by verbal stimuli.

This experiment was replicated with 20 simulated participants by following a procedure that was similar to the one used to replicate the experiment with images but with the following differences. During life each participant learned to suitably grasp the objects and also to associate the names of objects with their shape and identity. During the experiment, the participant learned to trigger a power or precision grip on the basis of the names of objects (without seeing their images). After this training the RTs of the participants were recorded after blocking all the learning processes. Note that the simulated participants "heard" the object names instead of reading them: this was based on the assumption that read words activate the internal auditory representation of words. This simplification was introduced to avoid building a sophisticated visual system for word reading which would have been overly complex but not relevant for the goals of this research.

Emphasised heading of paragraph: Experiment using object parts

In one of the experiments reported in Borghi et al (2004) (for ease of reference this will be called "experiment with object parts"), participants sat in front of a computer screen and responded using a response-box where the buttons were arranged from top to bottom. First they read a sentence related to an object being in a particular orientation, for example "There is a car in front of you", and then they pressed the central button to see the name of another object on the screen (e.g., "roof", "wheels", or "parking place"). The task was to verify if the object in the sentence included the second object as one of its parts: in the "yes-is-up" condition, the participants had to press the upper button to indicate "yes" and to press the lower button to indicate "no". In the "yes-is-down" condition the response mapping was reversed. During the experiment, RTs were recorded.

The results of the experiment revealed that during language comprehension words work as cues that assist in accessing spatial information (such as the location of the object parts) and related action information (such as how to interact with the object and its parts). For example, the word "roof" related to "car" activates a spatial representation of the roof with respect to the car and so prepares one to interact with it by moving the hand up. This preparation facilitates a "yes response" in the yes-is-up condition and interferes with it in the yes-is-down condition.

To replicate this experiment, 20 simulated participants were presented with the names of two large objects ("car", "doll"; for ease of reference these will be called "whole objects"), four parts of them ("roof", "wheels", "head", "feet"; for ease of reference these will be called "object parts"), and two unrelated objects ("plum", "mandarin"; these were not used in the original test but were used here to check that the system could also learn to reach objects located in a central position within the working space). During the experiment, half of the participants (first group) were trained to respond with a yes-is-up reaching movement whereas the other half of the participants (second group) were trained to respond with a yes-is-down reaching movement. In this way, half of the times participants from both groups had to perform a response towards a button located in a position congruent with the position of the object part normally experienced during life, and the other half of the times they had to perform a response towards a button located in a position incongruent with it. After this training, the participants' RTs were recorded after blocking all learning processes.

Similarly to what happened in the experiments with images and words, and before undergoing the experimental training and test, the simulated participants learned to associate a suitable kind of reaching (e.g., moving up) to the images of the parts of the objects shown in a commonly-experienced position (e.g., moving to a central position to reach the "car" and the "plum", moving up to reach the "roof", and moving down to reach the "wheels"). This learning phase also allowed the participants to form associations between the names of objects and the neural internal representations of both their commonly-experienced spatial location and category.

Title level 2: The body of the simulated participants

The neural network model controlled the body of a 3D simulated participant endowed with a visual system, a human-like 3-segments/4-DOFs arm, and a 21-segments/19-DOFs hand (Figure 2). The simulated participant's arm and hand had the same parameters of the humanoid robot iCub (http://www.icub.org). The simulator was built on the basis of the 3D physics simulation library *Newton*^{TM}. The visual system of the participant was formed by a simulated "eye" (a 630×630 pixel RGB camera with a 120° pan and a 120° tilt angle, see Figure 2). This was mounted 25 cm above the shoulder, leaned forward 10 cm, and looked down at the working plane where the arm performed actions.

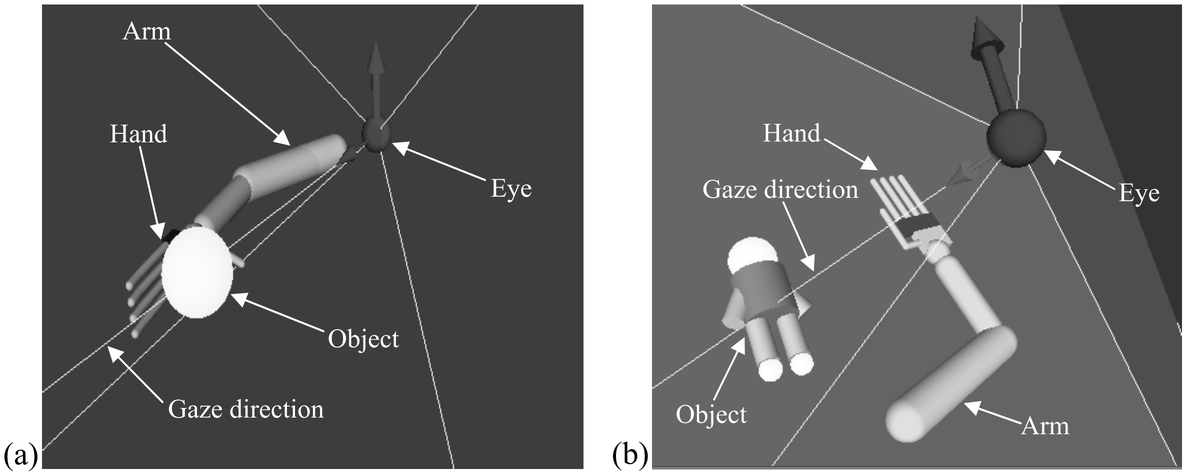

Figure 2: (a) The simulated arm, hand, and eye interacting with a simulated apple. (b) The simulated arm, hand, and eye interacting with a simulated doll. In both graphs, the line passing through the object indicates the gaze direction whereas the other four lines indicate the scope of the eye visual field. In both cases the arm and hand work on the plane.

The eye was controlled by a hardwired "focussing reflex" that led it to foveate the barycentre of objects with a specified red or green or blue colour. In particular, each of the desired pan and tilt angles of the eye was computed as follows:

$$a[t + \Delta t] = a[t] + 120\frac{(315 + c)}{630} \tag{1}$$

where a is the pan or tilt angle to be computed at time t, 315 is the centre of one dimension of the image (in pixels), 630 is the image size (in pixels), c is the weighted average of the (x, y) position within the retina (ranging in [-315, 315]) of the pixels having the desired colour (the average has weights equal to the activation of the pixels), and 120 is the movement range of the camera in each dimension (in degrees). The reflex is in line with the current neuroscientific literature suggesting that primates tend to foveate the target objects with which they interact and that their brain tends to exploit gaze centred reference frames for sensorimotor coordination (see Shadmehr & Wise, 2005, for a review).

During the experiments with images and with words, the system was exposed to eight different objects (their images, or their images and names, respectively in the two experiments): two large natural objects (apple: diameter 36 mm; potato: 32 mm), two small natural objects (grape: 11 mm; raisin: 10 mm), two large artefacts (cup 34 mm, mallet: 30 mm), and two small artefacts (needle: 4 mm; eraser: 10 mm). These objects are shown in Figure 4.

During the experiment with object parts, the system was exposed to eight different objects (their images and their names): two whole objects (car: positioned in the working plane at -10 cm to the left of the shoulder and 20 cm in front of it; doll: -10 cm, 25 cm), four object parts (roof: -10 cm, 30 cm; wheels: -10 cm, 15 cm; head: -10 cm, 30 cm; feet: -10 cm, 15 cm), and two "unrelated" objects (plum: -10 cm, 20 cm; mandarin: -10 cm, 20 cm). These objects are shown in Figure 13.

Note that the fact that the object parts were shown in isolation amounts to assuming that the system could filter out the other parts of the "whole object" on the basis of a suitable object-based attention mechanism not explicitly simulated here (Logan, 1996; Roelfsema, Lamme, & Spekreijse, 1998).

During the simulated experiments, the system was asked to reach three "buttons", in particular three positions in space located at -10 cm to the left of the shoulder and respectively 30, 20, and 10 cm in front of it.

To simplify the analysis of results, the images encoded in the visual units of the system were caused only by the objects and not by the hand. This amounts to assuming that the system could also filter out the hand on the basis of the object-based attention mechanism mentioned above. In this respect, it is relevant that seeing an object is sufficient to activate the representations of its affordances in the PC and pre-activate motor neurons in PMC (in particular, "canonical neurons"; Rizzolatti & Craighero, 2004; Creem & Proffitt, 2001; Martin, 2007; Bub, Masson, & Cree, 2008).

In the experiments with images and with words, the model controlled only 2-DOFs of the hand: one for the thumb, whose DOFs were all controlled proportionally to a first command issued by the model, and one for the four same-sized fingers, controlled as a whole "virtual finger" proportionally to a second command issued by the model (cf. Arbib, 2002). In these experiments, the DOFs of the wrist and arm were kept still as requested by the tasks of the target experiments. During the simulation of the participants' interactions with objects during life, objects were kept fixed in space to avoid that they slipped away from fingers during closure, similar to what happens in grasping objects placed on a horizontal plane.

In the experiment with object parts, which required reaching for different buttons located at different positions organised from top to bottom on a table in front of the participants, the model controlled only 2-DOFs of the arm and so worked on the plane. In this experiment the DOFs of the wrist and hand were kept to fixed angles so to have the hand aligned with the forearm in a straight position suitable for reaching the buttons, as requested by the task.

The activation of the grasping output map of the model (lateral premotor cortex - PMCl) encoded the desired hand posture, whereas the reaching output map of the model (dorsal premotor cortex - PMCd) encoded the desired arm posture. These postures were used to set the hand and arm "muscle models" in terms of "equilibrium points" (Feldman, 1986; Flanagan, Ostry, & Feldman, 1993). In particular, similarly to what done by Berthier, Rosenstein, and Barto (2005), single muscle models were simulated as simple Proportional Derivative controllers (PDs; these models capture the main elastic and damping properties of muscles, cf. also Bullock & Grossberg, 1989). The equation of a PD controller is as follows (Sciavicco & Siciliano, 1996):

$$\boldsymbol{T} = \boldsymbol{K_P}\tilde{\boldsymbol{q}} - \boldsymbol{K_D}\dot{\boldsymbol{q}} \tag{2}$$

where $\boldsymbol{T}$ is the vector of muscle torques applied to the joints, $\boldsymbol{K_p}$ is a diagonal matrix with elements equal to 300, $\tilde{\boldsymbol{q}}$ is the difference vector between the desired angular positions and the current angular positions of joints, $\boldsymbol{K_D}$ is a diagonal matrix with elements equal to 10, and $\dot{\boldsymbol{q}}$ is the vector of current angular speed joints. The use of this abstract muscle model implies that the difference between the desired joint angles set by the system and the actual joint angles is used to set a proportional torque which diminishes such a difference. In addition to this, the model also generates a torque opposite to the movement and proportional to the joint angular speed which results in a dumping effect which increases arm stability. The action of the PDs was also assumed to be integrated by a gravity compensation mechanism here implemented by simply ignoring the effects of gravity on the arm and hand.

Title level 2: TRoPICALS: Neuroscientific constraints, architecture, and functioning

This section has two goals. First, it reviews the neuroscientific literature which was used to constrain the overall architecture of TRoPICALS, in particular to decide which components to include in the model and which particular function to assign to them. This review is introduced now and then developed in the following sub-sections in relation to the specific components of the model. Secondly, it illustrates (in the sub-sections) how the functions of the brain areas included in the model were "translated" into specific computational mechanisms. Figure 3 shows the architecture of the model. The model is formed by ten components each corresponding to a brain area and represented by a neural map of 21×21 neurons.

The choice of the components of the architecture of the model used in the first and second experiment is broadly constrained on the basis of brain imaging evidence provided by Grèzes, Tucker, Armony, Ellis, and Passingham (2003). Their work shows that portions of parietal, premotor, temporal, and frontal cortex of humans performing the experiment of Tucker and Ellis (2004) are active during its execution. The architecture of the model is also constrained at a finer level on the basis of other sources of information on the various brain areas corresponding to the model components. The main sources of information used to draw these constraints are reported in Table 1. This table also shows the additional constraints used to formulate the extended version of the model presented in Figure 17.

The architecture of the model is formed by a dorsal neural pathway and a ventral neural pathway. The dorsal pathway is in turn formed by two pathways: (a) a visual-parietal-premotor pathway underlying grasping behaviour; (b) a somatosensory-parietal-premotor pathway underlying reaching behaviours (Matelli, Luppino, Murata, & Sakata, 1994). In this respect, several physiological experiments with monkeys have shown that there are parietal-premotor neural pathways that receive their input from visual and somatosensory areas and control "grasping" (i.e., the movements of the hand; Rizzolatti, Luppino, & Matelli, 1998) "reaching" (i.e., the movements of the proximal arm; Wise, Boussaoud, Johnson, & Caminiti, 1997), and "overt attention" (i.e., the movements of the eye; Colby & Goldberg, 1999). These pathways usually involve partially segregated sub-areas of the parietal and the premotor cortex and form whole functional systems dedicated to the control of different "actuators" or portions of them (Rizzolatti, & Luppino, 2001). These pathways have homologues in human cortical areas which subserve similar functions, as shown by various experiments based on brain imaging techniques (Culham & Kanwisher, 2001, Simon, Mangin, Cohen, Bihan, & Dehaene, 2002).

In the model, the ventral visual-prefrontal-premotor pathway integrates information about object category and context (in particular, the "experimental context" vs. the "ecological context", see below) to exert a top-down control on action. In the brain, the ventral pathway involves cortical areas performing high-level visual processing (Felleman & Van Essen, 1991; Van Essen et al., 2001; Grill-Spector & Malach, 2004). This information reaches various areas of PFC and provides them with information for setting high-level goals and sub-goals of action (Fuster, 2001). Interestingly, various frontal areas (e.g., both PFC and PMC areas) are also the target of phonetic information originating from the auditory compartments of the superior temporal cortex (Romanski et al., 1999). In humans, this connectivity might allow frontal areas (and especially PFC ones) to integrate visual and linguistic information so as to exert a suitable top-down control on action (Pulvermüller, 2005), as assumed by the model presented here.

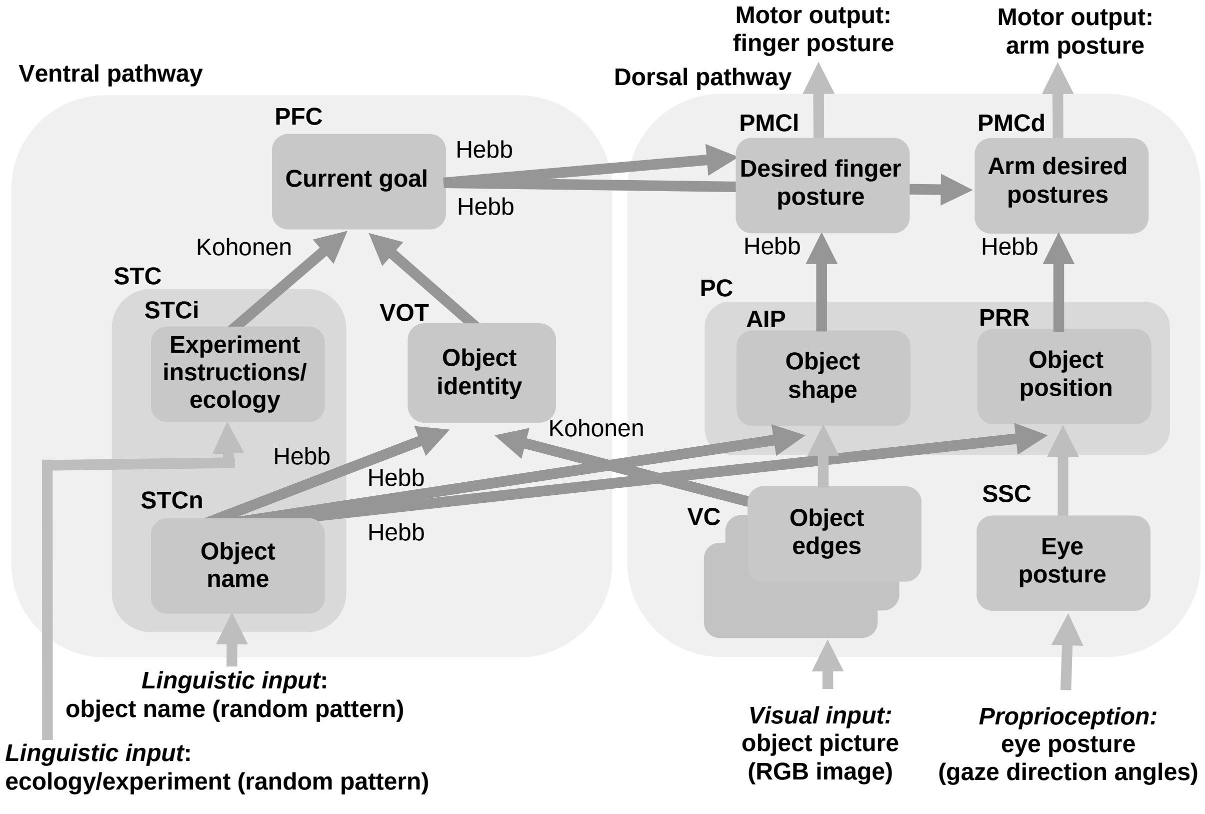

Figure 3: Architecture of TRoPICALS. The boxes indicate the components of the model. The label inside each box indicates the type of information encoded by the component, whereas the acronym at its top-left corner indicates the brain anatomical area putatively corresponding to it (the acronyms are explained in the label of Table 1). Light and dark grey arrows respectively indicate connections which were hardwired and connections which were updated by learning processes based on a Hebb covariance learning rule or a Kohonen learning rule. The input of the model is formed by three RGB visual neural maps (VC) and a somatosensory map (SSC). Downstream the VC and the SSC, the model divides into two main neural pathways: the dorsal pathway, which implements suitable sensorimotor transformations needed to perform action on the basis of perception, and the ventral pathway, which allows flexible control of behaviour thanks to the biasing effects exerted by the PFC on action selection. In turn, the dorsal pathway is formed by a pathway controlling grasping and a pathway controlling reaching.

Table 1

Review papers (see within them for detailed references) drawn from the neuroscientific literature on humans and monkeys in support of the brain neural pathways linking the various components of the model. Cortical areas of monkeys that are homolog to human cortical areas are indicated in brackets. Symbols: AIP: anterior intraparietal sulcus; FEF: frontal eye field; PEF: posterior eye field (LIP: lateral intraparietal area); PFC: prefrontal cortex; PMCd: premotor cortex controlling the proximal arm (PMCd: premotor cortex, dorsal division; or F2: portion F2 of dorsal premotor cortex); PMCl: premotor cortex controlling the hand digits (F5: portion F5 of ventral premotor cortex); PRR: parietal reach region (MIP: medial intraparietal area); SSC somatosensory cortex; STC: superior temporal cortex; VC: visual cotex; VOT: ventral occipito-temporal region (ITC: inferotemporal cortex).

From	**To**	**References on humans**	**References on monkeys**
VC	AIP	Simon, Mangin, Cohen, Le Bihan, & Dehaene (2002) Culham & Kanwisher (2001)	Rizzolatti, Luppino, & Matelli (1998)
VC	PEF (LIP)	Simon, Mangin, Cohen, Le Bihan, Dehaene (2002) Culham & Kanwisher (2001)	Colby & Goldberg (1999)
VC	VOT (ITC)	Grill-Spector & Malach (2004)	Felleman & Van Essen (1991)
SSC	PRR (MIP)	Simon, Mangin, Cohen, Le Bihan, & Dehaene (2002) Culham & Kanwisher (2001)	Wise, Boussaoud, Johnson, & Caminiti (1997)
SSC	PEF (LIP)	Simon, Mangin, Cohen, Le Bihan, & Dehaene (2002) Culham & Kanwisher (2001)	Colby & Goldberg (1999)
VOT (ITC)	PFC	Fuster (2001)	Webster, Bachevalier, & Ungerleider (1994)
VOT (ITC)	PC	Essen, Lewis, Drury, Hadjikhani, Tootell, Bakircioglu, & Miller (2001)	Webster, Bachevalier, & Ungerleider (1994)
AIP	PMCl (F5)	Simon, Mangin, Cohen, Le Bihan, & Dehaene (2002) Culham & Kanwisher (2001)	Rizzolatti, Luppino, & Matelli (1998)
PRR (MIP)	PMCd (PMCd or F2)	Simon, Mangin, Cohen, Le Bihan, & Dehaene (2002) Culham & Kanwisher (2001)	Wise, Boussaoud, Johnson, & Caminiti (1997)
PEF (LIP)	FEF	Simon, Mangin, Cohen, Le Bihan, & Dehaene (2002) Culham & Kanwisher (2001)	Colby & Goldberg (1999)
PFC	PMC	Fuster (2001)	Rizzolatti, Luppino, & Matelli (1998) Wise, Boussaoud, Johnson, & Caminiti (1997) Colby & Goldberg (1999)
STC	PFC	Pulvermüller (2005)	
STC	VOT	*Indirect pathways*	
STC	AIP	*Indirect pathways*	
STC	PRR	*Indirect pathways*	
STC	PEF	*Indirect pathways*	

Emphasised heading of paragraph: Visual cortex (VC): encoding of object edges

In the monkey (Felleman & Van Essen, 1991, Van Essen et al., 2001) and the human brain (Van Essen et al., 2001, Grill-Spector & Malach, 2004; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001), early stages of the visual cortex (VC) are organised hierarchically. This hierarchy underlies a processing of visual information which leads to extracting increasingly abstract information from visual images, from simple edge detection (Hubel, 1988) to complex feature recognition (Vinberg & Grill-Spector, 2008). Information elaborated in VC is important for both object recognition processes taking place in the ventral pathway and for the sensorimotor transformations guiding action and taking place in the dorsal pathway (see below).

In the model, VC is formed by three maps and receives the visual signal supplied by a simulated camera. In particular, the neurons of the three maps have an activation which ranges in [0, 1] and encode information about shape and colour of the foveated object obtained through three distinct Sobel filters (Sobel & Feldman, 1968), one for each map. Each filter performs edge detection on the basis of the image of the object formed by only one colour (either red, green, or blue). These processes represent edge detection performed by the retina and by the succeeding early stages of the visual cortex in primates. The model assumes that the eye always foveates the target on the basis of the focussing reflex.

Emphasised heading of paragraph: Somatosensory cortex (SSC): encoding of eye posture

In monkey (Colby & Goldberg, 1999; Wise, Boussaoud, Johnson, & Caminiti, 1997) and human brain (Simon, Mangin, Cohen, Bihan, & Dehaene, 2002; Culham & Kanwisher, 2001), the associative parietal cortex areas that form the parietal-premotor pathways which control movements of the hand, the arm, and the eye get important information from the somatosensory cortex (SSC). Such parietal areas implement important visuomotor transformations suitable for the control of various actuators via their connections to PMC. In this respect, information from SSC allows the PC to suitably adjust visual information with respect to different reference frames on the basis of the state of relevant parts of the body (e.g., the position of the head with respect to the arm while this is controlled; Andersen, Snyder, Bradley, & Xing, 1997; Colby, 1998; Platt & Glimcher, 1999; Shadmehr & Wise, 2005). Proprioception might also furnish the parietal cortex information readily usable for action. For example, in normal conditions participants performing a reaching action usually fixate the target object with the eyes. The gaze direction thus furnishes a “clean” information on the object position which is readily usable by the arm (Allport, 1987; Balkenius, 2000; DeSouza et al., 2000).

In line with this, the SSC component of the model encodes the proprioception of the eye in terms of its gaze direction angles. These allow the PC-PMC pathway controlling the arm to know the position in space of the target of the reaching action. In particular, the neurons of SSC have different “preferred” gaze angles assigned to them on the basis of their positions in the neural map (these positions lay on the vertexes of a regular grid) and the assumption that the two dimensions of the map encode the two gaze angles. Depending on the eye position, the neurons are activated on this bases through a Gaussian function on the basis of the distance between their preferred angles and the actual gaze angles:

$$a_i = \exp\left[-\frac{\left\| v_g - v_i \right\|^2}{2\sigma^2} \right] \qquad (3)$$

where a_i is the activation of neuron i of the map, $||v_g - v_i||$ is the distance between the gaze angles and the neuron preferred angles seen as two vector points in the neural map 2D space (the measure unit of this space is the distance between two neighbouring neurons), σ is the width of the Gaussian function (σ was set to 0.6). This way of encoding information in the map is in line with the “gain fields approach” which assumes that brain uses neural maps, and combines their activation in a multiplicative fashion, to encode information such as eye or limb postures (Pouget, Dayan & Zemel, 2000; Pouget & Latham, 2002; Shadmehr & Wise, 2005).

Emphasised heading of paragraph: Anterior intraparietal sulcus (AIP): detection of object shape

In the monkey PC, areas such as the anterior intraparietal area in the intraparietal sulcus (AIP; Rizzolatti, Luppino, & Matelli, 1998; Murata, Gallese, Luppino, Kaseda, & Sakata, 2000), and the caudal portion of the intra parietal cortex (cIPS; Sakata, Taira, Kusunoki, Murata, & Tanaka, 1997), encode information about features of objects which are important to guide manipulation, for example shape, orientation, 3D aspects, and tactile aspects of objects. AIP neurons also discharge during goal directed actions performed with the hand and mouth (e.g., grasping, holding, and manipulation). AIP receives important connections from PFC (Borra et al., 2008; Baumann, Fluet, & Scherberger, 2009) and so it possibly plays a high-level role in the hierarchy of action selection and performance. On the basis of this evidence, many authors claim that AIP plays a central role in encoding object affordances (Fagg & Arbib, 1998; Oztop, Bradley, & Arbib, 2004). The human homologue of monkey AIP might equally be the anterior intraparietal area which is activated by the manipulation of objects and the view of graspable objects (Culham & Kanwisher, 2005; Simon, Mangin, Cohen, Hihan, Dehaene, 2002).

In the model, AIP encodes information about object shape (cf. Oliver & Thompson-Schill, 2003; Oliver, Geiger, Lewandowski, & Thompson-Schill, 2005;). In particular, the neurons forming this map are activated with the average activation of the topologically correspondent RGB edge-encoding neurons of VC. Note that the representation of the “shape” of objects used here should be considered as a “proxy” of the various object properties encoded in the real AIP. Moreover, the PFC-PMC influence on action selection used in the model should be considered a proxy of the whole influence that PFC exerts on PMC and AIP. These abstractions were imposed by the need to focus on the few properties of objects necessary to investigate compatibility effects while avoiding a detailed representation of other object properties underlying the actual performance of grasping (on this, cf. Fagg & Arbib, 1998; Oztop & Arbib, 2002; Oztop, Bradley, & Arbib, 2004).

A further assumption model about AIP relates to the fact that, due to the activation process described above, large objects would tend to activate a much larger number of neurons with respect to small objects. This goes against the empirical evidence showing that the opposite happens in real brains, namely that the prefrontal cortex, rostral cingulate cortex, parietal cortex, and premotor cortex (but not primary motor cortex) tend to have a larger activation with small objects than with large ones (Ehrsson et al., 2000). This is likely due to the fact that the interaction with small objects requires a more sophisticated and detailed control of fingers in order to perform accurate precision grips, and so requires the activation of neurons not required to perform gross power grips. Although the level of abstraction of the model did not allow representing these differences, we used a coefficient s to scale the activation of AIP neurons based on the size of objects (s was set to 1 for small objects and to 0.3 for large objects and was used to multiply the activity of all neurons of AIP). Although this does not reproduce the *causes* of the different activation of PC by large and small objects, it allows reproducing the resulting differential activation at a *phenomenon* level. This allows regulating the activation of AIP so as to avoid the undesired effects on RTs which would be caused by the unrealistic large activation of AIP by large objects (see the Appendix for an indication of the effects of the manipulation of s on RTs).

Emphasised heading of paragraph: Parietal reach region (PRR): detection of object position

In monkeys, the medial intraparietal area (MIP) plays a key role in reaching (Wise, Boussaoud, Johnson, Caminiti, 1997). Neurons from this area encode visual stimuli, somatosensory information on arm movement and positioning, and gaze direction. It has been proposed that the human homologue of the monkey MIP is the “parietal reach region” (PRR; Andersen, Essick, & Siegel, 1985; Simon, Mangin, Cohen, Bihan, & Dehaene, 2002; Culham & Kanwisher, 2001) which is activated during reaching movements and is modulated by eye position (DeSouza et al., 2000).

In the model, the neurons of the PRR are a simple copy of SSC. In this respect, the PRR map was added for future extensions of the model to represent the processes performed by this area to support reaching movements. Indeed, in real brains this region implements fundamental sensorimotor transformations underlying reaching, for example the remapping of the object position in space from the eye reference frame to the limb reference frame on the basis of eye position in the orbit and the head orientation (Colby, 1998; Platt & Glimcher, 1999; Shadmehr & Wise, 2005; Jeannerod, Arbib, Rizzolatti, & Sakata, 1995). At the stage of development of the model in which the eye and head of the system do not move these complex remappings have not been simulated. However, they might be necessary in the future development of the model.

Emphasised heading of paragraph: Premotor cortex, lateral division (PMCl): encoding of desired finger postures

In monkeys, the PMC area F5 has been extensively studied as it seems to encode a rich repertoire of actions such as precision, power, and side grasps not only when these are performed but also when they are observed (Rizzolatti, Luppino, & Matelli, 1998; Luppino & Rizzolatti, 2000; Rizzolatti & Craighero, 2004). The human homologue of the monkey F5 has been linked to area 44 in the lateral premotor cortex, here referred to as PMCl for simplicity (Rizzolatti, Luppino, & Matelli, 1998).

In the model, the neurons of PMCl control hand movements in terms of the two desired angles issued to the finger joints. These angles, encoded by the two dimensions of the map, are “read out” from the map as a weighted average of the position of neurons within it, suitably remapped onto the angles space, with the weights of the average corresponding to the activation of the neurons. Note that the operation of “information reading out” from a neural map is the opposite of the operation of “information encoding” onto a neural map seen in this section in relation to SSC. The two operations are used when information is encoded in neural maps with “population codes” (Pouget, Dayan, & Zemel, 2000).

PMCl supports the selection of postures on the basis of a dynamic competition involving the leaky neurons which form it. These neurons have short-range lateral excitatory connections and long-range lateral inhibitory connections (Erlhagen & Schöner, 2002). When input signals from AIP and PFC activate neurons of PMCl, these tend to accumulate activation, form clusters (due to the lateral excitatory connections) and, at the same time, suppress the formation of other clusters (via lateral inhibitory connections). This dynamic process continues until a cluster succeeds in suppressing all other clusters, overcomes a threshold and triggers the hand movement based on the reading out of the map. Mathematically, each PMCl leaky neuron is activated as follows (the formula is presented for two generic PMC and PC areas as it is used for the interactions between all PC and PMC specific areas):

$$\begin{gathered} s_j[t] = \overbrace{\sum_{i=1}^{N} w_{(PMC\rightarrow PMC)\,ji}\, a_{PMC\,i}[t]}^{Inner} + \overbrace{\sum_{i=1}^{N} w_{(PC\rightarrow PMC)\,ji}\, a_{PC\,i}[t]}^{Dorsal\ stream} + \overbrace{\sum_{i=1}^{N} w_{(PFC\rightarrow PMC)\,ji}\, a_{PFC\,i}[t]}^{Ventral\ stream} \\ u_j[t+\Delta t] = \left(1-\frac{\Delta t}{\tau}\right) u_j[t] + \frac{\Delta t}{\tau} s_j[t] \qquad a_j[t+\Delta t] = \left[\tanh\left[u_j[t+\Delta t]\right]\right]^{+} \end{gathered} \tag{4}$$

where $s_j[t]$, $u_j[t]$ and $a_j[t]$ are respectively the input signal, the activation potential, and the activation of neuron j at time t, Δt is the integration time step (Δt was set to 0.01 sec, which implies that 100 steps of simulation correspond to 1 sec of real time), τ is a time constant (τ was set to 0.5 sec), tanh[.] is the hyperbolic tangent function, and $[.]^{+}$ is the identity function returning 0 for negative values. The *Inner* component of the input signal accounts for signals received from PMC lateral connections having hardwired connection weights $w_{(PMC\rightarrow PMC)}$. These weights, excitatory for connections between neighbouring neurons and inhibitory for connections between distant neurons, are set to fixed values on the basis of a Gaussian function and an inhibition term as follows:

$$w_{(PMC \to PMC)ji} = \exp\left[-\frac{\left\| v_j - v_i \right\|^2}{2\sigma^2} \right] - I \qquad (5)$$

where $w_{(PMC \to PMC)ji}$ is the weight between two neurons i and j of the map, $||v_j - v_i||$ is the Euclidean distance between the two neurons in the map "neural space" (the measure unit being the distance between two neighbouring neurons), σ is the width of the Gaussian function (σ was set to 0.6), and I is the inhibition term (I was set to 0.9). The *Dorsal stream* component of the input signal accounts for the signals received from PC neurons through connections having weights $w_{(PC \to PMC)}$. Finally, the component *Ventral stream* accounts for the signals received from PFC neurons through connections having weights $w_{(PFC \to PMC)}$. Note that the weights $w_{(PFC \to PMC)}$ were kept within the interval [0, 0.3] and the weights $w_{(PC \to PMC)}$ within the interval [0, 0.1] as PFC signals need to overwhelm affordance-related signals when necessary (Miller & Cohen, 2001). Within the model, reaction times correspond to the time requested by at least one neuron from the winner cluster of PMC to reach an action-triggering threshold (Erlhagen & Schöner, 2002; the threshold value was set to 0.7).

An important caveat on the RTs produced by the model is that they rely upon various coefficients which could not be set to realistic values drawn from the properties of real neurons. This was due to the relatively high abstraction level of the neurons of the model (firing rate neurons; cf. Dayan & Abbott, 2001). For this reason, the model was intended to reproduce only the pattern of qualitative results of the target experiments, and not their quantitative levels (e.g., in relation to the specific RTs exhibited by the real participants; cf. Erlhagen & Schöner, 2002). The Appendix presents a qualitative analysis on the effects that some key parameters of the model produce on its RTs.

Emphasised heading of paragraph: Premotor cortex, dorsal division (PMCd): encoding of desired arm postures

In monkeys, the dorsal premotor cortex (PMCd) plays an important role in the performance of reaching and pointing movements (Wise, Boussaoud, Johnson, & Caminiti, 1997; Cisek & Kalaska, 2005; Shadmehr & Wise, 2005). The human homologue of the monkey PMCd might be the dorsal 6aα area in the dorsal premotor cortex, here referred to as PMCd for simplicity (Rizzolatti, Luppino, & Matelli, 1998; Wise, Boussaoud, Johnson, & Caminiti 1997).

The neurons of PMCd encode the output of the system related to the control of the arm in terms of desired angles of the arm joints. Similarly to what is done for PMCl, these angles are mapped onto the two dimensions of the map and are "read out" from the map as a weighted average of the position of neurons in it (suitably remapped onto the joint space) with the weights of the average corresponding to the activation of the neurons. PMCd supports the selection of arm postures on the basis of a dynamic competition similar to the one implemented by PMCl (described above) and based on short-range lateral excitatory connections and long-range lateral inhibitory connections (see Equation 5; here Δt was set to 0.01 sec, τ was set to 0.3 sec, σ was set to 0.6, and I was set to 0.9). The reading out from the map, and the consequent action execution, are performed when a cluster of the map succeeds to suppress all other clusters and to overcome a certain threshold (set to 0.7).

Emphasised heading of paragraph: Ventral occipito-temporal cortex (VOT): encoding of object identity

In the monkey brain, the inferior temporal cortex (ITC; Logothetis, Pauls, & Poggio, 1995) is located at the highest level of the ventral visual pathway and plays an essential role in visual object recognition (Felleman & Van Essen, 1991; Van Essen et al., 2001). In humans, various areas within the ventral visual pathway have an activation which responds to highly abstract patterns (e.g., faces, places, houses) and is rather invariant with respect to various aspects of images such as location, orientation, and luminance (Grill-Spector & Malach, 2004; Vinberg & Grill-Spector, 2008). Among these areas, the ventral occipito-temporal cortex (VOT) seems to play a key role in object recognition.

In the model, the VOT encodes the "identity of objects" on the basis of their colour and shape. Note that here "identity of objects" should be intended as "category of objects": the former expression was used instead of the latter as the simulated setup involves only one non-noisy item for each object category. In accordance with visual physiology findings (Tanaka, 1996), from lower levels (VC) to higher levels of the visual hierarchy (e.g., VOT) the size of receptive fields and stimulus selectivity of neurons increase, whereas their visual topography is progressively lost. In the model, this was in part captured by recoding the information of VC into VOT on the basis of a self-organising map ("SOM"; Kohonen, 1997). Each unit of the map is activated as follows:

$$a_j = \exp\left[-\frac{\left\| a_{VC} - w_{(VC \rightarrow VOT)j} \right\|^2}{2\sigma^2} \right] \qquad (6)$$

where a_j is the activation of VOT neuron j, a_{VC} is the vector of activations of VC neurons, $w_{(VC\rightarrow VOT)j}$ is the vector of the connection weights linking VC neurons to VOT neuron j, σ is the size of the Gaussian function used to compute the activation (σ was set to 0.55). SOMs tend to form weights such that for any given input pattern the activation function causes a hill-shaped spatially-segregated cluster of active neurons within the map, whereas the other neurons remain silent. Moreover, similar input patterns tend to generate similar clusters of active neurons whereas substantially different input patterns tend to activate non-overlapping clusters.

This modality of functioning of SOMs allow using them as a proxy of the representation of objects in high-level visual cortex which has a level of abstraction suitable for the purposes of this work. Indeed, SOMs were chosen as: (a) they tend to form categories of stimuli based on population-codes, similarly to what might happen in the brain (Pouget, Dayan, & Zemel, 2000); (b) although they are sensitive to the spatial location of patterns within the image, they tend to respond to similar stimuli with the activation of similar clusters of units, while ignoring differences of few features (Kohonen, 1997): research on object categorisation carried out in the last ten years is showing that higher stages of visual processing tend to develop different areas dedicated to different broad classes of objects (e.g., faces, places, words, and houses; cf. Grill-Spector & Malach, 2004); (c) they are a computational device which is capable of developing categories of objects autonomously, so they are suitable to mimic processes of categorisation learning; (d) they perform a useful orthogonalisation of input patterns which aids downstream learning processes (e.g. based on Hebb learning rules); (e) they are computationally robust (i.e., tolerate noise) and at the same time parsimonious, and this eases modelling.

Emphasised heading of paragraph: Superior temporal cortex for nouns (STCn): encoding of object names

In humans, the superior temporal cortex (STC) is the locus of the primary auditory cortex (A1), of the auditory belt region of Wernicke's area (WP), and of the auditory parabelt region of Wernicke's area (WPB) (Pulvermüller, 2005). These regions play a key role in language comprehension as they send various efferent connections to prefrontal and premotor regions where some aspects of the sensorimotor representations of word referents might be encoded (Romanski et al., 1999; Young, Scannel, Burns, Blakemore, 1994; Hilgetag, Burns, O'Neill, Scannell, & Young, 2000). In this respect, it is important to consider that in the model the STC components are also connected to regions of PC. It is reasonable to assume that words can also recall aspects of their referents which are encoded in such regions (e.g., spatial and haptic aspects of stimuli). The authors are not aware of direct anatomical connections which might support these processes but there might be indirect neural pathways to support them.

In the model, STC encodes both linguistic information about names of the seen objects and linguistic information about the instructions of the experiment. To differentiate these two kinds of information two neural maps were used which were assumed to represent two distinct sets of neuron clusters of STC: STCn, corresponding to neuron clusters encoding the (phonological aspects of the) names of objects, and STCi, corresponding to neuron clusters encoding the (phonological aspects of the) experiment instructions or, alternatively, the ecological conditions. Note that the distinction between STCn and STCi is a *functional* one, not an anatomical one, and is due to the lack of empirical data on where and how the nouns of objects and the experiment instructions are encoded within STC. STCn and STCi respectively project to the VOT and PFC to represent the fact that functionally the broad context experienced by the participants (experiment/ecological condition) likely exerts a relatively strong influence on the goals of actions encoded in PFC, and the fact that object nouns have relatively strong links with their sensorimotor counterparts.

The neurons of STCn are activated with a different pattern for each different object (each pattern is formed by 20 randomly chosen neurons set equal to one and the rest set equal to zero). During the simulated tests, STCn neurons are activated gradually. This gradual activation represents, at a *phenomenon* level, the slow processing of the complex linguistic processes which in the real experiments likely activate the phonetic representations of words on the basis of read words. This gradual activation was introduced to avoid having biased RTs due to the lack of a specific simulation of the aforementioned linguistic processes which would have required overly complex mechanisms non needed for the scope of this research.

Emphasised heading of paragraph: Superior temporal cortex for instructions (STCi): encoding of the experiment instructions or of the ecological conditions

In the model, the neurons of STCi were activated with either one of two random patterns (each pattern was formed by 20 randomly chosen neurons set equal to one and the rest set equal to zero). These two patterns were assumed to correspond to either the instructions of the three target experiments which asked participants to perform a “categorisation grasp”/“decision reach”, or to an ecological context requiring the participants to perform a grasp or a reach suitable for the object type.

Emphasised heading of paragraph: Prefrontal cortex (PFC): encoding of current goal

In the model the neurons of the prefrontal cortex (PFC) are assumed to encode information about the current goal of action depending on both the task (STCi) and the identity of the object (VOT), similarly to what is done in Deco and Rolls (2003). Computationally, PFC neurons are activated again according to a Kohonen activation function as the one used for the VOT (Equation 6). The use of Kohonen networks to also represent processes taking place in PFC is justified by the studies which show that this cortical area is involved in highest-levels of visual processing, categorization, and organisation of behaviour (Miller, Freedman, & Wallis, 2002; Freedman, Riesenhuber, Poggio, & Miller, 2003; Shima, Isoda, Mushiake, & Tanji, 2007).

These assumptions lead the system to form representations for all possible combinations of the eight objects used in each experiment and their two sub-tasks (ecological versus experimental conditions). A more parsimonious representation would have been possible, for example by allowing PFC to represent the two categories of objects required by the target experiments (e.g., natural vs. normal in the experiments with images and words) or the ecological experiment (e.g., large vs. small). However, this solution would have required introducing specific mechanisms, such as a supervised training mechanism, to allow the system to form the two categories before tackling the experimental task. The former solution was preferred as it was simpler (this usually eases the interpretation of results) and it was neutral with respect to the study of compatibility phenomena.

Title level 2: TRoPICALS: Learning mechanisms

The model was trained in three learning stages which roughly correspond to the main relevant learning phases experienced during life and during the experiment by real participants. These learning stages are now explained in detail and, when necessary, they are explained separately for the three experiments with images, words and object parts. Before training, the weights of all the plastic connections of the model were set to values randomly drawn within a uniform distribution within [0, 0.1].

Emphasised heading of paragraph: Phase 1. Learning to interact with objects during life: experiment with images and experiment with words

Before the experiments with images and words, the system underwent a first learning phase which simulated learning to grasp objects experienced by real participants during life. This learning phase led the model to acquire the affordance-based behaviour within the dorsal stream (in particular, this process updated AIP-PMCl connection weights; cf. Rizzolatti, Luppino, & Matelli, 1998) and to form the identity of objects within the ventral stream (in particular, this process updated the VC-VOT connections; cf. Rodman, 1994).

The training was performed by repeatedly presenting each of the eight objects to the model (the simulation steps related to the presentation of one object is here referred to as "trial"). On the perceptual side, at each presentation VC performed colour-based edge detection of the object image and PC performed colour-independent edge detection (i.e., extracted the shape of objects) by averaging the activation of RGB neurons having the same position in the three maps of VC.

AIP-PMCl connection weights were updated on the basis of a "motor babbling" process and a Hebb covariance learning rule. This allowed the dorsal pathway to acquire the capacity to perform the correct grasp depending on the shape of objects. Motor babbling processes (von Hosten, 1982; Caligiore et al., 2008) are general learning processes for which the production of rather unstructured behaviours allow the formation of basic associations between sensory representations and motor representations (Piaget, 1952). Here the motor babbling process involved these phases at each step of the trial: (a) one object was set close to the hand palm of the system; (b) the neurons of PMCl were activated on the basis of a Gaussian function (the shape and parameters of the function were set as in Equation 3) encoding specific desired angles of the hand aperture: such angles were progressively decreased of two degrees per DOF at each step of the trial; (c) the desired hand postures encoded by PMCl were issued to the PD models ("muscles") and these produced suitable torques which progressively closed the hand on the object during the trial; (d) in the step in which the thumb, index, and medium finger tips were in touch with the object (see Figure 4) the activation patterns of AIP and of PMCl were associated on the basis of the covariance Hebb rule (see Equation 7 below).

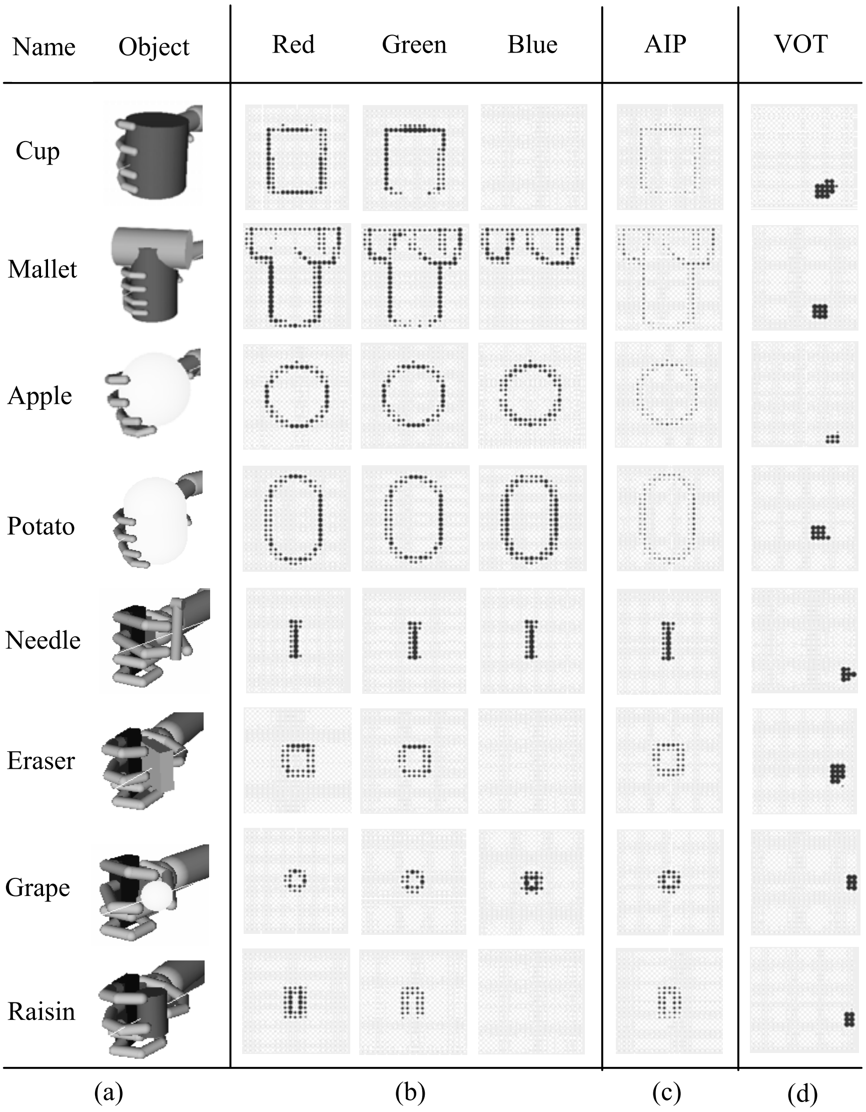

Figure 4: Experiment with images. The columns of this graph, from left to right, report: (a) The object name, appearance, and hand grip on it; (b) The activation of the VC (three RGB maps) caused by the eight objects; (c) The activation of the AIP encoding the shape of objects; (d) The activation of the VOT encoding the identity of objects. Notice the different grips for the different objects, the activations of the VC caused by the variety of shape and colours of the objects, the encoding of shape in the AIP, and the abstract representation of the identity of the objects in the VOT.

The activation of PMCl with progressively smaller apertures is a simplification of the random activation of the map which would have been caused by a motor-babbling process producing the execution of random grip actions (the use of the regular activation of the map instead of a random process leads to similar effects and speeds up learning; cf. Caligiore et al., 2008). Triggering learning when the hand has a three-finger grip on the object can be considered a simple kind of reinforcement learning process (cf. Berthier, Rosenstein, & Barto, 2005; Bonaiuto & Arbib, in press; Ognibene, Rega, & Baldassarre, 2006).

When learning was triggered, the all-to-all connection weights between AIP and PMCl neurons were updated on the basis of a Hebb covariance learning rule (Sejnowski, 1977; Dayan & Abbott, 2001) to form associations between the perceived shape of the object (AIP) and the corresponding hand posture (PMCl). Formally, the Hebb rule was as follows:

$$\Delta w_{ji} = \eta\left(a_j - \bar{a}_j\right)\left(a_i - \bar{a}_i\right)\left(w_{max} - \left|w_{ji}\right|\right) \qquad \bar{a}[t + \Delta t] = (1 - \xi)\bar{a}[t] + \xi a \tag{7}$$

where η is a learning rate (η was set to 4), w_{max} keeps the connection weights within a small range (w_{max} was set to 0.1), a_j is the activation of PMCl neuron j, a_i is the activation of AIP neuron i, $\bar{a}$ refers to the moving decaying averages of the activation of neurons a_j and a_i, denoted respectively with $\bar{a}_j$ or $\bar{a}_i$ (ξ was set to 0.8; these averages were set at zero at the beginning of each trial). This rule strengthens the connections between each set of neurons which have both an activation above or both an activation below their own average activation, and weakens their connections in the other cases. Note that although simple, this process allowed the system to store various grips with different apertures depending on the different size of objects.

During motor babbling, the ventral stream acquired the capacity to categorise objects on the basis of their appearance. In particular, VC-VOT connection weights formed the categories of objects on the basis of a Kohonen learning rule (Kohonen, 1997; Ritter, Martinetz, & Schulten, 1992):

$$\Delta w_{ji} = \mu \cdot \exp\left[-\frac{\left\| v_{j^*} - v_j \right\|^2}{2\sigma^2} \right] \cdot \left(a_i - w_{ji}\right) \tag{8}$$

where μ is a learning rate (μ was set to 1), v_j is the vector point corresponding to the position in the VOT of the VOT neuron j whose incoming weights are being updated, v_{j*} is the vector point corresponding to the position in the map of the VOT neuron with maximum activation ("winning neuron"), σ is the width of the Gaussian activation function of VOT neurons (σ was set to 0.55), a_i is the activation of the VC neuron i, w_{ji} is the connection weight between the VC neuron i and the VOT neuron j.

Emphasised heading of paragraph: Phase 1. Learning to interact with objects during life: experiment with object parts

Before the experiments with object parts, the model underwent a learning phase which simulated learning to reach objects experienced by real participants during life. This learning phase involved acquiring the reaching capability within the dorsal stream (in particular, this process updated PRR-PMCd connections) and forming the identity of objects within the ventral stream. As in the experiment with images, the eight objects of the experiment were repeatedly presented to the system (trials). During these presentations, SSC encoded the eye gaze direction and the PRR the object position. At the same time, the ventral stream acquired the capacity to categorise objects on the basis of their appearance by updating VC-VOT connection weights (Equation 8) as done for the experiments with image and words.

PRR-PMCd connection weights were updated on the basis of a second "motor babbling" process and the same Hebb covariance learning rule used in the experiments with images and words (Equation 7; in this case w_{max} was set to 0.4 and η was set to 38). This allowed the dorsal pathway to acquire the capacity to perform the correct reaching actions depending on the positions of objects in space. The motor babbling process involved these phases (here one trial lasted few steps): (a) the hand was located in one position randomly chosen from three different positions on the working plane representing the locations in space of the buttons to be pressed during the experiment; (b) the eye focused the hand barycentre on the basis of the focussing reflex; (c) the neurons of PMCd were activated on the basis of a Gaussian function centred on the current arm posture (Equation 3 was used to this purpose); (d) after few steps required by the focussing reflex to stabilise the eye gaze, the object location (encoded in the PRR) and the arm posture (encoded in PMCd) were associated on the basis of the covariance Hebb rule (Equation 7).

The activation of PMCd with the postures of the arm corresponding to certain positions in space is an abstraction of a random activation of PMCd by the motor babbling process which would cause the performance of random reaching actions (the activation of the map in correspondence to only the three spatial positions simplifies and speeds up learning, but the mechanism can allow the model to learn to reach any position in the working space; cf. Ognibene, Rega, & Baldassarre, 2006, and Caligiore, Parisi, & Baldassarre, 2007).

Emphasised heading of paragraph: Phase 2. Learning to name objects: experiment with words and experiment with object parts

In the experiment with words and the experiment with object parts the first learning phase (illustrated in the previous sections) was followed by a second phase which mimicked the acquisition in the early life of the association between the internal representations of object names and the internal representations of the perceived object features. In the experiment using objects names this learning phase allowed the model to develop STCn-AIP connection weights, that is the associations between object names and the representations of their shape. In the experiment with object parts it allowed the model to develop STCn-PRR connection weights, that is the associations between object names and the representations of their experienced position in space. In both experiments this learning phase also allowed the model to develop STCn-VOT connection weights, that is the associations between object names and object identities.

During the learning phase the model interacted several times with objects (trials). In the experiment with words, the following processes took place at each trial (which in this case lasted only one step): (a) the model perceived an object and this activated VC; (b) VC activated AIP; (c) VC activated the VOT; (d) the object noun activated STCn; (e) the Hebb covariance learning rule (Equation 7) was used to update the weights of STCn-AIP and STCn-VOT connections.

In the experiment with object parts, these processes took place at each trial (lasting only one step): (a) the model perceived an object and this activated VC and SSC; (b) SSC activated the PRR; (c) VC activated the VOT; (d) the object noun activated STCn; (e) the Hebb covariance learning rule (Equation 7) was used to update the weights of STCn-PRR and STCn-VOT connections. Note that in this experiment, when object parts were presented the VOT was activated not only on the basis of the image of the object part but also with the image of the relative whole object ("car" for the object parts "roof" and "wheels"; "doll" for the object parts "head" and "feet"). Indeed, during learning in real life the real subjects likely experienced the whole objects before their parts. This implied that, after learning, when parts of objects were named the model activated not only the representation of the knowledge related to them but also the representation of the knowledge on the whole corresponding object. This assumption is important as it allowed the model to solve the task, that is to indicate if object parts belong to the whole objects or not.

During training, the maximum value w_{max} of STCn-AIP, STCn-PRR and STCn-VOT connection weights was set respectively to 0.1, 0.2 and 0.3 whereas their respective learning rates were set to 3, 15, 16. The reason for these choices is that STCn-VOT connections need to be stronger than those of STCn-AIP and STCn-PRR to allow the ventral pathway to exert control on actions and, when necessary, to overwhelm the signals from the dorsal pathway.

Emphasised heading of paragraph: Phase 3. Learning to accomplish the experimental tasks: experiment with images, experiment with words, and experiment with object parts

The third learning phase mimicked the acquisition, within the ventral pathway, of knowledge necessary for accomplishing the psychological experiment. The learning processes of this phase were intended to abstract the formation of associations between stimuli and responses in prefrontal areas, happening during the initial phase of the experiment, on the basis of trial-and-error processes and instruction-guided learning. This training phase also allowed the system to develop a PFC control during the execution of the "ecological tasks" learned in phases 1 and 2 (note that in ecological conditions the prefrontal and dorsal control signals always match). In the model, this PFC control was not necessary to execute the ecological tasks, as the dorsal pathway was sufficient to correctly execute the actions suggested by the object affordances. However, this acquisition was simulated because it was useful to obtain the prediction of the model presented in a following section and because it is biologically plausible (as the ventral pathway, fundamental for the psychological experiment, is also present during the execution of the ecological task).

The learning phase involved repeated interactions with the objects (trials) in either the experimental or the ecological condition. At each step of the trials the following processes took place: (a) the model perceived an object and this activated VC (and also SSC in the experiment with object parts) and hence AIP (or the PRR in the experiment with object parts) and the VOT; (b) STCi was activated with the specific pattern corresponding to either the psychological or the ecological condition; (c) the hand (or arm in the experiment with object parts) performed the grip (or the reaching action) requested by the ecological or the psychological tasks: in the psychological task the grip (or reach) was performed on the basis of the dorsal pathway responding to the affordances of objects; in the experimental task it was assumed that the correct grip (or reach), dependent on the experimental instructions and stimuli, was performed thanks to memories and processes related to such instructions (these were not explicitly simulated as not relevant for this research); (d) the connection weights between STCi and the VOT, considered as a whole input neural layer, and PFC, considered as an output neural layer, were updated using the Kohonen algorithm (Equation 8); (e) the weights of the PFC-PMCl connections (or PFC-PMCd connections in the experiment with object parts) were updated on the basis of the Hebb covariance learning rule (Equation 7; in the equation, w_{max} was set to 0.3 and 0.65 for respectively PFC-PMCl and PFC-PMCd connections, and η was set to 5 for both).

The training of (STCi VOT)-PFC connection weights allowed the system to acquire suitable representations of "goals" within PFC, that is the representations that are necessary to decide which action corresponds with which combination of psychological task and object identity. The training of the PFC-PMC connection weights allowed the system to associate the representation of the particular combination of task and object identity, encoded as a whole in PFC, to the suitable action encoded in PMC and required by the psychological or the ecological task.

Title level 1: Results

This section reports and discusses the results of the simulations directed to replicate and account for the three target experiments. The section also presents some predictions related to the experiment with images and the experiment with object parts.

Title level 2: The experiment with images

Title level 3: TRoPICALS reproduces and provides an interpretation of the results of the experiment with images of Tucker and Ellis (2004)

To replicate the results obtained by Tucker and Ellis (2004) using images, the simulated participants underwent the first learning phase mimicking the acquisition of grasping skills happening "during life" (i.e., before the experimental test). To this purpose, each simulated participant learned to grasp objects for 16,000 cycles (each object presentation lasted 200 cycles). This learning process allowed the participant to become able to trigger, via the dorsal pathway, grasping actions based on the affordances evoked by the seen objects (i.e., a power grip for large objects and a precision grip for small objects). In parallel, in the ventral pathway the VOT formed eight different clusters of neurons representing the eight different objects of the task (Figure 4). After these learning processes, when the system sees an object VC activates and encodes its shape and colour, AIP activates and encodes the object shape (Figure 4), and finally PMCl gains activation until some of its neurons reach the action-triggering threshold and cause the execution of either a power or a precision grip suitable for the object size through the simulated “muscles” and hand (Figure 5).

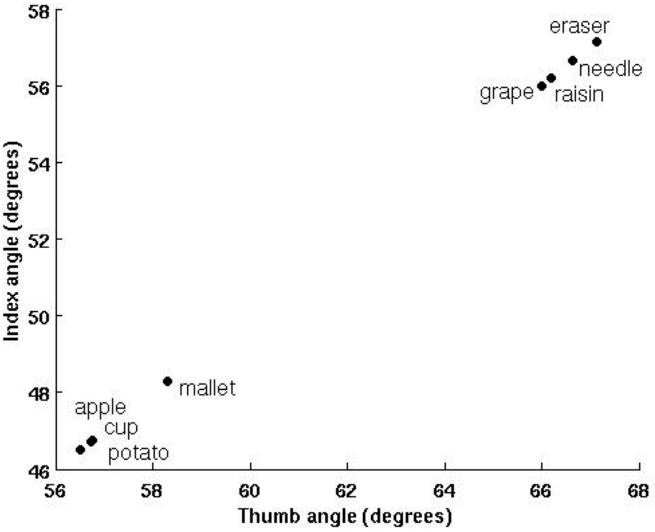

Figure 5: Experiment with images: angles of hand aperture stored by one simulated participant in correspondence to the eight objects of the experiment (note: small angles correspond to large apertures). Notice how the objects tend to cluster into two categories: small objects and large objects.

This first learning phase was followed by a further learning phase which completed the training of the ventral pathway and mimicked both learning to accomplish the experimental task on the basis of the instructions and learning to exert prefrontal control during the performance of the ecological task. This second learning procedure was based on the presentation of the eight objects for further 16,000 cycles. During these presentations, STCi was activated with either one of the two patterns corresponding to the ecological grasping task or the categorisation experimental task. When STCi was activated with the ecological task pattern, all the participants performed grasping actions corresponding to the size of the objects. When STCi was activated with the experimental task pattern, half of the participants performed the categorisation actions corresponding to the “natural-power grip” and “artefact-precision grip” associations and the other half performed the actions corresponding to the reversed associations. During these training phases, the eight objects caused the reactivation of the eight different VOT neuron clusters learned during life and PFC learned, with the Kohonen algorithm, to activate sixteen different neuron clusters corresponding to the combinations of the two ecological/experimental tasks with the eight objects (Figure 6). Note how this rich representation of objects and different contexts allows PFC to bias action selection in any possible way, for example in the way established by the arbitrary rule of the psychological experiment (e.g., power grip for artefacts and precision grip for natural objects or, vice-versa, power grip for natural objects and precision grip for artefacts).

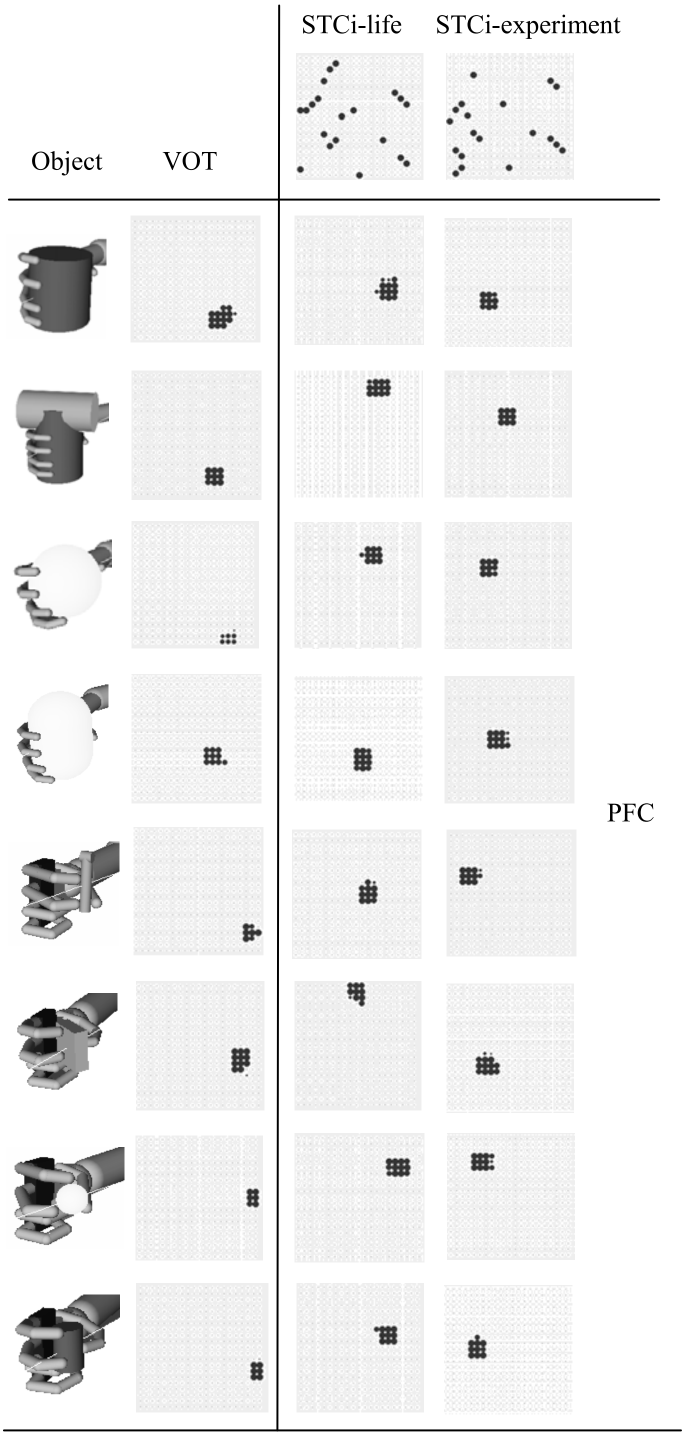

Figure 6: Experiment with images: activation of the PFC caused by the eight different VOT activations (corresponding to the eight objects) and the two STCi activations (ecological condition experienced during life and experimental condition experienced during the psychological test). Notice the different representations of the various contexts and objects within the PFC: the richness of such representations is at the basis of the potential of the PFC to flexibly associate any action to any condition.

These learning processes allow the system to reproduce the main result of the experiments of

Tucker and Ellis (2004). Figure 7 compares the RTs of the simulated 20 participants and the RTs exhibited by the real participants in the target experiment. While these data were collected the learning processes were blocked. A two-way ANOVA on RTs was performed with the factors “congruency” (congruent vs. incongruent) and “object size” (large vs. small). In agreement with the experiments run with real participants, RTs were faster in congruent than in incongruent trials (*F*(1, 16)=3.16, $p = 0.080$) whereas there was not a significant difference between small and large objects (*F*(1, 16)=0.00, $p = 0.997$).

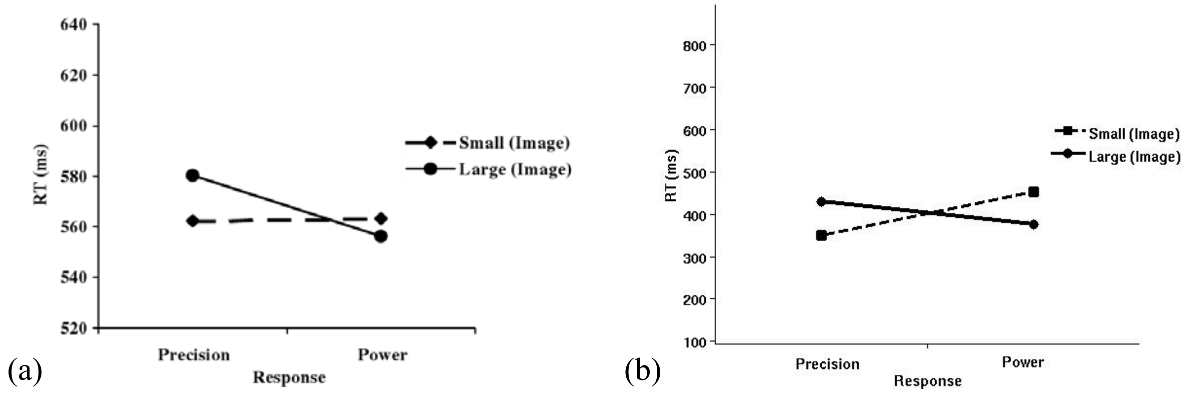

Figure 7: Experiment with images: reaction times (y-axis) versus kind of grip (x-axis). (a) Average reaction times exhibited by the real participants of the experiments of Tucker and Ellis (2004) (copyright of Elsevier, Acta Psychologica). (b) Average reaction times exhibited by the simulated participants. Notice the compatibility effect: large objects tend to speed up RTs of power grips whereas small objects tend to speed up RTs of precision grips.

The model provides a detailed account of this result. During the experimental test, STCi is activated with the pattern corresponding to the categorisation task. In incongruent trials, the ventral pathway (VC-(STCi and VOT)-PFC) evokes an action which is different from the one evoked by the dorsal pathway (e.g., a power grip to categorise a raisin as natural) via a suitable bias issued by PFC to PMCl, thus causing a conflict within the latter (Figure 8a). As the PFC-PMCl signal is stronger than the AIP-PMCl signal, the bias from PFC wins the competition (e.g., by triggering a power grip) but the RTs are relatively long. In fact, when the PFC- and the AIP-induced clusters mismatch they lead to a slower charge of PMCl leaky neurons which will win the competition, so these neurons take longer to reach the threshold required to trigger the corresponding action. On the contrary, in congruent trials (Figure 8b) the signals from PFC and AIP match and so converge on the same action representations within PMCl and so cause relatively fast RTs.

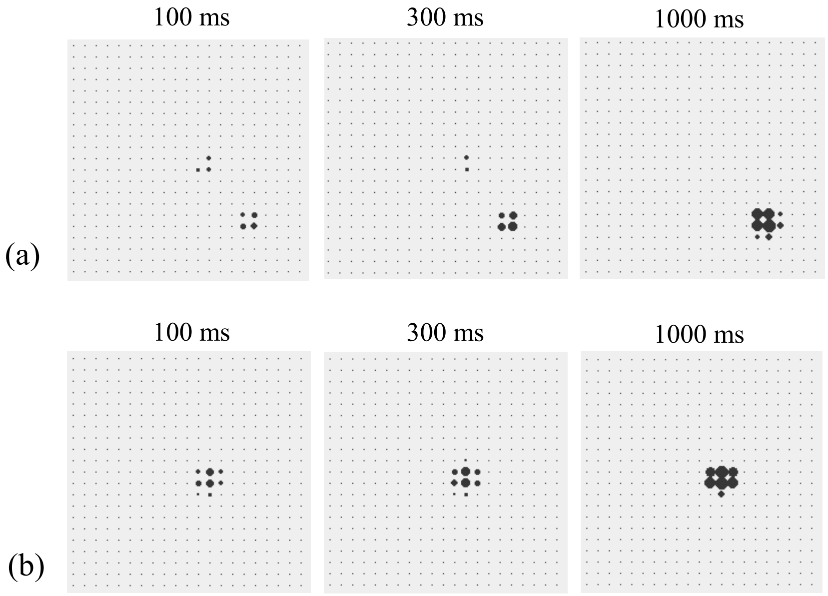

Figure 8: Experiment with images. (a) Activation of the PMC in an incongruent trial: the biases from the PFC and the PC cause two different clusters of neurons to compete until the cluster caused by the PFC suppresses the cluster caused by the PC and triggers the suitable action. (b) Activation of the PMC in a congruent trial: the biases from the PFC and the PC overlap and cause the formation of only one cluster of neurons. The graphs report the activation of the PMC after 100, 300, and 1000 ms.

Title level 3: Predictions of the model: effects of prolonged training on reaction times

The model produced some predictions which might be tested in experiments with human participants. The first prediction concerns the effects of a very long training, based on the experiment with objects, on the performance of participants. In this respect, it is important to notice that in the model "learning during life" mimics the slow learning processes underlying the formation of habits (Salmon & Butters, 1995), whereas "learning during the experiment" mimics the fast learning/problem solving processes taking place in prefrontal areas (Pasupathy & Miller, 2005; Seger & Cincotta, 2006). The simulation leading to the prediction mimics a condition where the task of the psychological experiment is performed so many times ("overtraining") that the knowledge acquired by the ventral pathway has time to be also acquired by the dorsal pathway. The prediction is that this overtraining causes the following effects: (a) the size of the hand aperture when grasping objects in ecological conditions will increase for small objects and decrease for large ones; (b) compatibility effects will diminish, in particular the differences of RTs in congruent versus incongruent conditions will decrease.

The prediction was obtained using the following procedure. First, six simulated participants went though the learning phases illustrated in the above sections (in particular, training of the dorsal pathway based on object affordances and training of the ventral pathway based on the experimental and ecological conditions). Then, they performed the experimental test for 16,000 cycles while the learning process involving the dorsal pathway was reactivated (recall that this learning was blocked to collect the data reported in Figure 7). In each trial, when an object image was presented and the neural competition taking place in PMCl reached the action-triggering threshold, AIP-PMCl connection weights were trained on the basis of the Hebb covariance learning rule (Equation 7).

This training led the winning action encoded in PMCl, which was coherent with the PFC bias and hence with the experimental task, to associate with the object representation in AIP. In congruent cases this training further strengthened the dorsal tendency to select grasping actions coherent with the size of objects, whereas in incongruent cases it caused the dorsal pathway to slowly learn, to some extent, to perform actions coherent with the experimental task (for example to use a precision grip with a potato if the task required such grip to classify the object as natural).

Figure 9 reports the index and thumb angles produced by the model in an ecological experiment before and after the overtraining, averaged for the incompatible cases and averaged for the six simulated participants. The graph shows that after training the hand aperture for large objects decreases whereas for small objects it increases. Two one-way ANOVAs with participants as random factors were performed, one on large and one on small objects, in which the independent factor was the training (before vs. after training) and the dependent variable was the distance between the thumb and the other fingers. Both ANOVAs were significant: for the large objects the fingers distance was reduced after training ($F(1, 5) = 256.26$, $p < 0.0001$) whereas for the small objects the finger distance increased after training ($F(1, 5)=101.74$, $p < 0.0001$). This result shows that the dorsal pathway acquires the tendency to perform actions in agreement with the experimental task instead of the size of objects. Interestingly, the system continues to correctly perform the ecological task thanks to the PFC bias which remains capable of distinguishing the experimental and ecological conditions on the basis of the input from STCi.

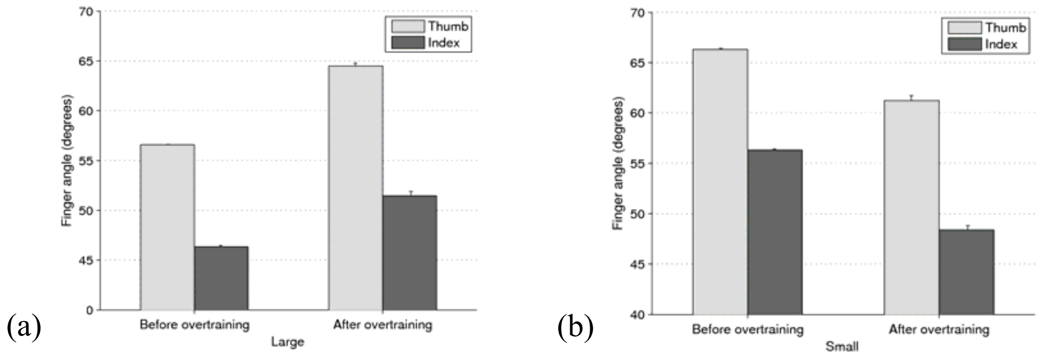

Figure 9: Prediction of the model in relation to the experiment with images: average angles of the hand aperture (Thumb and Index) with incompatible objects before and after the overtraining of the dorsal pathway with the experimental task (recall that large angles correspond to small hand apertures). (a) Averages for large objects. (b) Averages for small objects. Notice how after the overtraining the hand aperture decreases for large objects and increases for small objects, so getting more similar to the responses requested by the experimental task.

We also measured the RTs before and after the training with the same modalities illustrated in the above sections and used to reproduce the compatibility effects (so including both compatible and incompatible cases). The results of this measurement, reported in Figure 10, show that the overtraining leads to an attenuation of the compatibility effects as the interference with the execution of the experimental task caused by the dorsal pathway progressively fades away. In the ANOVA performed with the within-participant factors of “grip” (precision vs. power), “object” (small vs. large) and Training (before vs. after) the 3 way interaction was significant ($F(1,5)=6.91$, $p < 0.05$). Post-hoc Newman Keuls tests showed that with both small objects and with large objects the difference between compatible and non compatible grip was statistically significant before training (respectively, $p = 0.004$; $p = 0.039$), whereas it was not significant after training. If confirmed by tests run with real participants, these results would further support the hypothesis that the causes of compatibility effects reside in the automatic response tendencies of the dorsal pathway.

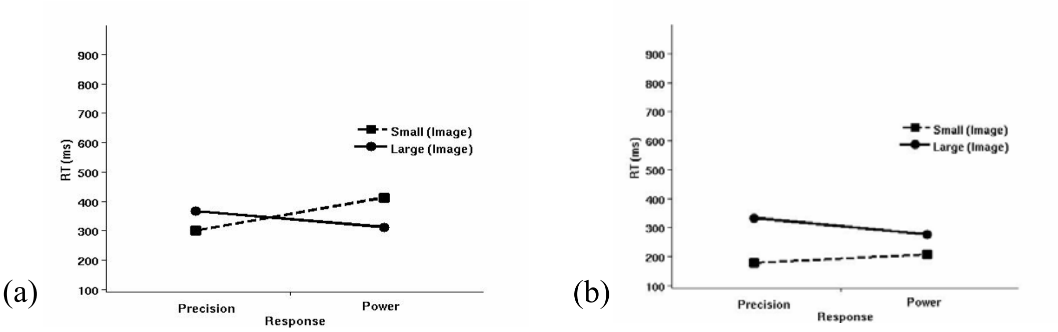

Figure 10: Prediction of the model in relation to the experiment with images. (a) Reaction times before overtraining the dorsal pathway with the experimental task. (b) Reaction times after the overtraining.

Title level 3: Discussion

The four principles incorporated by TRoPICALS are at the core of the interpretation of the target experiments furnished by the model. In this respect, the differences in the RTs obtained in congruent and incongruent conditions can be explained in terms of the cooperation/competition of the biases exerted on action selection by the ventral and dorsal pathways (first principle). In congruent conditions the two pathways tend to select the same action and so the bias from PFC incorporating the goals of the participants, here related to the experimental task (second principle), sums up with the motor signal produced by the affordance, automatically elicited by the object via the PC. This causes PMC to rapidly accumulate evidence in favour of one same action and to produce fast RTs (third principle). On the contrary, in incongruent conditions the signals of the two pathways mismatch and as a result the ventral pathway has to suppress the tendency of the dorsal pathway to automatically elicit actions that are congruent with the affordances of objects but are not in line with the participants' goals. This leads PMC to inhibit the desired response and decide on the correct action. This process brings about a delay and therefore produces longer RTs (third principle).

The account presented here supports the theoretical positions for which in particular conditions motor information and affordances tend to be automatically activated independently of the goals of participants (cf. Borghi et al., 2007; Castiello, 1999; Creem & Proffitt, 2001; de' Sperati & Stucchi, 1997). As assumed in the model, the dorsal system is a good candidate for encoding visuomotor representations based on physical characteristics of objects, and for biasing the activation of low-level relevant object-action associations, based on past history of the individuals' interactions with objects (Tucker & Ellis, 2001). In particular, when goals are unspecified, or only loosely specified (as it happens in ecological conditions) such low-level visuomotor transformations receive a high priority and guide behaviour. When present, more complex semantic knowledge about the objects, suitably integrated with external directions or rules (as it happens in the model during the psychological test), can override the biases automatically generated by affordances (Goodale & Humphrey, 1998; Wallis, Anderson, & Miller, 2001).

This "action-based account" of compatibility effects contrasts with (or, at a more abstract level, refines) other "disembodied" accounts, for example those based on putative "abstract codes". These constructs have been again used, for example, to explain the Simon effect (Lu & Proctor, 1995). The idea is that target stimuli automatically activate abstract "location codes" and these codes tend to influence responses if they have spatial components, thus resulting in a facilitation or an interference effect (Kornblum, Hasbroucq, & Osman, 1990). This account faces the difficulty of explaining why abstract codes related to objects *automatically* activate spatial components of responses (Tucker & Ellis, 1998). Action-based accounts such as the one presented here overcome this difficulty thanks to the core assumption on the continuity existing between spatial aspects of sensory representations and spatial aspects of motor representations within visuomotor transformations. Indeed, such action-based accounts propose that semantic representations inherently include motor components (Borghi et al., 2004; Barsalou, Santos, Simmons, & Wilson, 2008).

However, the explanation of compatibility effects presented here bears much in common with the "dual-route models" which are used to explain compatibility effects related to the Simon effect (e.g., see the same Kornblum, Hasbroucq, & Osman, 1990, which, as aformentioned also proposed the idea of location codes; see also De Jong, Liang, & Lauber, 1994; Toth et al., 1995; Kornblum, Stevens, Whipple, & Requin, 1999; cf. also the last sections on this). Although differing in details, dual-code models agree that different aspects of stimuli activate two different processing pathways, one which is slow and is based on working memory and one which is fast and is based on long-term memory. The congruency effect derives from the fact that the fast automatic activation primes the code for the correct response in congruent trials whereas it hinders it in incongruent trials. The model presented here is an advancement with respect to dual-code models in several aspects: (a) it is formulated in an "fully expanded" and operational form (i.e., in terms of an explicit computational model) whereas the dual-code models have been formulated only verbally: this allows for reproduction of the detailed mechanisms behind the generation of different RTs and for the formulation of specific predictions; (b) it is based on specific neuroscientific hypotheses on the brain architecture and mechanisms behind the "dual route" hypothesis; (c) it is tested in an embodied setup; (d) it reproduces the learning processes leading to the acquisition of the target behaviours (Kornblum, Stevens, Whipple, & Requin, 1999, propose a connectionist model but with hand coded connection weights: although very interesting, the model cannot reproduce/predict effects such as those reported below in the section related to the prediction of the effects of overtraining with the task of the experiment of Borghi, Glenberg, & Kaschak, 2004); (e) it represents a theoretical framework with the potential of integrating the explanations of different compatibility effects (see last sections on this).

The model also leads to the formulation of some predictions on the possible effects caused by overtraining in the psychological experiment task. The model shows that overtraining might lead the participants to exhibit less pronounced compatibility effects. Namely, the dorsal pathway would learn the experimental task and so would exert a lower interference on the decisions performed by the ventral pathway. Kinematics measures on the hand aperture of the model indicated that with overtraining the dorsal pathway actually converges towards the grips requested by the experimental task, and RT measures indicated that the compatibility effect diminishes after a long training session. If confirmed by similar tests on real participants, this prediction would corroborate the hypothesis put forward by the model for which the automatic reactivation of knowledge within the dorsal pathway facilitates/interferes with the execution of the experimental task and this is the major cause of the compatibility effects illustrated in Tucker and Ellis (2004). Note that the prediction shares a resemblance with recent findings on the effects of repetition of actions on processing abstract and concrete words (Glenberg, Sato, & Cattaneo, 2008).

An important caveat about the prediction, however, is that in real brains the existence of connections between the ventral and the dorsal pathway (Webster, Bachevalier, & Ungerleider, 1994), or any possible sensory input reaching the dorsal pathway and informing it on the experimental/ecological context, might lead the dorsal pathway to acquire the capacity to properly respond to *both* experimental and ecological tasks. As currently the model does not have any of these connections, when the dorsal pathway acquires the capability of performing the experimental tasks it also looses the capability of performing the ecological one. This aspect of the model might be improved in the future, for example by establishing a communication between the ventral and dorsal pathways before they converge within PMC or by letting the dorsal pathway to access context information through other pathways. Note, however, that this simplification of the model does not affect the relevance of the prediction according to which a prolonged learning of the dorsal pathway during the experimental task would cause a reduction in compatibility effects.

Title level 2: The experiment with words

Title level 3: TRoPICALS reproduces and provides an interpretation of the results of the experiment with words of Tucker and Ellis (2004)

In order to replicate the results obtained by Tucker and Ellis (2004) with object names, the simulated participants underwent three learning phases: (a) the learning phase simulating “life” and directed to acquire the affordances within the dorsal pathway and object identities within the VOT; (b) the learning phase mimicking the acquisition of simple linguistic associations during life; (c) the learning phase involving the formation of suitable goals within PFC, and suitable top-down biases on actions within PFC-PMCl connections. The first and third of these learning phases were identical to those illustrated in relation to the experiment with object images so only the second learning phase will be explained in further detail. This learning phase lasted 4,000 cycles (200 cycles per object) and led the system to form associations between the names of objects represented in STCn and their shape and identity representations activated by VC respectively in AIP and the VOT. These associations were formed on the basis of the Hebb covariance learning rule (Equation 7).

These three learning phases allowed for the reproduction of the Tucker and Ellis (2004) experiment run with word stimuli. In order to mimic the presentation of the names of objects without their sight, STCn was activated with words patterns whereas VC was not activated (recall that the word patterns were activated progressively from zero and one to simulate the time requested by the processing of words: the effect of this is shown in Figure 11a). Figure 12 shows the results reported by Tucker and Ellis (2004) and the results of the same tests run with the simulated participants. For comparison, the two diagrams of the figure report not only the results of the experiment with words, but also those of the experiment with images. The figure shows that the model reproduces the main congruency effect even if the visual stimuli (the seen objects) are substituted with linguistic stimuli (the names of objects).

A two-way ANOVA on RTs was performed with two factors, “congruency” (congruent vs. incongruent) and “object size” (large vs. small). The congruency factor resulted to be statistically significant: the RTs were faster in congruent than in incongruent trials ($F(1, 16) = 8.84$, $p = 0.004$), whereas there was no significant difference between small and large objects ($F(1, 16)=1.35$, $p = 0.249$). This result is due to the fact that the linguistic input causes patterns of activation in the VOT and AIP similar to those caused by visual inputs (“internal simulations”). This can be seen by comparing Figure 11, which reports the internal activations caused by words, and Figure 4, which reports the internal activations caused by the objects corresponding to those words. Due to this similarity, such “internal simulations” cause activations which diffuse along the dorsal and ventral pathways and cause congruency phenomena similar to those caused by the referents of words.

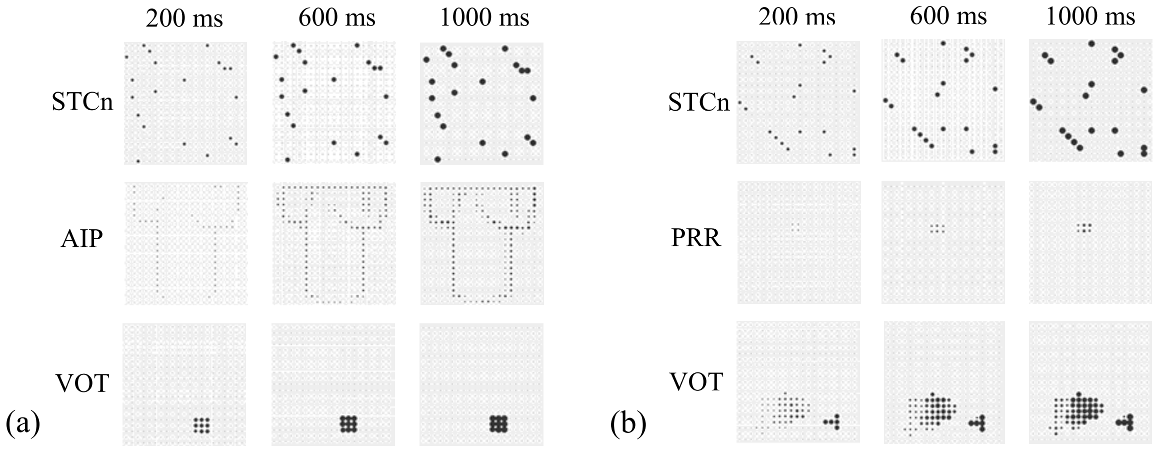

Figure 11: Examples of activations caused by object names after 200, 600 and 1000 ms. (a) Experiment with words: activations caused by the name "mallet". (b) Experiment with object parts: activations caused by the name "feet". Notice the progressive activation of the areas in time.

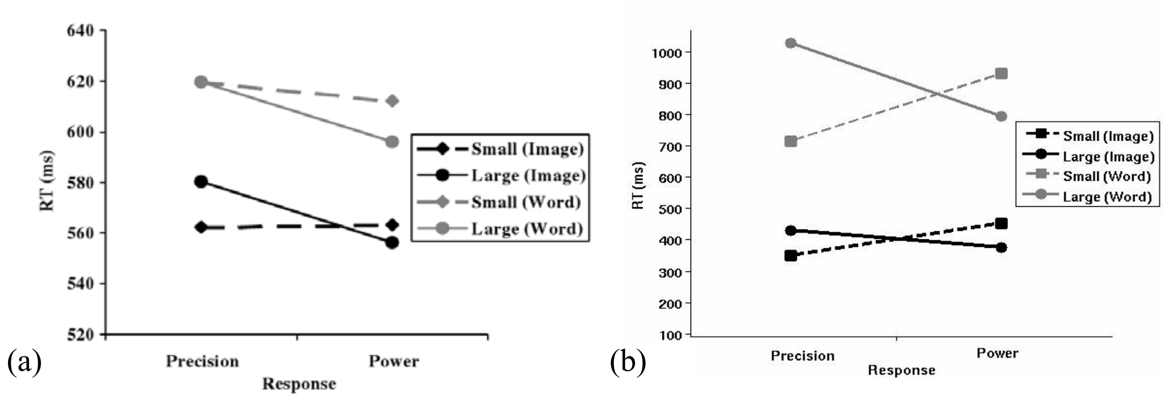

Figure 12: Experiment with words: reaction times (y-axis) versus kind of grip (x-axis). For comparison each graph also reports the reaction times of the experiment with images. (a) Average reaction times exhibited by the participants of the original experiments by Tucker and Ellis (2004) (copyright of Elsevier, Acta Psychologica). (b) Average reaction times exhibited by the simulated participants. Notice how the object names cause a congruency effect similar to that caused by the object images.

Title level 3: Discussion

The mechanisms used to simulate simple associative linguistic processes allowed TRoPICALS to give an explanation of compatibility effects obtained with object names by Tucker and Ellis (2004). The key idea underlying these results is that linguistic representations of words are grounded in the sensorimotor system. Thus they can activate the internal representations of the referents of words (Glenberg & Robertson, 2000). As a consequence, the activation of these representations by words triggers action selection processes affected by the same congruency phenomena caused by the direct experience of objects (or their images).

These results contribute to understand the detailed mechanisms which might underlie theories that propose action-based views of language (Glenberg & Kaschak, 2003; Arbib, 2005; Gallese & Lakoff, 2005; Barsalou, Santos, Simmons, & Wilson, 2008; Gallese, 2008; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008). According to these theories, linguistic representations (e.g., related to the phonological or articulatory aspects of language) are tightly coupled with other non-linguistic sensorimotor representations, and so can give rise to a number of relevant phenomena on the basis of simple associations. Although the whole complex spectrum of language phenomena cannot be accounted for in these terms, these associations seem to be at the right level of abstraction to give an economical account of compatibility effects.

Title level 2: The experiment with object parts

Title level 3: TRoPICALS reproduces and provides an interpretation of the results of the experiment of Borghi, Glenberg, and Kaschak (2004) with object parts

To replicate the results obtained by Borghi, Glenberg, and Kaschak (2004) with object parts, ten simulated participants underwent all the three learning phases as described in relation to the experiment with words. These learning phases involved: (a) the acquisition of affordances within the dorsal pathway (SSC-PRR-PMCd; 6,000 cycles), and object representations in VC-VOT connection weights (20,000 cycles); (b) the acquisition of simple linguistic associations (4,000 cycles) in STCn-VOT and STCn-PRR connection weights; (c) the formation of suitable goals in (STCi VOT)-PFC connection weights, and suitable actions within PFC-PMCd connection weights (6,000 cycles).

One main difference with the experiment with words was due to the involvement of the SSC-PRR-PMCd neural pathway instead of the VC-AIP-PMCl pathway. Figure 13 shows the activations of VC and the PRR caused by the objects and the activations of the VOT after learning. Another important difference relates to the training of the ventral pathway in the second learning phase involving STCn-PRR and STCn-VOT connections. These were trained in the same way as in the experiment with words with the important difference that when the object parts were presented they were associated, within VOT, with both the representation of the identity of the object parts and with the representation of the identity of the whole object they belonged to (i.e., “roof” and “wheels” were associated with their own representations and with the “car” representation; “head” and “feet” were associated with their own representations and with the “doll” representation). Figure 11b shows an example of activation of the PRR and the VOT caused in time by the name “doll” after this training (recall that name patterns activated progressively). The effect of the particular training of the VOT can be seen in Figure 14, clearly showing how names of object parts activate not only their own representations but also the representations of the whole objects they belong to (cf. the activations of the VOT shown in this figure with those shown in figure 13 and elicited by object images). Figure 14 also shows how the third training phase led PFC to develop different representations which depended not only on objects but also on the condition faced by the system (experiment vs. ecology).

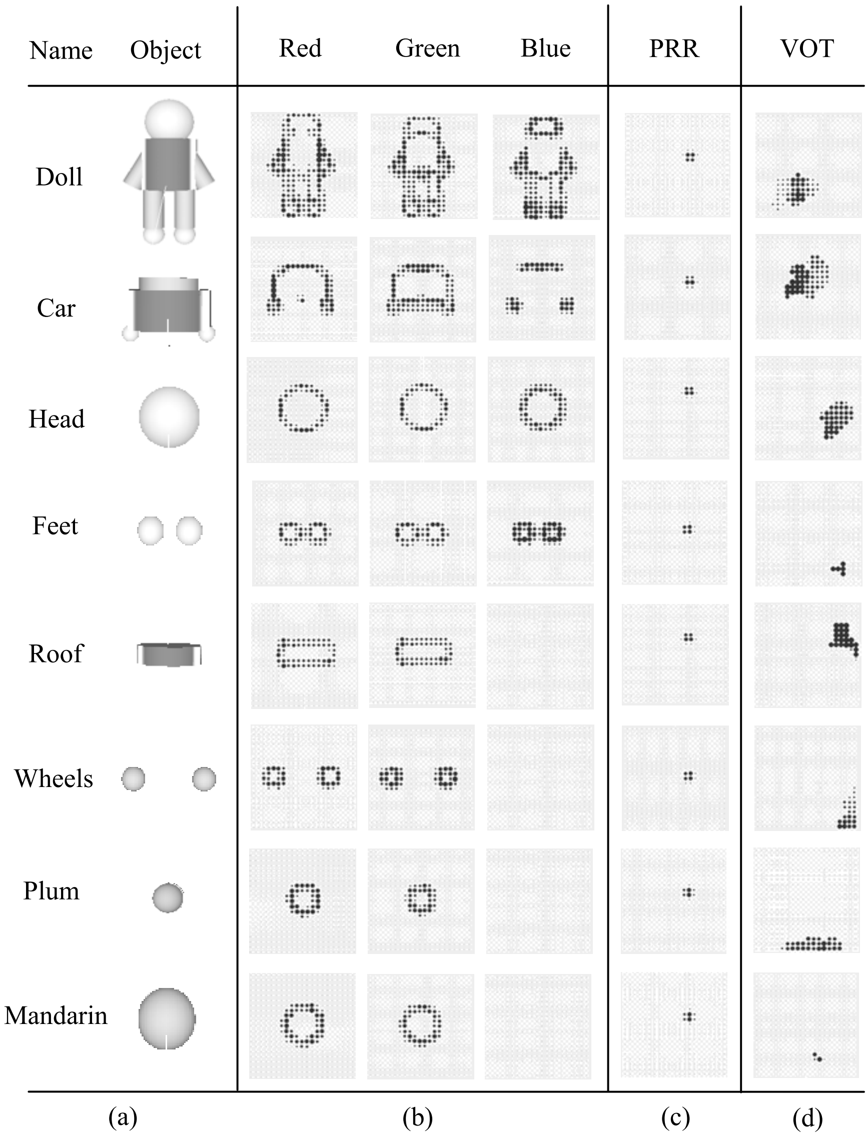

Figure 13: Experiment with object parts. The columns of the figure report from left to right: (a) The object name and appearance; (b) The activation of the VC (i.e., the three RGB maps) encoding the colour-based edges of the eight objects; (c) The activation of the PRR encoding the position of objects (note that the activation of this map differs minimally in correspondence to different positions of objects because the map covers the whole space reachable by the arm); (d) The activation of the VOT encoding the identity of objects. Notice how the PRR and the VOT represent respectively the position and identity of different objects in distinct ways.

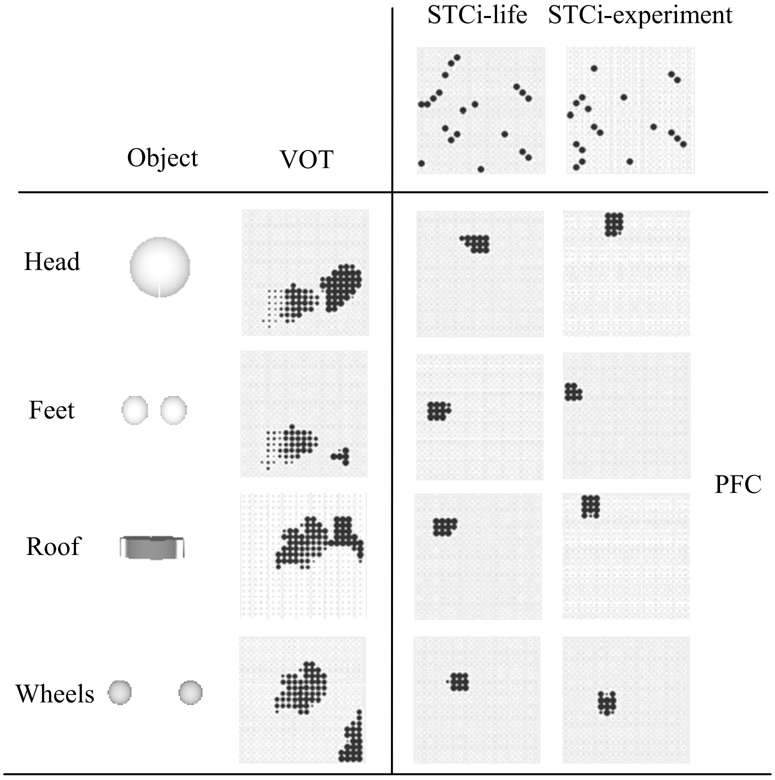

Figure 14: Experiment with object parts: activation of the PFC caused by four different VOT activations (corresponding to the four object parts) and the two STCi activations (corresponding to the functioning of the model in the "simulated life" and during the simulated psychological experiment). Notice how within the VOT the object parts reactivate both their own representation and the representation of the whole object they belong to.

The three training phases allowed the system to acquire the capabilities of reproducing the target experiment with object parts of Borghi, Glenberg, and Kaschak (2004). To see this, STCn was activated with words patterns corresponding to the names of the object parts (roof, wheels, head and feet) and the RTs were recorded. Figure 15 shows the results of the tests carried out with real participants and the results of the same tests carried out with the simulated participants. The figure shows that the model reproduces the target congruency effect. A two-way ANOVA on RTs with participants as random factors and "part location" (upper vs. lower part) and "response location" (yes-is-up or yes-is-down) as factors was performed. As in the target experiment, the interaction was significant (*F* (1, 4) = 20.77, p < 0.02). Newman-Keuls post-hoc showed that with upward movements upper parts (*M* = 900) produced significantly faster RTs than lower parts (*M* = 1102; *p* = 0.05), whereas the opposite was true with downward movements (*M* = 1042 vs. *M* = 924). In the last case the difference only approaches significance (*p* = 0.07), but this is probably due to the low number of the involved participants.

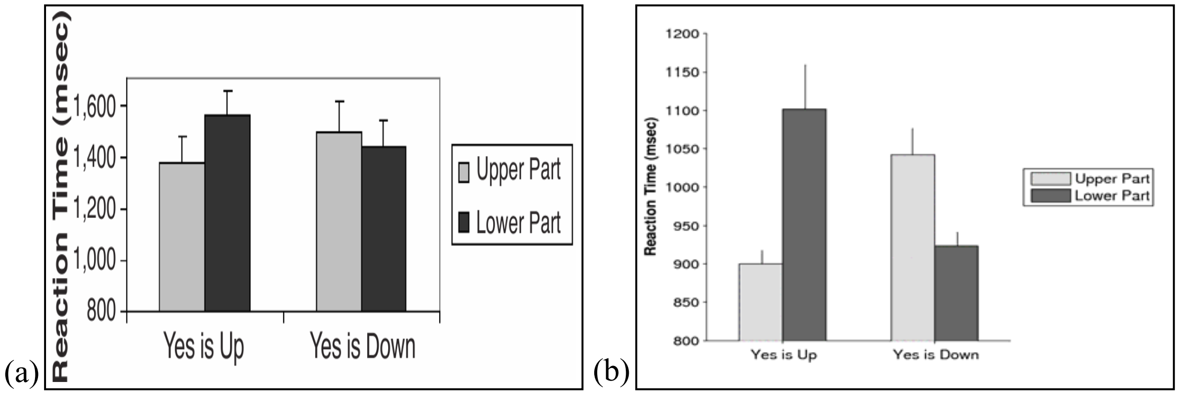

Figure 15: Experiment with object parts: reaction times (y-axis) versus kind of requested answer (yes-is-up" and "yes-is-down") and location of target objects (up or down). (a) Average reaction times exhibited by the participants of the original experiments by Borghi, Glenberg, & Kaschak (2004) (copyright of Memory & Cognition, Psychonomic Society, Inc.). (b) Average reaction times exhibited by the simulated participants. Notice how the positions of the buttons favour the RTs of the answers requested by the psychological decision task ("yes-is-up/no-is-down" or "yes-is-down/no-is-up") when they are spatially congruent with them, whereas they cause an interference in the case of incongruence.

The obtained result can be explained as follows. In the ventral pathway, the names of object parts cause a reaching motion directed to the buttons as requested by the experimental task. At the same time, in the dorsal pathway the stimuli tend to cause a reaching action directed to the spatial location that those parts had during the ecological training. In incongruent trials, the ventral pathway (STCn-(STCi and VOT)-PFC-PMCd) evokes a different action with respect to the dorsal pathway (STCn-PRR-PMCd), for example a "yes-up" answer to say that wheels are part of the car. This causes a conflicting signal arriving to PMCd. As the PFC-PMCd signal is stronger than the PRR-PMCd signal, the bias from PFC wins the competition but the RTs are relatively long. On the contrary, in congruent trials the two signals are congruent and so cause relatively fast RTs.

Title level 3: Predictions of the model: compatibility effects caused by the embodied representations of words

Thr version of the model used to reproduce the experiment with object parts produced specific predictions which might be tested in experiments with real participants. These predictions refer to the possible effects that the perception of a first word, referring to an object having a particular spatial location (for example the upper part of an object), might have on the RTs of the action needed to indicate the spatial location of an object indicated by a second word. For example, the name "roof" indicated as a first word should speed up the "up response" used to indicate that the object indicated by a second word, for example "head", is located "up", even if the two words refer to different objects. On the contrary, it should slow down the "down response" used to indicate that the object related to a second word, for example "feet", is located "down". This happens because the first word automatically activates internal simulations involving a particular spatial portion of the referent object (e.g., the upper part). This increases or decreases the RTs needed to trigger actions having a spatial component respectively compatible or incompatible with the one which was initially simulated.

In detail, the predictions of the model were obtained as follows. First, six simulated participants went though the three learning phases illustrated in the previous section. After these trainings, the model perceived a first word indicating an object part for 0.8 sec (this short time prevented the system from triggering a reaching action related to the first word). This presentation was followed by the presentation of a second word indicating another object part. In particular, we used couples of words where the first word referred to the object parts related to the "doll" ("head" and "feet") and the second word referred to the object-parts related to the "car" ("roof" and "wheels"). In this way the two object parts were not related to the same object: this is intended to represent a more general experimental condition for which the only semantic relationship existing between the two object parts presented in sequence is in terms of their spatial locations. The task asked the participants to indicate the spatial location of the object related to the second word by performing a reaching movement directed either to the "up button" or to the "down button"

Figure 16 shows the RTs of the model for triggering the requested actions in the cases in which the second names of objects were preceded by names of objects with compatible or incompatible spatial locations. The ANOVA performed on RTs showed a compatibility effect ($F(1, 4) = 7.28$, $p = 0.054$): the RTs of the reaching actions used to indicate the spatial location of the second-word object sped up or slowed down when preceded by words referring to objects with respectively compatible and incompatible spatial locations. The reason for this is that the presentation of the first word affects the selection of the response to the second word as follows: (a) the first word activates the spatial location of the indicated object within the PRR; (b) the PRR activates PMCd and this starts to accumulate activation in favour of a reaching movement directed to such spatial location; (c) if the spatial locations of the first and second words/objects are compatible, the presentation of the second word causes an activation of the PRR and PMCd similar to the one caused by the first word and this results in relatively fast RTs. On the other hand, if the spatial locations of the first and second words/objects are compatible they cause different activations in the PRR and PMCd and the second word has to suppress the previous activation in PMCd and this produces relatively slow RTs.

This prediction has some similarities with the results of the experiment run by Zwaan and Yaxley (2003). This experiment involved the presentation of pairs of words referring to parts of larger objects (e.g., attic-basement). Participants were required to judge whether the two words were semantically related or not. RTs were significantly faster when the words were presented in an iconic relation with their referents (e.g., attic presented above basement) than in a reverse-iconic relation (basement above attic).

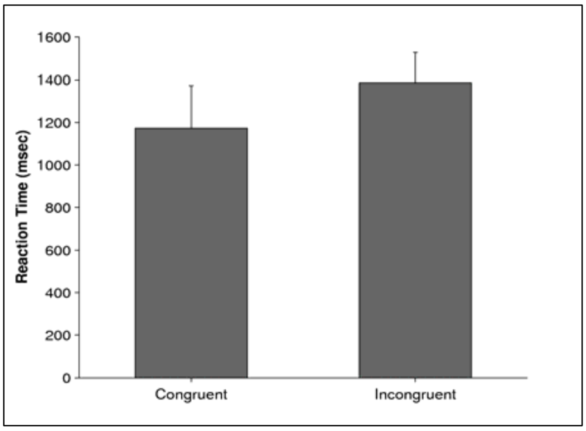

Figure 16: Prediction of the model in relation to the experiment with object parts: reaction times for a button located in a position corresponding to the position of an object part in two conditions: when the name is preceded by the presentation of a name of an object having same spatial location ("Congruent") or different spatial location ("Incongruent"). The reaction times include the display time of the first name (800 ms). Notice how the spatial location of the object indicated by the first name speeds up the RTs of the answer to the second name when it is congruent with the spatial location of the referent of the second word, and vice-versa when it is incongruent with it.

Title level 3: Discussion

The model was able to reproduce the results of the experiment with object parts. This result is based on some hypotheses that have relevant theoretical implications. Consider the way in which word learning occurs within the model. Words get associated with the sensorimotor experiences related to their referents. For example, learning the word "doll" implies that the representation of its phonetic form gets associated with the internal representations of some aspects of its referent (the object doll), namely the object identity within the VOT and the usually-experienced spatial location of the object within the PRR. Importantly, each object part, such as "head" (part of the object doll), is associated with *both* the representation(s) of its referent, i.e. the identity of the object head and its common location within the whole object doll, as well as with the identity of the whole object doll of which it is part. This implies that, due to their grounded relations during experience, for which one object is experienced as part of another, the words "doll" and "head" become indirectly associated via the grounded internal representations of the objects they refer to.

Now, compare this mechanism with one of the most influential current proposals on the representation of meaning of words, the Latent Semantic Analysis (LSA; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998; Landauer, McNamara, Dennis, & Kintsch, 2007). This proposal would likely account for the representation of the part-whole relation. According to the LSA framework, the meanings of words are based on the co-occurrence of linguistic forms of words in text and speech. As a consequence the part-whole relation would be captured by the formation of direct links between the representations of the phonetic forms of words, and not on the basis of the links between representations of objects grounded in experience. Note that the aim here is not to criticize the LSA model: links as those proposed by LSA surely form within the brain (cf. Pulvermüller, 2005), and the formation of direct associations between representations of words likely play an important influence on their meaning and the way they are acquired. Indeed, it has been shown that the LSA approach is able to account for many psychological processes, such as semantic priming effects and memory retrieval processes (Kahana, 1996). However, the model proposed here is able to reproduce the targeted experimental results relying only on sensorimotor grounding of words. This represents an operational hypothesis on the possibility that some important aspects of the meaning of words, such as the whole-part relations, are based on links between words and grounded internal representations of object characteristics, and not only on verbal "word-word" associations.

Title level 1: TRoPICALS: Comparison with other models, possible extensions to account for other compatibility effects, and future work

Title level 2: TRoPICALS and related models

Recently, the literature has proposed various neural-network models which are based on one or more of the four core principles incorporated in TRoPICALS. This section presents a critical review of some of these models, in particular of those which are more closely related to TRoPICALS or which tackle issues related to congruency effects. This comparison has three goals. First, it aims at further clarifying the nature of each of the *four principles* incorporated in TRoPICALS and at highlighting their importance for explaining higher-level cognition and, in particular, compatibility phenomena. Secondly, it illustrates the novelty of TRoPICALS related to the fact that, in contrast with other models, the four principles operate in an *integrated fashion* within it. Last, it identifies a number of relevant mechanisms exploited in other models which might be used to further develop TRoPICALS so as to use it to account for other compatibility effects (as indicated in the next two sections). Note that the reader not interested in these issues can skip this and the next two sections, and directly access the "Conclusions" section, as they are rather independent with respect to the rest of the paper.

The review presented in the section groups the models on the basis of the principle, among the four principles, which is most important for them. Table 2 presents an overview of the models, in particular it summarises the principles they incorporate and the extent to which they satisfy the specific constraints of the computational embodied neuroscience approach. Each model is first briefly illustrated, together with its strengths and limits, and then its relevance is evaluated with respect to the goals mentioned above. The review starts by evaluating the strengths and limits of TRoPICALS itself.

Table 2

Models related to TRoPICALS. The table groups the models depending on which of the four principles incorporated in TRoPICALS they emphasise (bold “**V**”s highlight the grouping). Moreover, the table highlights the degree of attention they give to the four constraints: (a) empirical evidence on brain (Brain); (b) empirical evidence on behaviour (Behav.); (c) embodiment realism (Embod.); (d) reproduction of learning processes underlying the target behaviours (Learn.). A capital “V”, a small “v”, or an empty space represent respectively a large, limited, or null attention given to the principles and the constraints.

	Principle 1: Ventral/ dorsal neural pathways	**Principle 2: Top-down biasing via PFC**	**Principle 3: Neural dynamic competition**	**Principle 4: Language and situated simulation**	**Brain**	**Behav.**	**Embod.**	**Learn.**
TRoPICALS	**V**	**V**	**V**	**V**	**V**	**V**	**v**	**V**
Emphasis on P1								
Fagg et al., (1998)	**V**		V		V	V		
Oztop & Arbib, (2002)	**V**	V	v		V	V	v	v
Gupta & Noelle, (2007)	**V**	V	v		V	v	v	V
Oztop et al., (2004)	**V**		v		V	V	V	v
Borghi, Di Ferdinando & Parisi (2002)	**V**		v			v	v	v
Emphasis on P2								
Deco & Rolls, (2003)		**V**	V		V	V		V
Hazy et al., (2007)		**V**	v		V	V		V
Botvinick et al., (in press)		**V**	v		V	v		V
Polk et al., (2002)		**V**	v		V	V		v
Tsiotas et al., (2005)		**V**				v	v	v
Emphasis on P3								
Erlhagen & Schöner, (2002)		V	**V**		V	V		v
Stoianov et al., (2005)	V		**v**		V	V		V
Plaut et al., (1996)			**v**	v		v		v
Erlhagen & Bicho, (2006)		V	**V**		V	v	V	v
Cisek, (2007)	V	V	**V**		V	V		
Ognibene et al., (2006)			**V**		V	V	v	V
Emphasis on P4								
Garagnani et al., (2008)		V		**V**	V	V		V
Mayor & Plunkett, (2010)				**V**	V	v		V
Mirolli & Parisi, (2005)	V			**V**			v	V
Cangelosi et al., (2006)				**V**			v	v

TRoPICALS was deliberately built on the basis of the four principles and these principles work in a closely integrated fashion within it. In relation to the constraints of computational embodied neuroscience, the overall architecture of the model was constrained with specific neuroscientific data (see Table 1 for an overview). However, further work is needed to improve the biological plausibility of the *micro-architecture* of its constituent parts (e.g., the SOM networks used in the model offer only a rough functional representation of the processes taking place in the VOT and PFC). The model has been validated with the reproduction of the results of three psychological experiments. However it still requires some extensions, and to be challenged with the results of further experiments on compatibility effects, to soundly prove its real potential for accumulation. Although the model has a remarkable potential for tackling "embodied tests", thanks to its integration with a sophisticated humanoid simulated participant, the experiments replicated so far represent a relatively modest challenge for such capability (e.g., they imply only simple "stylised" grasping actions), and the model architecture needs to be further developed to tackle more complex tasks (e.g., involving realistic grasping actions and more than 2-DOFs for reaching; specific mechanisms for doing this might be drawn from Caligiore et al., 2008, and Oztop & Arbib, 2002). Finally, learning plays an important role in the functioning of the model, and particular attention was spent in selecting biologically plausible learning processes (in particular motor babbling, Hebb learning rules, and non-supervised learning rules based on self-organisation).

Emphasised heading of paragraph: Models stressing principle 1:ventral/dorsal neural pathways

The FARS model (Fagg & Arbib, 1998) aims at studying the control of grasping in non-human primates and is strongly based on the neurophysiological findings of Sakata and Rizzolatti (Jeannerod, Arbib, Rizzolatti, & Sakata, 1995; Sakata, Taira, Murata, & Mine, 1995; Rizzolatti, Fogassi, & Gallese, 1997; Rizzolatti, Luppino, & Matelli, 1998). The model pivots on the cortical area F5, part of PMC, and the anterior intraparietal cortex (AIP). According to the model, AIP converts the information processed by the dorsal and ventral neural pathways into a set of representations of different "grasp affordances". These affordances are then forwarded to F5 where the most suitable one is chosen given the task constraints.

Starting from the FARS model, Oztop and Arbib (2002) focus on the linkage between the perception of object-related affordances and the generation of specific grasping actions. The dorsal visual stream (parietal cortex) extracts parametric information about the object being attended to and usable for implementing actions. However, it does not ''know'' what the object is: it can only conceive a set of possible affordances associated with the object. The ventral stream, by contrast, recognizes what the object is and passes this information to PFC. The PCF can then bias F5 to choose the affordance appropriate to the task at hand on the basis of the organism's current goals and the recognition of the nature of the object.

The FARS model and its developments represent a fundamental milestone for the modelling of affordances and their underlying neural mechanisms. The model gives a very detailed account of the formation and selection of affordances. It also stresses the importance of the bottom-up formation of affordances (putatively taking place within the dorsal pathway) and the role of a top-down pathway involving PFC allowing for selection of one of the affordances on the basis of the task at hand. In this respect, an important difference with TRoPICALS is that FARS has the goal of explaining in detail how the dorsal and ventral pathways cooperate to select the most suitable affordance in standard conditions. Instead, TRoPICALS aims to show how affordances are automatically activated both when they are useful and when they cause interferences for the task at hand. In this respect, the two models illustrate complementary phenomena related to affordances.

The computational neural-network model proposed by Gupta and Noelle (2007) is based on the hypothesis that there are two largely distinct neural pathways that respectively guide controlled and automatic behaviours. The model uses the Leabra modeling framework (O'Reilly, 1996) which incorporates two ways of modifying the strength of connections: (a) an error correction learning algorithm; (b) a Hebbian correlation learning rule. The network controls a two joint planar arm to reproduce human experiments in which the participants had to learn sequences of key pressing on a keyboard of nine keys. The controlled pathway learns more rapidly than the automatic one and the automatic pathway in isolation cannot produce correct motor sequences. Moreover, the controlled pathway is able to compensate the errors of the automatic pathway. This model is particularly relevant as it tackles the issue of the mechanisms underlying automatic versus controlled attentive behaviours, and in particular the dependencies of the learning mechanisms involving them: this issue is closely linked with the issue of the automatic versus top-down selection of affordances tackled here (Miller & Cohen, 2001) and an integrated explanation of the two classes of phenomena should be sought in the future.

The Infant Learning to Grasp Model (ILGM; Oztop, Bradley, & Arbib, 2004) is one of the first computational models on infant grasp learning constrained with data from the literature on infant motor development. The ILGM proposes four hypotheses about infant grasp learning: (a) infants acquire the skill to orient their hand towards a target during the early postnatal period rather than innately possessing it; (b) infants are able to acquire grasping skills before they develop an elaborate adult-like object visual analysis capability; (c) action opportunities afforded by the environment are an important factor which shapes infant grasp development; (d) the inability of young infants to pre-orient their hands could be explained by the lack of visual capabilities for extracting object affordances rather than motor immaturity (cf. Lockman, Ashmead, & Bushnell, 1984). The model is relevant for this work for at least two reasons. First, it uses an approach directed to constrain models of affordances on the basis of data from developmental psychology. The use of this type of data might allow challenging and improving the learning processes used by the TRoPICALS (cf. Paletta, Fritz, Kintzler, Irran, & Dorffner, 2007). Second, it informs us on how to improve the realism of the model related to both the representation and development of affordances in the parietal-premotor pathways. For example ILGM is capable of autonomously developing the capability of performing various types of grasps depending on the object affordance (e.g., precision, power, and side grasps).

In Borghi, Di Ferdinando and Parisi (2002) (see also Di Ferdinando & Parisi, 2004) a neural network model was requested to reply with different answers to the same objects in distinct contexts. The neural network had a feed-forward architecture formed by many layers organised in sequence. The simulations showed that with training the layers closer to the actuators developed internal activation patterns which strongly correlated with the actions to be performed. On the contrary, the layers closer to the sensors developed activation patterns which correlated more strongly with the features of objects and less so with the actions to be performed. This model is relevant for this work as it shows a general phenomenon for which activation patterns of areas closer to the output areas tend to develop representations more closely related to actions to produce whereas areas closer to the input areas tend to develop representations more closely dependent on sensations. These important gradients are not currently captured by the learning processes of TRoPICALS for which the different areas are either influenced by perception (VOT, PFC, AIP, and PRR) or by action (PMCl and PMCd). Future work should introduce more biologically plausible learning processes to have such gradients if they reveal important to account for some aspects of compatibility effects.

Emphasised heading of paragraph: Models stressing principle 2: top-down biasing by prefrontal cortex

Deco and Rolls (2003) describe a model based on integrate-and-fire neurons which is used to investigate how the neurons in the primate PFC provide a neural substrate for mapping stimuli and responses in a flexible fashion, based on context and rules. The behaviour of the model is compared with the results obtained in experiments where monkeys are requested to accomplish object or spatial discriminations by performing specific oculomotor responses. The model contains different populations of neurons ("pools") which respond to stimulus category and position ("sensory pools"), or to combinations of sensory properties of stimuli ("intermediate pools"), or to responses (left or right; "premotor pools"). The pools are arranged hierarchically, are linked by associative synaptic connections, and use global inhibition, implemented via inhibitory interneurons, to implement competition. The model allows a direct comparison with the neurophysiological data but it is not tested within an embodied system. This is one of the most sophisticated models on the complexity and importance of PFC in performing action selection on the basis of complex contexts and behavioural rules. The organisation of the functions played by the ventral pathway of TRoPICALS is similar to that of the PFC functions of this model, but it has a larger computational power due to the use of Kohonen networks.

To understand how PFC is involved in biased competitions and action selection mechanisms Hazy, Frank, and O'Reilly (2007) have proposed a strongly bio-constrained model (named "PBWM" – prefrontal cortex, basal ganglia, working memory) which aims at giving an account of the strict relationship existing between the BG and PFC. The BG modulates PFC representations, in particular to activate information in working memory. This allows PFC to develop more abstract representations. The algorithms used in the model are those of Leabra (see above) and a "k-winners-take-all" algorithm (kWTA). The model is relevant as it highlights that PFC actually does not work in isolation but forms a whole system with the BG. Indeed, this is true not only for PFC but also for PMC itself as both form closely integrated parallel loops with distinct regions of the BG (Kandel, Schwartz, & Jessel, 1998). These loops seem to play a fundamental role in the selection of actions at various levels of abstraction or, as illustrated in the model, in supporting the selection of the contents of working memory. In order to extend TRoPICALS to account for some compatibility effects it might be useful to represent BG in an explicit fashion as they can finely regulate various aspects of PFC bias or can endow it with working memory properties (McNab & Klingberg, 2008).

Recently, Botvinick, Niv, and Barto (2009; see also Botvinick, 2008) have proposed a computational model of the hierarchical organisation of action within the framework of hierarchical reinforcement learning (HRL; Baldassarre, 2002; Barto & Mahadevan, 2003). The model proposes an hypothesis for which the dorsolateral prefrontal cortex (DLPFC) could be involved in action selection mechanisms at an abstract level. The authors propose a possible mapping between the components and functions implemented by the HRL model and specific anatomical components and functions of the brain. In particular, they suggest a functional correspondence between the "actor" components of the model, on one side, and the DLPFC (this forms abstract representations of actions) working in synergy with the dorsolateral striatum portion of the BG ("DS": this forms detailed representations of actions), on the other side. Moreover, they also suggest a functional correspondence between the "critic" components of the model, on one side, and the ventral striatum portion of the BG (VS) working in synergy with the mesolimbic dopaminergic system (involved in the production of learning signals), on the other side. Representations within PFC correspond to "option identifiers" in HRL ("options" allow a hierarchical organisation of actions at various levels of abstraction, from those with abstract goals, e.g. "preparing coffee", to those with more specific goals, e.g. "adding sugar") whereas DS implements the details of the selected stimulus-responses corresponding to option-specific action policies of the model. These mechanisms give an account of the role of PFC in representing actions at multiple nested levels of temporal abstraction and show how the prefrontal representations do not directly implement policies but instead select among stimulus-response pathways implemented downstream. This model is particularly important as it presents a biologically plausible way of implementing the trial-and-error acquisition of hierarchical actions, a feature which might be relevant for extending TRoPICALS in future work. Note that the capacity of implementing trial-and-error learning of action is a second fundamental function of the BG aside the aforementioned putative role in the selection of working memory contents. One or both functions might be implemented in TRoPICALS if accounting for some compatibility effects will require simulating in a more detailed way the acquisition of more complex grasping and reaching actions or the functioning of working memory processes.

Polk and colleagues (Polk, Simen, Lewis, & Freedman, 2002) have developed a model to simulate the Tower of London task ("TOL"). This task requires human participants to move three coloured disks on three rods until they match a given final goal configuration. The model aims to understand the top-down control exerted on action by internally-generated sub-goals and an externally-provided final goal. The model is based on the combination of a bottom-up mechanism (a purely data-driven production system) and a top-down mechanism (a goal modulated system). The simulations were used to test a specific hypothesis about the role of the DLPFC in the TOL. The hypothesis suggests that this cortical area represents internally-generated sub-goals which bias the competition for execution of legal moves and so allow the system to achieve the final goal. This model is relevant as it shows that the top-down biasing of PFC can be generated by processes which go well beyond simple task-based biasing. For example, this may be seen in planning processes which recall sub-goals in working memory on the basis of previous experience or mental simulations. This capability of projection into the future is another fundamental property of PFC (Miller & Cohen, 2001) which might be relevant for accounting for other compatibility effects.

The model proposed by Tsiotas, Borghi, and Parisi (2005) represents a first attempt to replicate the results of the experiment by Tucker and Ellis (2001) using a computational neural network model (for related neural network models replicating the compatibility effect found by Tucker & Ellis, 1998, see Parisi, Borghi, Di Ferdinando, & Tsiotas, 2005, and Borghi, Di Ferdinando, & Parisi, accepted). The model is rather abstract in that its architecture is not directly inspired by specific neuro-anatomical hypotheses and the arm of the simulated organism has no realistic dynamical properties. The model weights are updated on the basis of a genetic algorithm (Mitchell, 1999): the results obtained with the model are based on the assumption that reaction times of real participants can be reproduced in terms of the number of evolutionary generations required by the model to develop the requested behaviours. This idea is exploited to show that learning to act in congruent conditions requires a lower number of generations compared to incongruent conditions. In incongruent condition the system has to learn to use the top-down action-selection bias related to the task to suppress the bottom-up action-selection signals related to objects.

Emphasised heading of paragraph: Models stressing principle 3: neural dynamic competitions

The dynamic neural field approach used by Erlhagen and Schöner (2002) to simulate reaction times captures two important aspects which are often ignored in the simulations of reaction times based on classical connectionist neural networks: (a) the competitive nature of information processing in human brain; (b) the effects of response metrics on reaction times (e.g., the distance between the directions of two or more alternative pointing movements significantly affects reaction times in tasks where participants have to select one). As the authors show with a number of specific examples, these two characteristics are crucial to account for various phenomena related to RTs collected in experiments where body movements are important. Models based on simple feed-forward neural networks do not have a similar expressive power. This can be seen in the fact that in works using these type of neural networks RTs are often reproduced in terms of output errors (e.g., see Seidenberg & McClelland, 1989; Cohen, Dunbar, & McClelland, 1990). For these reasons, we decided to use dynamic field neural networks in TRoPICALS to reproduce RTs. Notwithstanding these advantages, however, it has to be mentioned that the dynamic neural field approach has two important limits (cf. Pouget & Latham, 2002). First, it uses neural "maps" with a number of dimensions equal to the number of parameters to be represented, and as a consequence the number of neurons of the map increases *exponentially* with such number and so it soon becomes computationally intractable. Second, to the authors' knowledge the dynamic field approach has been mainly used with networks whose connectivity is *hardwired* and is strongly dependent on the *topological relations* between neurons, and so its applicability is limited by these conditions.

Neural competitive mechanisms similar to those of the dynamic field approach have also been used in "Boltzmann Machines" to account for two compatibility phenomena (Stoianov, Umiltà, & Zorzi, 2005): the Simon effect and the SNARC effect. Boltzmann Machines are neural networks formed by stochastic units that learn on the basis of Hebbian rules (Ackley, Hinton, & Sejnowski, 1985). Stoianov, Umiltà, and Zorzi (2005) have used these models to show that both the Simon and the SNARC effect arise at the response-selection level where all input signals converge after having gone through different processing pathways. Although the model is not embodied and is not closely constrained on the basis of neuroanatomical evidence, its use of the competitive stochastic mechanisms of Boltzmann Machines to mimic RTs is surely one of the most interesting alternatives to the dynamic neural field approach.

Plaut, McClelland, Seidenberg, and Patterson (1996) use a recurrent neural network where learning plays a key role to model domains with multiple regularities (e.g., English word reading). The analysis of the solutions found by the system shows how various aspects of language, such as word frequency and spelling-sound consistency, influence naming latencies. In these simulations the authors define RTs in terms of time needed by the system to reach a stable dynamical state. This is another alternative to the use of dynamic field networks and Boltzmann Machines for modelling RTs. Given their variable features (e.g., hardwired versus learned weights; topological versus unconstrained architecture, etc.) but also their common inherently-competitive nature, a closer future comparison of these methods might lead to important insights with respect to modelling decision making processes and RTs.

Erlhagen and Bicho (2006) show, both theoretically and through examples, how dynamic neural fields can be used as a general design tool for building biologically plausible cognitive architectures for controlling autonomous robot. The work shows how the neural field approach can give a notable behavioural flexibility to robots and functions well in continuous and noisy domains. However, it also (implicitly) shows the limits of the approach in terms of dimensionality of the problem domains which can be tackled through it (see above).

Cisek (2007) uses a dynamic neural field model to propose a general "affordance competition hypothesis" for action selection. According to this hypothesis, the brain processes sensory information to simultaneously specify several potential actions that are currently available. These potential actions compete against each other for further processing until a single response is selected. In particular, the model suggests that the dorsal visual system specifies actions that compete under the biasing influences provided by prefrontal regions and the basal ganglia. The work shows how the model can reproduce the activation of frontal motor areas found in the brain of real monkeys engaged in motor decision tasks. In doing so, it shows that dynamic neural field models capture an important property of the frontal cortex activations underlying motor decision making: the coincidence of the locus of the *decision* about which action to execute and the locus of the *setting of the parameters* of the selected action. Notice that this property is also captured by TRoPICALS.

Drawing constraints from anatomical and physiological evidence from research in the motor system of primates, Ognibene, Rega, and Baldassarre (2006) built an embodied model to reproduce and account for behavioural data from monkeys that learn by trial-and-error to solve pointing tasks. To this purpose, the model proposes a way to integrate the dynamic neural field approach with a continuous reinforcement learning algorithm. Indeed, trial-and-error processes, which can be captured with reinforcement learning actor-critic models, are fundamental learning mechanisms for organisms (Barto, 1995; Sutton & Barto, 1998; Baldassarre, 2002; Mannella & Baldassarre, 2007). The model in particular proposes a way in which the dynamic neural field approach used in TRoPICALS could be extended to include trial-and-error learning processes along with the learning processes used here based on motor babbling and Hebb rules. This might be an important step as the latter mechanisms, although biologically plausible, have strong computational limits if not integrated with other mechanisms (Caligiore et al., 2008).

Emphasised heading of paragraph: Models stressing principle 4: language and situated simulation

The computational model proposed by Garagnani, Wennekers, and Pulvermüller (2008) investigates the neural mechanisms underlying results of recent EEG/MEG studies. These studies have revealed that brain responses to the same speech sounds differ if the stimuli are presented in different task contexts (Pulvermüller & Shtyrov, 2006; Pulvermüller, Shtyrov, Kujala, & Näätänen, 2004; Sinai & Pratt, 2002). The authors build a neuroanatomically-grounded neural-network model of the left perisylvian cortex, and use it to simulate early word acquisition processes by means of synaptic-weight adaptation based on a neurobiologically-realistic Hebbian learning rule. The network is able to autonomously form stable distributed neural representations ("cell assemblies") for words and reproduces brain responses to various task contexts measured in real experiments. The model is first to provide a unifying account, at the cortical-circuit level, of different neurophysiological data on language (Pulvermüller & Shtyrov, 2006; Pulvermüller, Shtyrov, Kujala, & Näätänen, 2004; Sinai & Pratt, 2002). In this respect, it suggests different ways in which the linguistic capabilities of TRoPICALS might be expanded.

Mayor and Plunkett (2010) propose a neuro-computational account of early lexical development composed of two self-organising maps respectively performing object and acoustic categorisation. As maps emerge, associative connections between the maps are strengthened on the basis of the co-occurrence of objects and labels, capturing the joint attentional activities between infants and caregivers. Once maps have gained sufficient experience with objects and words, the model successfully generalises labels to novel objects of similar kind by following a single labelling event in a manner that resembles taxonomic responding in infants. Both the models of Garagnani Wennekers, and Pulvermüller (2008) and Mayor and Plunkett (2010) show how various phenomena involving language can be captured on the basis of associative mechanisms similar to those exploited here. These mechanisms are very powerful when they work in certain conditions (e.g., in a variety of contexts or in social environments).

Mirolli and Parisi (2005) describe a neural network model of early language acquisition with an emphasis on how language can lead to the formation of more accurate internal object categories in the child. Language associative processes are captured with two separate networks that are responsible for non-linguistic sensorimotor mappings and for recognizing and repeating linguistic sounds. These networks form Hebbian associations between their internal layers and this captures the associative properties of language which link word representations to sensorimotor representations. The model shows that associative processes similar to those used in TRoPICALS allow words to recall internal representations of objects, but also that objects can recall the corresponding verbal representations and thus produce the corresponding phono-articulatory movements. The model also shows that associative processes such as those used here can cause phenomena which go beyond the simple associations between words and the representations of their referents, for example they can lead to the modification (and possibly improvement) of such representations.

Cangelosi, Hourdakis, and Tikhanoff (2006) trained a neural network in a robot to exploit symbols received in input to reproduce as output a categorical representation of the objects the symbols refer to. The authors propose this system as a model of language acquisition and show how a robot that is guided by it is able to acquire new concepts of actions via linguistic instructions. This model shows that associative processes involving words and categorical representations of objects can be used to transfer the compositionality properties of language to sensorimotor representations. As flexible compositionality is a hallmark of language, the mechanisms proposed by the model might useful to account for some compatibility effects involving language.

Title level 2: TRoPICALS as a general model for explaining compatibility effects

One of the major claims of the paper is that, thanks to the generality of the four general principles it incorporates, TRoPICALS has the potential to account for several different experiments on compatibility effects. Note that this does not imply that the model can account for other experiments *as it is*, but rather that it can do so if suitably developed under various secondary (although important) aspects *by pivoting on the four fundamental principles* on which it is based. The potential of the principles for explaining compatibility effects derives from their *synergic interplay* within TRoPICALS. In this respect, the mechanism of the double neural route of the control of behaviour is an important precondition for explaining how different sources of information can act in *synergy or in contrast*. This dual route allows the biasing effect which PFC exerts on action selection, based on the context and the goals and needs of the participants, to *overwrite low-level automatic tendencies* when needed, for example those that are related to affordances. The dynamic competition mechanism leads affordances (dorsal pathway) and the prefrontal bias (ventral pathway) to cause *different reaction times* in congruent and incongruent conditions due to the accumulation of evidence in favour of same or different actions. Finally, the associative properties of language allow words to activate the related "simulations", and this allows the three previous mechanisms to *extend their effects to word stimuli*.

Although specific modelling work is needed to fully demonstrate the claim on the generality of the model, we now support it by envisaging the possible ways in which TRoPICALS can be extended to account for various other experiments on compatibility effects. The first type of results which can be modelled with TRoPICALS refer to the versions of the experiment with images where the target object is presented together with a second distracter object (see Ellis, Tucker, Symes & Vainio, 2007). This experiment aims to investigate how the speed of participants' responses (e.g., to classify as either "round" or "square", with a precision or power grip, a target 3D object generated on a computer screen) are influenced by the congruence or incongruence of the distracter size with the requested responses. An interesting result of these tests is that the congruency of the distracter with the target actually *slows down* the response to the target instead of favouring it. The authors of the experiment interpret these results on the basis of the involved attentional processes. Indeed, the experiment can be addressed with TRoPICALS by extending it with attentional capabilities which allow it to scan the various components of the images. If modified in this way, the model could account for the results on the basis of the hypothesis, suggested by the authors of the target experiments, for which the suppression of the perceptual processing of the distracter, and/or the suppression of the action automatically elicited by it, would interfere with the execution of the responses requested by the experiment when the stimulus and the distracter evoke similar affordances.

A second type of result which can be tackled using TRoPICALS is related to the *Simon effect* (Simon, 1990; Kornblum, Stevens, Whipple, & Requin, 1999). This refers to the phenomenon for which the spatial location of target stimuli can influence the reaction times of spatially defined responses even if the location of stimuli is irrelevant for the task. The Simon effect and the affordance-based compatibility effects might have some underlying mechanisms in common, in particular when one considers actions with a strong spatial component (e.g., reaching vs. grasping). Indeed, the Simon effect might be addressed with TRoPICALS by exploiting reaching capabilities and an overt (or covert) attention mechanism with effects similar to those exploited here to account for the experiment with object parts (recall that this experiment involved reaching responses). Such attention mechanism would give a spatial bias to action thus favouring or interfering with the responses requested by the experiment depending on their spatial components (cf. Stoianov, Umiltà, & Zorzi, 2005, and Kornblum, Stevens, Whipple, & Requin, 1999). Note that this way of modelling attention mechanisms is in line with an embodied vision of cognition and the "action-based view of attention" for which the mechanisms underlying covert attention processes, eye movements, and arm-hand movements are strongly related (Allport, 1987; Rizzolatti, Riggio, Dascola, & Umiltá, 1987; Neumann, 1990; see Balkenius, 2000, for the principles that might be exploited for building models within an action-based view of attention, and Ognibene, Balkenius, & Baldassarre, 2008, for a specific model based on such principles).

Simon effects have been also obtained with iconic/symbolic stimuli like left/right pointing arrows (Pellicano, Lugli, Baroni, & Nicoletti, 2009) or with biological stimuli like cartoon faces with their eyes gazing to the left or to the right hand side (Ansorge, 2003). These results might be obtained with the extensions of TRoPICALS illustrated above in relation to the Simon effect, in particular a top-down attention mechanism capable of driving attention to specific space locations with respect to the currently foveated object (see Ognibene, Balkenius, & Baldassarre, 2008, for a model which focuses attention and action in certain regions of space on the basis of visual cues).

TRoPICALS also has the potential to account for *correspondence effects* observed when pictures of real objects are presented by manipulating their spatial orientations. In a typical condition (Tucker & Ellis, 1998; see also Glenberg et al., 2009), stimuli are pictures of graspable objects (e.g., a teacup) presented in an upright or inverted position and with a handle oriented to the left or right. Participants are instructed to press a left (or right) key using the two hands if they detect upright objects and a right (or left) key if they detect inverted objects while ignoring the orientation of the handle. The results show that the objects automatically activate reach-to-grasp actions which favour the responses of the hand congruent with the handle orientation (Phillips & Ward, 2002). TRoPICALS might account for these results on the basis of the same extensions and principles mentioned for the Simon effect (reaching capabilities and attention for action), and the addition of a second arm and a "second" dorsal pathway to control it similar to the ones used here. In fact, although in the experiment the objects are presented centrally, the affordances they elicit have spatial components, for example in relation to the direction of the motion of reaching actions directed to the handle positioned either on the left or on the right side of the cup. These spatial components might affect one-hand responses (e.g., they might affect pressing a left/right button) or even the use of the limbs to execute actions (e.g., left/right arm, as in Tucker & Ellis, 1998).

Another compatibility effect which TRoPICALS might explain, which poses interesting but subtle challenges in the area of word semantics related to sensorimotor representations and evaluations, was carried out by Meier and Robertson (2004). These authors performed an experiment where participants were asked to evaluate words presented on a computer screen as either "positive" or "negative". The authors found that evaluations of positive words were faster when words were presented at a top rather than at a bottom position in the screen, whereas evaluations of negative words were faster when words were presented at a bottom rather than at a top position. On this basis, the authors inferred that positive evaluations pre-activate actions directed to higher areas of visual space whereas negative evaluations pre-activate actions directed to lower areas of visual space. In order to address this experiment with TRoPICALS, one might try to extend it with learning processes which allow it to associate positive/negative values of words with the production of certain gestures (e.g. these might be socially learned, e.g. they might be gestures associated with movements towards high/low portions of space to express positive/negative values). The automatic activation of the representations of these gestures might then facilitate or interfere with the execution of the actions requested by the target experiment.

TRoPICALS has also the potential to account for the recent literature related to the *approach/avoidance effect* (see Niedenthal, Barsalou, Winkielman, Krath-Gruber, & Ric, 2005). This effect refers to the fact that positive and negative words automatically trigger approach or avoidance actions (Chen & Bargh, 1999; Markman & Brendl, 2005; van Dantzig, Pecher, & Zwaan, 2008). For example, Chen and Bargh (1999) found that reading positive words evokes movements towards one's own body (attraction), whereas processing negative words activates avoidance movements (repulsion). Recent evidence demonstrated that approach/avoidance actions are defined in terms of their outcomes: positive words trigger actions aimed at reducing the distance with the stimulus, whereas negative words elicit actions that increase such distance (approach vs. avoidance movements; van Dantzig, Pecher, & Zwaan, 2008). In addition, Freina, Baroni, Borghi, & Nicoletti (2009) found that the compatibility effects between the emotional connotation of words (e.g., "spider" or "cake") and the direction of arm movement found by Chen and Bargh (1999) are affected by the hand posture (e.g., faster RTs for triggering movements away from the body when the word "cake" is perceived if the hand is open than if the hand is holding a ball; faster RTs for triggering movements away from the body when the word "spider" is perceived if the hand is holding a ball than if the hand is open). TRoPICALS can potentially account for these experiments as the core mechanisms underlying them might be similar to those underlying the compatibility effects with positive/negative words. In particular, the model should be pre-trained (learning "during life") to produce approach/avoidance movements (van Dantzig et al., 2008) or grasping/throwing-away movements (Freina, Baroni, Borghi, & Nicoletti 2009) in correspondence to positive (e.g., useful) objects and negative (e.g., damaging) objects, and to associate suitable words to the representations of objects and actions so developped.

In summary, the different experiments on compatibility effects illustrated in this section should be explained in an incremental fashion through TRoPICALS by suitably improving some aspects of its architecture and, at the same time, by leveraging on the four principles it incorporates. In the long run, we expect that the effort to account for such a variety of different experiments will force/allow us to identify the common cognitive principles and neural mechanisms underlying them. This is expected to furnish increasingly comprehensive accounts of compatibility effects, to give an important contribution to their theoretical grasp, and to lead to the production of precise testable predictions.

Title level 2: Future development of the model

Figure 17 presents a scheme which summarises the main components and connections, and the corresponding putative brain areas, that TRoPICALS should be equipped with to account for the experiments illustrated in the previous section. Notice how this architecture is an extension of the architecture of the model implemented here and shown in Figure 3. The scheme highlights that one of the major changes of the model involves the addition of the important cognitive functionality related to the control of overt attention. This would involve adding a further dorsal pathway to the current two dedicated to the guidance of grasping and reaching actions. The existence of a dorsal pathway dedicated to the control of overt attention, that is the control of eye movements directed to identify targets for accurate visual processing, is in line with neuroscientific evidence. In monkeys, this pathway could be formed by the lateral intraparietal area (LIP), integrating information on the spatial aspects of objects with the current positioning of the eye and head, and the premotor region of the frontal eye field (FEF), important for the control of voluntary eye movements (Rizzolatti, Luppino, & Matelli, 1998; Colby & Goldberg, 1999). The human homologs of these monkey areas might be the “parietal eye field” region (PEF) and the same FEF (Culham & Kanwisher, 2001; Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002).

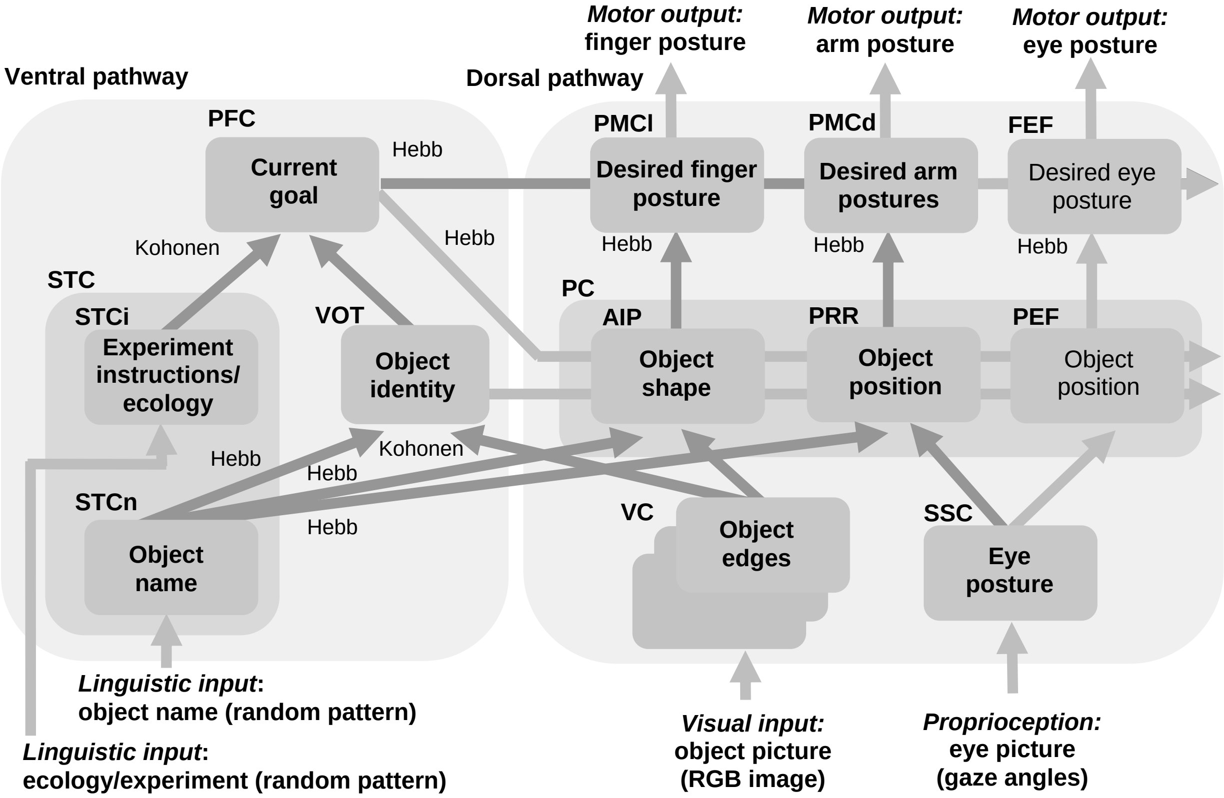

Figure 17: Scheme of brain areas, functionalities, and neural pathways which compose TRoPICALS at the moment (areas with bold names, and connections represented with dark grey arrows) and those which should be added to it (areas with plain names, and internal connections represented with light grey arrows) in order to account for the

compatibility effects investigated with the various experimental paradigms analysed in the text. Arrows indicating the input and output information flows are also reported in light grey.

Based on these components, the attention focus would be guided by both top-down and bottom-up processes. Top down processes would guide attention based on task demands involving working memory processes taking place in PFC and producing effects, via suitable connections, in premotor cortex and parietal cortex (Posner & Petersen, 1990; Fuster, 2001; Hayhoe & Ballard, 2005; Knudsen, 2007). Bottom-up processes would guide attention based on the perceptual saliency of the various regions of the scene and objects (cf. Wheeler & Treisman, 2002, and Itti & Koch, 2000, Knudsen, 2007, on this; see Ognibene, Balkenius, & Baldassarre, 2008, for the proposal of an architecture that integrates bottom-up and top-down attention processes; also see Böhme & Heinke, 2009, for a model in which the attention focus depends on the affordances and contact points suitable for grasping of the objects). Attention might have the effect of enhancing the activation of neural representations of stimuli under the attentional focus (Fischer & Boch, 1981; Culham, Cavanagh, & Kanwisher, 2001; Knudsen, 2007; Gee, Ipata, Gottlieb, Bisley, & Goldberg, 2008) and hence might modulate the RTs due to the increased neural activation fuelling the competition taking place in PMCl and PMCd components of the model.

Note that accounting for some compatibility effects might require a covert attention mechanisms instead of an overt one if attention operates without eye movements (e.g., possibly in Ellis, Tucker, Symes, & Vainio, 2007). In humans, covert attention might rely upon neural systems and mechanisms similar to those exploited by overt attention, as suggested by the “premotor theory of attention” for which covert attention is oriented to a given point when the oculomotor programme for moving the eyes to that point is ready to be executed (Rizzolatti, Riggio, Dascola, & Umiltá, 1987). As a consequence covert attention might be modelled on the basis of mechanisms similar to those mentioned above in relation to overt attention.

Future development of the model might also aim to detail some components of the models which are currently very abstract. The first possible specification regards the current simplified representation of object features within AIP. This component of the model now only encodes the shape of objects but in brain AIP encodes a much richer range of object features together with important motor elements (e.g., hand-manipulation-related neurons in AIP were found to be sensitive to the type of grip, object size and orientation, object 3D appearance, etc., see Murata, Gallese, Luppino, Kaseda, & Sakata, 2000). These issues have been reproduced and investigated in models focussed on the study of affordances and grasping (Fagg & Arbib, 1998; Oztop & Arbib, 2002; Oztop, Bradley & Arbib, 2004). Future developments of the model might develop a richer representation of the object features within AIP and this might help to account for some compatibility effects (e.g., those depending on the hand posture and state, see Freina, Baroni, Borghi, & Nicoletti, 2009) or it might help to find a specific mechanism to reproduce the suitable activations of some model components when the system performs precision or power grips (for example, recall that in many areas of brain precision grips produce a larger activation than power grips: in the model this effect is now reproduced only at a phenomenon level to avoid biasing RTs in wrong ways).

Another part of the model that might be refined in future work is the PRR. This component plays a key role in implementing complex sensorimotor transformations which transform sensorial information to a suitable level of abstraction and to formats expressed in reference frames suitable for the motor control of limbs, body parts, and the orientation of sensors themselves (Colby, 1998; Platt & Glimcher, 1999; Shadmehr & Wise, 2005; Jeannerod, Arbib, Rizzolatti, & Sakata, 1995). The PRR might be refined to incorporate some of these processes if this is required to account for some compatibility effects, for example to model experiments where the duration of action execution is relevant (e.g., see Rubichi, & Pellicano 2004).

There is also a more general issue related to the encoding of motor-related information in AIP and PRR. Currently, the activation patterns in most components of the model (i.e., not only in AIP and the PRR but also in all components of the ventral pathway) are affected by perception but not by action. The reason is that such representations are either hard coded (VC, SSC, STCi, STCn, AIP, PRR), or develop on the basis of non-supervised learning algorithms such as the Kohonen learning rule (VOT, PFC). In real brains such representations are heavily based on action (Jeannerod, Arbib, Rizzolatti, Sakata, 1995; Rizzolatti, Fogassi, & Gallese, 1997; Rizzolatti, Riggio, Dascola, & Umiltá, 1987; Fuster, 2001). This might be due to both the typical bi-directionality of brain cortico-cortical connections or to the learning processes which favour the emergence of action-based internal representations (e.g., supervised learning and trial-and-error learning). Even if this has not been reproduced yet in the current version of the model, its introduction is compatible with the model architecture and will be done in future work.

Another aspect of the model which needs to be developed concerns the locus where PFC exerts a top-down control on action and perception. At the moment this effect is implemented only on the basis of direct connections of PFC with PMC. Contrary to this, in real brains these connections are indirect (they involve supplementary motor areas, SMA, cf. Rizzolatti, Luppino, & Matelli, 1998). Moreover, such an assumption ignores the fact that an important means through which PFC influences motor behaviour is represented by PFC-PC connections (Oztop, Bradley, & Arbib, 2004; Borra et al., 2008; Baumann, Fluet, & Scherberger, 2009). These connections might allow PC to contribute to select actions on the basis of the selective enhancement of some aspects of object affordances represented by it. This is also in line with the literature which ascribes an important role for attention selection mechanisms to the PFC-PC pathway (Knudsen, 2007). In general in relation to the issue of the early versus late control exerted by PFC on the selection of actions and affordances, we are in favour of the proposal of Cisek (2007) and Miller and Cohen (2001) for which PFC contributes to select actions at multiple stages of processing, for example both at the level of the PC and PMC. According to this position, further developments of the model might move part of the PFC influence on affordances/action selection from PMC to PC if this is relevant to account for some compatibility effects (e.g., those involving attention, cf. Ellis, Tucker, Symes, & Vainio, 2007).

Another aspect which is currently not represented in the model concerns the possible connections supporting a “dialogue” between the ventral and the dorsal neural pathways before they converge within the frontal areas. For example, it is known that ITC has anatomical connections with some regions of the PC (Webster, Bachevalier, & Ungerleider, 1994). This aspect might be taken into consideration in future work as it might allow overcoming of some of the limits of the current model, for example the current impossibility of the dorsal pathway to acquire the capacity of responding to objects as requested by the experimental task, and not as suggested by the object affordances, when the two differ.

Similarly, the way the model mimics the associations between representations of words and representations of their referents is now rather abstract. In particular, the model is based on direct connections linking STC to PC whereas these areas might be connected only indirectly via frontal areas (Young, Scannell, Burns, Blakemore, 1994; Pulvermüller, 2005) or via back-connections to early perceptual areas (Kosslyn, 1994). This aspect of the model might be developed in future work to have a higher biological realism. Moreover, as already mentioned in the sections related to the experiments involving language, linguistic processes are represented here only on the basis of very simple linguistic-sensorimotor associations which cannot reproduce several important aspects of language, for example the effects of the formation of linguistic-linguistic associations and the fact that the formation and the nature of linguistic and semantic (sensorimotor) internal representations might strongly influence each other.

Title level 1: Conclusions

This paper presented a model, TRoPICALS, developed to provide a unified interpretation of compatibility effects caused by the automatic activation of neural representation corresponding to object affordances. The core of TRoPICALS is based on four principles: (a) the principle for which actions are triggered on the basis of the operation of two neural pathways, the ventral pathway which processes information used to support overall decisions (in particular at the level of ventro-occipital and temporal cortex and, at a later stage, prefrontal cortex) and the dorsal pathway which processes information used to support the online performance of actions (in particular at the level of parietal and premotor cortex); (b) the principle for which the prefrontal cortex exerts a top-down bias on action selection guided by high-level goals, linguistic instructions, social rules, etc.; (c) the principle for which the selection of one action from the action repertoire is accomplished on the basis of a biased dynamic competition between the neural representations of such actions guided by the accumulation of evidence supporting them; (d) the principle for which the acoustic/visual representations of words are capable of reactivating sensorimotor representations normally activated by their referents and therefore to trigger perceptual and motor processes similar to those directly triggered by them. To the authors' knowledge, and as indicated by an extensive critical review of other models, TRoPICALS is the first model to integrate the four principles and to exploit their *synergic* interactions to furnish a coherent account of compatibility effects.

Thanks to the incorporation of these principles, TRoPICALS was able to replicate and provide a detailed account of the results obtained by Tucker and Ellis (2004) related to the congruency effect exhibited by real participants when dealing with object images. In particular, the dual route organisation of the model architecture allows the action representations automatically activated by object affordances via the dorsal neural pathway to either facilitate or contrast the execution of actions needed to accomplish the experiment and decided by the prefrontal cortex within the ventral pathway. This causes relatively fast or slow reaction times depending on the synergy or competition at the level of the premotor cortex between the two signals.

TRoPICALS also gives an interpretation of compatibility effects obtained with nouns of objects (Tucker & Ellis, 2004; Borghi , Glenberg, & Kaschak, 2004). The key idea underlying this account is that the acoustic/visual representations of words can activate associated sensorimotor representations in the brain ("simulations") generated and activated by the direct interaction of the system with the referents of the words, and therefore they are able to produce similar compatibility effects.

Following a methodological approach named computational embodied neuroscience, TRoPICALS was developed by taking into account four different sources of constraints directed to augment its biological plausibility and generality: (a) *neurobiological constraints*: the architecture of the model was constrained on the basis of neuroscientific knowledge on the cortical areas relevant for the performance of the target experiments; (b) *behavioural constraints*: the model was required to reproduce the results related to the behaviours exhibited by the participants of three experiments; (c) *embodiment constraints*: the model was required to function within a simulated artificial system equipped with a human-like visual system and an arm similar to the participants' ones; (d) *learning constraints*: the model was required to autonomously learn to accomplish the experimental tests and also to acquire skills associated with object affordances (this mimicked the participants' acquisition of skills before undergoing the experiments, an essential step to account for compatibility effects).

As shown on the basis of a theoretical analysis, the fulfilment of these constraints, and the generality of the four principles it incorporates, gives TRoPICALS the potential to explain, if suitable developed, many other experiments on compatibility effects, for example: Simon effects (both with objects' positions and with iconic or biological stimuli), correspondence effects mediated by both objects or language, and approach/avoidance effects. In this respect, the main contribution of this research is the proposition of a model which might be used both as a general theoretical framework and a flexible modelling tool having the potential of progressively producing a unified interpretation of the results of different experiments on affordance-related compatibility effects.

Title level 1: Appendix

This section reports a qualitative analysis of the effects of the variation of some parameters of the model on the results. This analysis varied some key parameters and evaluated the resulting variation of the RTs on the basis of some sampling simulations. The analysis could only assess some effects of the manipulations of the parameters, and so can report only qualitative trends observed during these manipulations, as a systematic collection of data was prevented by the time length of the simulations (given a particular parameter setting, running all the relevant trainings and tests required few hours). Among the various parameters of the model, those mentioned below have the strongest effects on RTs. The results of the analysis can be summarised as follows:

(a) A larger time constant τ regulating the dynamics of the speed of the competitions taking place in PMCl and PMCd (Equation 4) slows down all RTs.

(b) A high decision threshold of such competition causes slow RTs; moreover, it tends to cause more pronounced compatibility effects.

(c) A large width σ of the Gaussian function used to form clusters within PMCl and PMCd (Equation 3) causes clusters to overlap and so RTs become long.

(d) A small value of the lateral inhibition between clusters within PMCl and PMCd (I of Equation 4) tends to cause slow RTs. Too small values of this variable prevent the system from producing “clean” winning neuron clusters.

(e) Maximum values of PFC-PMC connection weights have to be relatively higher than maximum values of PC-PMC connection weights to let PFC to always control action and overwhelm PMC in incompatible cases.

(f) If the differential activation caused by small and large objects in AIP is not diminished (as done in the model), large and small objects produce different RTs; in particular, large objects cause faster RTs than small objects due to the fact that they tend to activated a larger number of neurons in AIP as they have a larger size, and so AIP sends a relatively larger activation to PMC.

Acknowledgements

This research was supported by the EU funded Projects ‘ROSSI – Emergence of communication in RObots through Sensorimotor and Social Interaction’, contract no. FP7-STREP-216125, and ‘IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots’, contract no. FP7-IP - 231722. We thank Kate Burke for helping in revising the English aspects of the paper.

References

Ackley, D., Hinton, G., & Sejnowski, T. (1985). A Learning Algorithm for Boltzmann Machines. Cognitive Science, 9, 147–169.

Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer, & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: Erlbaum.

Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields, *Biological Cybernetics*, *27*, 77–87.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 456–458.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience, 20,* 303–330.

Ansorge, U. (2003). Spatial Simon effects and compatibility effects induced by observed gaze direction. *Visual Cognition*, 10, 363–383.

Arbib, M. A. (1997). From visual affordances in monkey parietal cortex to hippocampo-parietal interactions underlying rat navigation. *Philosophical Transactions of The Royal Society B Biological Sciences, 352*, 1429–1436.

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* (pp. 229–254). London: Springer Verlag.

Arbib, M. A. (2005). *Beyond the Mirror*. Oxford: Oxford University Press.

Arbib, M. A. (Ed.) (2002). *The Handbook of Brain Theory and Neural Networks.* Cambridge, MA: The MIT Press.

Arbib, M. A., Bonaiuto, J. B., Jacobs, S., & Frey, S. (in press). Tool use and the distalization of the end-effector. *Psychological Research – Special issue on "Intentional action: from anticipation to goal-directedness"*.

Balkenius, C. (2000). Attention, habituation and conditioning: Toward a computational model. *Cognitive Science Quarterly*, *1*, 171–204.

Baldassarre, G. (2002). A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviours. *Journal of Cognitive Systems Research, 3,* 5-13.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37,* 407–419.

Barsalou, L.W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Biological Sciences, 358*, 1177-1187.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.). *Symbols, embodiment, and meaning* (pp. 245–284). Oxford, U.K.: Oxford University Press.

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 215–232). Cambridge, MIT Press.

Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems, 13*, 341–379.

Bastian, A., Riehle, A., Erlhagen, W., & Schöner, G. (1998). Prior information preshapes the population representation of movement direction in motor cortex. *NeuroRepo*rt*, 9,* 315–319.

Baumann, M. A., Fluet, M., & Scherberger, H. (2009). Context-specific grasp movement representation in the macaque anterior intraparietal area. *Journal of Neuroscience, 29,* 6436–6448.

Berthier, N. E., Rosenstein, M. T., & Barto, A. G. (2005). Approximate optimal control as a model for motor learning. *Psychological Review, 112*, 329–346.

Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2006). Extending a biologically inspired model of choice: multialternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society B, 362*, 1655–1670.

Böhme, C., & Heinke, D. (2009). Modeling visual affordances: The selective attention for action model (SAAM). In J. Mayor, N. Ruh, & K. Plunkett (Eds.), *Connectionist Models of Behaviour and Cognition II: Proceedings of the 11th Neural Computation and Psychology Workshop* (pp. 325–336). Singapore: World Scientific.

Bonaiuto, J. B., & Arbib, M.A. (in press). Augmented competitive queuing: Flexible motor programs and insight learning.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13*, 185–188.

Borghi, A.M.**,** Cimatti, F.(2009). Words as tools and the problem of abstract words meanings. In N. Taatgen & H. van Rijn (eds.). *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2304-2309). Amsterdam: Cognitive Science Society.

Borghi, A.M.**,** Cimatti, F.(2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia, 48*, 763-773.

Borghi A.M.**,** Di Ferdinando A., & Parisi D. (2002). The role of perception and action in object categorization. In J.A. Bullinaria, & W. Lowe (Eds), *Connectionist Models of Cognition and Perception*. Singapore: World Scientific, pp. 40-50.

Borghi, A. M., Di Ferdinando, A., Parisi, D. (accepted). Objects, spatial compatibility, and affordances: A connectionist study. Cognitive System Research.

Borghi, A. M., 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. M., & Kaschak, M. (2004). Putting words in perspective. *Memory & Cognition, 32*, 863–873.

Borghi, A. M., & Riggio, L. (2009). Sentence comprehension and simulation of object temporary, canonical and stable affordances. *Brain Research, 1253*, 117–128

Borra, E., Belmalih, A., Calzavara, R., Gerbella, M., Murata, A. Rozzi, S., & Luppino, G. (2008). Cortical connections of the macaque anterior intraparietal (AIP) area. *Cerebral Cortex, 18,* 1094–1111.

Botvinick, M. (2008). Hierarchical models of behaviour and prefrontal function. *Trends in Cognitive Sciences, 12,* 201–208.

Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition, 113*, 262-280.

Boulenger, V., Silber, B. Y., Roy, A. C., Paulignan, Y., Jeannerod, M., & Nazir, T. A. (2008). Subliminal display of action words interferes with motor planning: A combined EEG and kinematic study. *Journal of Physiology – Paris, 102*, 130–136.

Brooks, R. A. (1989). How to build complete creatures rather than isolated cognitive simulators. In K. VanLehn (Ed.), *Architectures for Intelligence* (pp. 225–239). Hillsdale, NJ: Lawrence Erlbaum Associates.Bub, D.N., Masson, E.J., & Cree, G.S., (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106*, 27-58.

Bub, D. N., Masson, E. J., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106*, 27–58.

Bullock, D., & Grossberg, S. (1989). Vite and flete: neural modules for trajectory formation and postural control. In W. Hershberger (Ed.), *Volitional Action* (pp. 253–298). Amsterdam: Elsevier.

Caligiore, D., Ferrauto, T., Parisi, D., Accornero, N., Capozza, M., & Baldassare, G. (2008). Using motor babbling and Hebb rules for modeling the development of reaching with obstacles and grasping. In R. Dillmann, C. Maloney, G. Sandini, T. Asfour, G. Cheng, G. Metta, A. Ude (Eds.), *International Conference on Cognitive Systems* (pp. E1–8). Karlsruhe: University of Karlsruhe.

Caligiore, D., Parisi, D., & Baldassarre, G. (2007). Toward an integrated biomimetic model of reaching. In *Proceedings of 6th IEEE International Conference on Development and Learning* (pp. 241–246). London: Imperial College.

Cangelosi, A., Hourdakis, E., & Tikhanoff, V. (2006). Language Acquisition and Symbols Grounding Transfer with Neural Networks and Cognitive Robots. In *Proceedings of IEEE Conference on Computational Intelligence* (pp. 1576–1582). Vancouver.

Castiello, U. (1999). Mechanisms of selection for the control of hand action. *Trends in Cognitive Sciences, 3*, 264–271.

Chen, M., & Bargh, J. A. (1999). Consequences of Automatic Evaluation: Immediate Behavioral Predispositions to Approach or Avoid the Stimulus. *Personality and Social Psychology Bullettin, 25*, 215–224.

Churchland, P. S., & Sejnowski, T. J. (1992). *The Computational Brain*. Cambridge, MA: MIT Press.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis, *Philosophical Transactions of The Royal Society B - Biological Science*s, *362*, 1585–1599.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, *45*, 801–814.

Clark, A. (1997). Being there – Putting brain, body and world together again. Cambridge, MA: MIT Press.

Cohen, J., Dunbar, K., & McClelland (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop task. *Psychological Review, 97,* 332–361.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron, 20,* 15-24.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience, 22*, 319–349.

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

Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology, 11,* 157–163.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron, 32*, 737–745.

Daw, N. D.; Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8,* 1704–1711.

Dayan, P., & Abbott, L. F. (2001). Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT Press.

De Jong, R., Liang, C. C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 731–750.

de'Sperati, C., & Stucchi, N. (1997). Recognizing the motion of a graspable object is guided by handedness. *NeuroReport, 8*, 2761–2765.

Decety, J. & Grèzes, J. (2006). The power of simulation : Imagining one's own and other's behaviora. *Brain Research, 1079*, 4-14.

Deco, G., & Rolls, E. T. (2003). Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *European Journal of Neuroscience, 18*, 2374–2390.

DeSouza, J. F., Dukelow, S. P., Gati, J. S., Menon, R. S., Andersen, R. A., & Vilis, T. (2000). Eye position signal modulates a human parietal pointing region during memory-guided movements. *Journal of Neuroscience, 20*, 5835–5840.

Di Ferdinando, A., & Parisi, D. (2004) Internal representations of sensory input reflect the motor output with which organisms respond to the input. In Carsetti, A. (ed.) Seeing and Thinking. Boston, Kluwer.

Ehrsson, H. H., Fagergren, A., Jonsson, T., Westling, G., Johansson, R. S., & Forssberg, H. (2000).

Cortical activity in precision- versus power-grip tasks: an fMRI study. *Journal of Neurophysiology, 83,* 528–536.

Ellis, R., Tucker, M., Symes, E., & Vainio, L. (2007). Does selecting one visual object from several require inhibition of the actions associated with nonselected objects? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 670–691.

Erlhagen, W. & Bicho, E. (2006). The dynamic neural field approach to cognitive robotics. *Journal of Neural Engineering, 3,* 36–5.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological review, 109*, 545–572.

Fagg, A. H., & Arbib, M. A., (1998). Modeling parietal-premotor interaction in primate control of grasping. *Neural Networks, 11*, 1277–1303.

Feldman, A. G. (1986). Once more on the equilibrium-point hypothesis (lambda model) for motor control. *Journal of Motor Behavior, 18*, 17–54.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Fischer, B., & Boch, R. (1981). Enhanced activation of neurons in prelunate cortex before visually guided saccades of trained rhesus monkeys. *Experimental Brain Research, 44*, 129-137.

Flanagan, J., Ostry, D., & Feldman, A. (1993). Control of trajectory modifications in target-directed reaching. *Journal of Motor Behavior, 25*, 140–152.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and temporal cortices during visual categorization. *Journal of Neuroscience, 23*, 5235–5246.

Freina, L., Baroni, G., Borghi, A.M., & Nicoletti, R. (2009). Emotive concept-nouns and motor responses: attraction or repulsion? *Memory and Cognition, 37*, 493-499.

Fuster, J. M. (1997). The prefrontal cortex: anatomy, physiology, and neuropsychology of the frontal lobe. Philadelphia, PA: Lippincott-Raven.

Fuster, J. M. (2001). The prefrontal cortex – an update: time is of the essence. *Neuron, 30*, 319–333.

Gallese, V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Social Neuroscience, 3*, 317–333.

Gallese, V. (2009). Motor abstraction: A neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research, 73*, 486-98.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: the role of the sensorimotor system in reason and language. *Cognitive Neuropsychology, 22*, 455–479.

Gallese, V., & Metzinger, T. (2003). Motor ontology: The representational reality of goals, actions, and selves. *Philosophical Psychology*, *13*, 365–388.

Gallese, V., Craighero, L., Fadiga, L., & Fogassi, L. (1999). Perception through action. *Psyche*, *5*. Retrieved from http://psyche.cs.monash.edu.au/v5/psyche-5-21-gallese.html

Garagnani, M., Wennekers, T., & Pulvermüller, F. (2008). A neuroanatomically-grounded Hebbian learning model of attention-language interactions in the human brain. *European Journal of Neuroscience, 27*, 492–513.

Gee, A. L., Ipata, A. E., Gottlieb, J., Bisley, J. W., & Goldberg, M. E. (2008). Neural enhancement and pre-emptive perception: the genesis of attention and the attentional maintenance of the cortical salience map. *Perception, 37,* 389–400.

Georgopoulos, A. P. (1995). Motor cortex and cognitive processing. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 507–517). Cambridge, MA: MIT Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61–105.

Glenberg, A. M., Becker, R., Klötzer, S., Kolanko, L, Müller , S., & Rinck, M. (2009). Episodic affordances contribute to language comprehension. *Language and Cognition, 1,* 113-135.

Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. In B. Ross (Ed.), *The Psychology of Learning and Motivation*, (pp. 93–126). New York: Academic Press.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high dimensional and embodied theories of meaning. *Journal of Memory and Language, 43*, 379–401.

Glenberg, A.M., Sato, M., Cattaneo, L., Riggio, L., Palumbo,D. & Buccino, G. (2008). Processing Abstract Language Modulates Motor System Activity. *The Quarterly Journal of Experimental Psychology*, 61, 905-919.

Glenberg, A.M., Sato, M., & Cattaneo, L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology,* 18, R290-R291.

Goodale, M. A., & Humphrey, G. K. (1998). The objects of action and perception. *Cognition*, *67* , 181–207.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15*, 20–25.

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *European Journal of Neuroscience, 17*, 2735–2740.

Grill-Spector, K., Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27,* 649–677.

Gupta, A., & Noelle, D. (2007). A dual-pathway neural network model of control relinquishment in motor skill learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*, 188–194.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society B, 362*, 1601–1613.

Hilgetag, C. C., Burns, G. A., O'Neill, M. A., Scannell, J. W., & Young, M. P. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philosophical Transactions of the Royal Society of London - B Biological Sciences, 355,* 91–110.

Hubel, D. H. (1988). *Eye, brain and vision*. New York. Scientific American Books.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40* (10-12), 1489-1506.

Jeannerod, M. (1994). The representing brain: neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17*, 187–246.

Jeannerod, M. (2007). *Motor Cognition: What actions tell the self.* Oxford: Oxford University Press.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformations. *Trends in Neuroscience, 18*, 314–320

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*, 103–109.

Kandel, E. R., Schwartz, J. H., & Jessel, T. M. (2000). Principles of Neural Science. New York, NY: McGraw-Hill.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience, 30*, 57–78.

Kohonen, T. (1997) *Self-Organizing Maps* (Second Edition). Berlin: Springer-Verlag.

Kohonen, T., & Oja, E. (1998). Visual feature analysis by the self-organising maps. Neural *Computing & Applications, 7*, 273–286.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility – A model and taxonomy. *Psychological Review, 97,* 253-270.

Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1. The time course of Stimulus-Stimulus and Stimulus-Response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of*

Experimental Psychology: Human Perception and Performance, 25, 688–714.

Kosslyn, S.M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT 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.

Kuperstein, M. (1988). A neural model of adaptive hand-eye coordination for single postures. *Science, 239,* 1308–1311.

Landauer, T. K., McNamara, D., Dennis, S., & Kintsch, W. (2007). Handbook of Latent Semantic Analysis. Mahwah, NJ: Lawrence Erlbaum Associates.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes, 25,* 259–284.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex, 11,* 287–297.

Lockman, J., Ashmead D.H., & Bushnell, E.W. (1984). The development of anticipatory hand orientation during infancy. *Journal of Experimental Child Psychology, 37,* 176–186.

Logan, G. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603–649.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5*, 552–563.

Lu, C.H., & Proctor, R.W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin and Review, 2*, 174-174.

Luppino, G., & Rizzolatti, G. (2000). The organization of the frontal motor cortex. *Physiology, 15*, 219–224.

Markman, A. B., & Brendl, C. M., (2005). Constraining theories of embodied cognition. *Psychological Science, 16*, 6–10.

Mannella, F., & Baldassarre, G. (2007). A neural-network reinforcement-learning model of domestic chicks that learn to localise the centre of closed arenas. *Philosophical Transactions of the Royal Society B – Biological Sciences, 362,* 383-401.

Mannella, F., Mirolli, M., & Baldassarre, G. (2010). The interplay of Pavlovian and instrumental processes in devaluation experiments: A computational embodied neuroscience model tested with a simulated rat. In Tosh, C.R. and Ruxton, G.D. (eds.), *Modelling perception using artificial neural networks.* Cambridge: Cambridge University Press.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology,* 58, 25–45.

Matelli, M., Luppino, G., Murata, A., & Sakata, H. (1994). Independent anatomical circuits for reaching and grasping linking the inferior parietal sulcus andinferior area 6 in macaque monkey. *Society for Neuroscience* Abstr 20: 404.4

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review, 117*, 1–31.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience, 11*, 103–107.

Meier, B. P., & Robertson, M. D. (2004). Why the sunny side is up: association between affect and vertical position. *Psychological Science, 15*, 243–247.

Meyer, J.-A., & Wilson, S. W. (Eds.) (1991). In *From Animals to Animats 1: Proceedings of the First International Conference on Simulation of Adaptive Behaviour*. MIT Press, Cambridge, MA.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of premotor cortex function, *Annual*

Review of Neuroscience, 24, 167–202.

Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002). The prefrontal cortex: categories, concepts, and cognition. *Philosophical Transactions of The Royal Society B Biological Sciences, 357*, 1123–1136.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia, 46*, 774–785.

Milner, D. A., & Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.

Mirolli, M., & Parisi, D. (2005). Language as an aid to categorization: A neural network model of early language acquisition. In A. Cangelosi, G. Bugmann, R. Borisyuk (Eds.), *Proceedings of the 9th Neural Computation and Psychology Workshop - Modelling Language, Cognition and Action* (pp. 97–106). Singapore: World Scientific.

Mitchell, M. (1999). *An Introduction to Genetic Algorithms.* Cambridge, MA: MIT Press.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology, 83*, 2580–2601.

Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research, 52*, 207–215.

Newell, A. (1973). You can't play 20 questions with nature and win. In W.G. Chase (Ed.), *Visual information processing* (pp. 135–183). New York: Academic Press.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krath-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*, 184–211.

O'Reilly, R. C. (1996). Biologically Plausible Error-driven Learning using Local Activation Differences: The Generalized Recirculation Algorithm. *Neural Computation, 8,* 895–938.

Ognibene D., Rega A., & Baldassarre G. (2006). A model of reaching integrating continuous reinforcement learning, accumulator models, and direct inverse modelling. In S. Nolfi, G. Baldassarre, R. Calabretta, J. Hallam, D. Marocc, J-A. Meyer, O. Miglino, & D. Parisi (Eds.), *From Animals to Animats 9: Proceedings of the Ninth International Conference on the Simulation of Adaptive Behavior*, (pp. 381–393). Berlin: Springer Verlag.

Ognibene, D., Balkenius, C., & Baldassarre, G. (2008). Integrating epistemic action (active vision) and pragmatic action (reaching): a neural architecture for camera-arm robots. In M. Asada, J.C.T. Hallam, J.-A. Meyer, J. Tani (Eds.), *From Animals to Animats 10: Proceedings of the Tenth International Conference on the Simulation of Adaptive Behavior* (pp. 220–229). Berlin: Springer Verlag.

Oliver, R. T., & Thompson-Schill, S. L. (2003). Dorsal stream activation during retrieval of object size and shape. *Cognitive, Affective, and Behavioral Neuroscience, 3*, 309–322.

Oliver, R. T., Geiger, E. J., Lewandowski, B. C., & Thompson-Schill, S. L. (2005). Involvement of the right inferior parietal lobule in shape retrieval is modulated by prior tactile experience with objects. *Journal of Vision, 5*, 610.

Oztop E., & Arbib M.A. (2002) Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics, 87*, 116–140.

Oztop, E., Bradley, N. S., & Arbib, M. A. (2004). Infant grasp learning: a computational model. *Experimental Brain Research, 158,* 480–503.

Paivio, A. (1986). *Mental representations: a dual coding approach.* New York: Oxford University Press.

Paivio, A.. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

Paletta, L., Fritz, G., Kintzler, F., Irran, J., & Dorffner, G. (2007). Learning to perceive affordances in a framework of developmental embodied cognition. In *Proceedings of 6th IEEE International Conference on Development and Learning* (pp. 110–115). London: Imperial College.

Parisi, D., & Schlesinger, M. (2002). Artificial Life and Piaget. *Cognitive Development,* 17, 1301–

1321.

Parisi, D., Borghi, A.M., Di Ferdinando, A., & Tsiotas, G. (2005). Meaning and motor actions: Behavioral and Artificial Life evidence. Comment to the target-article ‘From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics’, by M.A. Arbib. *Behavioral and Brain Sciences, 28*, 35–36.

Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature, 433,* 873–876.

Pellicano, A., Lugli, L., Baroni, G., & Nicoletti, R. (2009). The Simon effect with conventional signals: A time-course analysis. *Experimental Psychology, 56*, 219-27.

Pellicano, A., Iani, C., Borghi, A.M., Rubichi, S., & Nicoletti, R. (in press). Simon-like and functional affordance effects with tools: The effects of object perceptual discrimination and object action state. *Quarterly Journal of Experimental Psychology.*

Pezzulo, G. (2008). Coordinating with the future: the anticipatory nature of representation. *Minds and Machines, 18,* 179–225.

Phillips, J., & Ward, R. (2002). S–R correspondence effects of irrelevant visual affordance: Time-course and specificity of response activation. *Visual Cognition, 9*, 540–558.

Piaget, J. (1952). *The origins of intelligence in children*. I. U. Press, Ed. New York.

Platt, M., & Glimcher, P. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400*, 233–238.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Polk, T., Simen, P., Lewis, R., & Freedman, E. A. (2002). A computational approach to control in complex cognition. *Cognitive Brain Research, 15,* 71–83

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Posner, M., & Di Girolamo, G. (2000). Cognitive neuroscience: Origins and promise. *Psychological Bulletin, 126*, 873–889.

Posner, M., Pea, R., & Volpe, B. (1982). Cognitive neuroscience: Developments - Toward a science of synthesis. In J. Mehler, E.C.T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities* (pp. 251–276). Hillsdale, NJ: Lawrence Erlbaum.

Pouget, A., & Latham, P. (2002). Population codes. In M. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: The MIT Press.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing and population codes. *Nature Reviews Neuroscience, 1*, 125–132.

Prescott, T. J., Montes-Gonzalez, F., Gurney, K., Humphries, M. D., & Redgrave, P. (2006), A robot model of the basal ganglia: Behavior and intrinsic processing. *Neural Networks, 19,* 31–61.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature, 79*, 49–71.

Pulvermüller, F., & Shtyrov, Y. (2006) Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology, 79*, 49–71.

Pulvermüller, F., Shtyrov, Y., Kujala, T., & Näätänen, R. (2004) Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology, 41,* 106–12.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience, 89,* 1009–1023.

Ritter, H. J., Martinetz, T. M., & Schulten, K. J. (1992). *Neural Computation and Self-Organising Maps: An Introduction*, Reading, MA Addison-Wesley.

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience, 27,* 169-192.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron, 31*, 889–901.

Rizzolatti, G., Fogassi, L., & Gallese, V. (1997). Parietal cortex: from sight to action, *Current Opinion in Neurobiology, 7*, 562–567.

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control, 106*, 283–296.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia, 25*, 31–40.

Rodman, H. R. (1994). Development of inferior temporal cortex in the monkey, *Cerebral Cortex, 4*, 484–498

Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision.* Oxford: Oxford University Press.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function.* Oxford: Oxford University Press.

Rubichi, S., & Pellicano, A. (2004). Does the Simon effect affect movement execution? *European Journal of Cognitive Psychology, 16,* 825–840.

Roelfsema, P. R.; Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*, 376–381.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience, 2*, 1131–1136.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., & Tanaka, Y. (1997). The TINS Lecture. The parietal association cortex in depth perception and visual control of hand action. *Trends in Neuroscience, 20*, 350–357.

Sakata, H., Taira, M., Murata, A., & Mine, S. (1995). Neural Mchanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cerebral Cortex, 5*, 429–438.

Salmon, D. P., & Butters, N. (1995). Neurobiology of skill and habit learning. *Current Opinion in Neurobiollgy, 5*, 184–190.

Sciavicco, L., & Siciliano, B. (1996). *Modeling and Control of Robot Manipulators*, McGraw-Hill.

Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex*, *16*, 1546–1555.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology, 4*, 303–321.

Shadmehr, R., & Wise, S. P. (2005). *The Computational Neurobiology of Reaching and Pointing*. Cambridge, MA: The MIT Press.

Shima, K., Isoda, M., Mushiake, H., & Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex, *Nature, 445*, 315–318.

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology, 102*, 106–119.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor, & T. G. Reeve (Eds.). *Stimulus–response compatibility: An integrated perspective* (pp. 31–86). Amsterdam, North-Holland: Elsevier.

Simon, O., Mangin, J. F., Cohen, L., Bihan, D. L., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe.. *Neuron, 33*, 475–487.

Sinai, A., & Pratt, H. (2002) Electrophysiological evidence for priming in response to words and pseudowords in first and second language. *Brain and Language, 80*, 240–52.

Sobel, I., & Feldman, G. (1968). A 3x3 isotropic gradient operator for image processing, *Presentation for Stanford Artificial Project.*

Sparks, D. L., & Groh, J. M. (1995). The superior colliculus: A window for viewing issues in integrative neuroscience. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 565–584). Cambridge, MA: MIT Press.

Stoianov, I., Umiltà, C., & Zorzi, M. (2005). Computational investigations of the Simon and the SNARC effects. In M. Bucciarelli, B. G. Bara & L. Barsalou (Eds.), *Proceedings of the Twenty Seventh Annual Conference of the Cognitive Science Society*, (pp. 2080–2085). Mahwah, N. J.: Lawrence Erlbaum Associates.

Sutton, R. S., & Barto, A. (1998). Reinforcement Learning: An Introduction. Cambridge, MA: The MIT Press.

Tanaka, K. (1996). Inferotemporal cortex and object vision, *Annual Review of Neuroscience, 19*, 109–140.

Tipper, S. P., Howard, L. A., & Houghton, G. (2000). Behavioural consequences of selection from neural population codes. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and Performance XVIII* (pp. 223-246). Boston, MA: MIT Press.

Toth, J., Levine, B., Stuss, D., Oh, A., Winocur, G., & Meiran, N. (1995). Dissociation of processes underlying spatial S-R compatibility: Evidence of the independent influence of what and where. *Conscious Cognition, 4,* 483–501.

Ts'o, D., Gilbert, C. D., & Wiesel, T. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience, 6,* 1160–1170.

Tsiotas, G., Borghi, A. M., & Parisi, D. (2005). Objects and affordances: An Artificial Life simulation. In B. Bara, L. Barsalou, B. Bucciarelli (Eds.), *COGSCI2005. XXVII Annual Conference of the Cognitive Science Society* (pp. 2212–2217). Mahwah, N. J.: Lawrence Erlbaum Associates.

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*, 830–846.

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

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

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.). *Analysis of visual behavior,* (pp. 549–586). Cambridge, MA: MIT Press.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychological Review*, *108,* 550–592.

van Dantzig, S., Pecher, D., & Zwaan, R. A. (2008). Approach and avoidance as action effect. *The Quarterly Journal of Experimental Psychology, 61*, 1298–1306.

Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B., Bakircioglu, M., & Miller, M. I. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Research, 41,* 1359–1378.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: MIT Press.

Vinberg, J. & Grill-Spector, K. (2008). Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *Journal of Neurophysiology, 99*, 1380–1393.

von Hofsten, C. (1982). Eye-hand coordination in newborns. *Developmental Psychology, 18*, 450–461.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*, 953–956.

Webb, B. (1995). Using robots to model animals: a cricket test. *Robotics and Autonomous Systems, 16*, 117–134.

Webb, B. (2009). Animals versus animats: or why not model the real iguana? *Adaptive Behavior,* *17,* 269–286.

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex, 4,* 470–483.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal Experimental Psychology: General, 131*, 48–64.

Wise, S. P., Boussaoud, D., Johnson, P. B., & Caminiti, R. (1997). Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. *Annual Review of Neuroscience, 20,* 25–42.

Young, M. P., Scannell, J. W., Burns, G. A., & Blakemore, C. (1994). Analysis of connectivity: neural systems in the cerebral cortex. *Review in the Neurosciences, 5*, 227–250.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, *10,* 403–430.

Zorzi, M., & Umiltà, C. (1995). A computational model of the Simon effect. *Psychological Research, 58,* 193–205.

Zwaan, R.A., & Yaxley, R.H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10,* 954-958.