Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENCE, COGNITION AND TECHNOLOGY

Ciclo XXVI

Settore Concorsuale di afferenza: 11/E1

Settore Scientifico disciplinare: M-PSI/01

COGNITION IN CONTEXT: EVIDENCE ON AFFORDANCES AND VERBAL LANGUAGE

Presentata da: Andrea Flumini

Coordinatore Dottorato	**Relatore**
Prof. Giuliano Pancaldi	**Prof.ssa Anna M. Borghi**

Esame finale anno 2014

To my family
with all my love

INDEX

PART I - Theoretical framework and literature review

1. Embodied and grounded view of cognition

2. The affordance effect

3. Language and embodiment

PART II – Experimental work

4. Affordance-based compatibility effects in different contexts

5. Semantic categorization of conflict objects in different contexts

6. Activation of affordances by visual objects and their names

7. Simulating concrete and abstract words acquisition

8. The activation of the left-right mental timeline is context-dependent

9. Evidence of sound-symbolism with every-day objects figures

INTRODUCTION

In the following pages, I would like to give an overview of my doctoral thesis, presenting briefly the issues treated in each chapter. The first part of the thesis is composed of three chapters of theoretical and empirical introduction, with an extended review of the literature relevant for my experimental investigation. The second part is composed by seven chapters in which is reported the original research that I performed during the past three years. This research is aimed at demonstrating the high flexibility and contextuality of embodied and situated simulation processes.

Chapter 1 outlines at a very general level the theoretical context in which my empirical work on human cognition has been developed. First of all, I give a brief outline of the propositional theory (e.g., Fodor, 1975; Marr, 1982), which has been the dominant view in cognitive science in the last decades. Then, the contrasting perspective brought in cognitive science by embodied and grounded cognition theories is outlined. In this view, the meaningfulness of our psychological representations depends on the ability of the human body for sensing the world and acting in it (e.g., Barsalou, 1999; Gallese & Lakoff, 2005; Gallese, 2008; Gibbs, 2006; Glenberg, 1997; Lakoff & Johnson, 1999; Zwaan, 2004). As this approach strongly emphasizes the relation between knowledge and action, the reciprocity of the subjects and the environment where they are embedded is assumed as the basic structure that sustain the very same possibility of any human behavior, including cognition: this mutual relation is what embodiment calls *situatedness* (e.g., Barsalou, 2009). According to the embodied view, activations and re-activations by situated simulation in the brain systems are considered the basis of any cognitive process (Shapiro, 2010). Mental simulation is always characterized by situatedness itself, as it puts again the subject with his/her first-person perspective in the same context in which those experience were initially produced. After discussing in detail the notion of situated simulation, the chapter is concluded with a brief theoretical and empirical overview on the process thought to sustain situated simulations at the neural level. This brain mechanism is called motor resonance, and it is discussed reviewing briefly some relevant findings in behavioural and neuroscientific literature in respect to object, action and language processing.

Chapter 1 sustain that simulation is what underlies cognition and concepts, presenting several relevant findings in the literature in support of this assumption. Among this evidence, there is the Affordance effect. Chapter 2 provides an extensive theoretical and empirical introduction to this issue, as it is the first experimental topic of this thesis. Chapter 2 is opened with the discussion on the notion originally proposed by Gibson (1979), which defined affordances as properties in the environment relevant for the goals of an organism that acts in it. Then, different contemporary theorizations are presented starting from those proposals that, explicitly rejecting the notion of mental representation, described affordances as dispositional properties of the environment that need to be coupled with some dispositional properties of the acting organisms (e.g., Michaels, 1993), or as relations between the features of a situation and the capabilities of a subject (e.g., Chemero, 2003). The discussion follows focusing on alternative perspectives, that focused on the relations between affordances and conceptual knowledge (e.g., Borghi, 2005). In this view, the human ability to interact appropriately with external objects is tied to the capacity of categorizing them, storing conceptual information about them, and linking their concepts to names (e.g., Borghi, 2005). Also the notion of micro-affordances (Ellis & Tucker, 2000) as potential elements of specific actions is described. Intended as brain representations of action possibilities, micro-affordances are the outcome of the linkage between action-related responses and previously stored sensory-motor experiences. Along this line, a last view related to common coding approaches is examined; it proposes that an affordance is realized by allowing an automatic translation of perceptual object features to action that occurs in overlap with the prediction of the action effect features (Haazebroek et al., 2011). Thus, representational perspectives on affordances describe them as relations between situational features, individual abilities and subjective intentions. This relational nature is confirmed by a number of empirical findings (obtained with behavioral techniques as well as neurophysiological, neuroimaging, etc.) that are then presented along the chapter (e.g., Binkofski et al., 1999; Borghi et al., 2007; Chao & Martin, 2000; Craighero et al., 1999; Grafton et al., 1997; Grèzes & Decety, 2001, 2002; Rumiati & Humphreys, 1998; Pellicano et al., 2010; Phillips & Ward, 2002; Tipper et al., 2006; Tucker and Ellis, 2001). This research indicates that at the behavioural level the perception of manipulable objects leads to the activation of appropriate representations based on sensory-motor simulations. The neuroscientific evidence confirms that objects are represented in the brain as potential action patterns by the neural mechanism of motor resonance, indicating canonical neurons as the substrate that sustains affordances (e.g., Gallese & Sinigaglia, 2011; Martin, 2007; Rizzolatti et al., 1998). Subsequently, as objects, and in particular tools, can be endowed with multiple - and sometimes conflicting - affordances, the proposals to distinguish between manipulative and functional affordances (e.g., Bub et al., 2008) as well as between stable and variable affordances (Borghi & Riggio, 2009) are examined in depth. This discussion shows that the simulation elicited by the very same object varies across different contexts or under different action intentions (e.g., Borghi et al., 2012; Kalénine et al., 2013; Natraj et al., 2013), even if with single objects there seems to be an advantage for manipulative affordances activation in perceptual and motor tasks (e.g., Jax & Buxbaum, 2010). The chapter is closed extending the discussion to the motor resonance triggered by the observation of actions performed by other agents (e.g., Calvo-Merino et al., 2005, 2006). In this final section several findings are presented to suggest that our brain is able to give meaning to others actions by a motor simulation modulated by the similarity between the observed actions and the ones we are able to perform, being the resonance tuned to one's own motor skills and repertoire (e.g., Aglioti et al., 2008; Perani et al., 2001; Pezzulo et al., 2010; Tai et al., 2004).

On the whole, chapter 2 overview shows that the emergence of affordance-based compatibility effects is influenced by a number of action-relevant objectual features, as by action intentions and tasks goals. Thus, this theoretical and empirical introduction suggest a flexible and situational nature of the affordance effects, as the results indicate that sensory-motor simulations selectively integrate different properties in dependence of the current context.

Chapter 3 provides the theoretical and empirical introduction to the embodied and grounded cognition view applied to language. Indeed, verbal language is the second experimental topic of this thesis. Embodiment, following the neural exploitation (or reuse) principle (e.g., Anderson, 2010; Gallese, 2008), claims that during language processing is triggered a situated simulation that recruits the same sensory-motor areas active during the interaction with the objects/entities that language refers to. Thus, in order to comprehend words or sentences, an internal re-enactment of the state of the world that is linguistically described is activated by mean of multimodal simulations (e.g., Barsalou, 2008; Zwaan, 2004). In the chapter, sensory-motor representations evoked by language are described in detail starting from a review of behavioural and kinematics evidence on words and sentences processing. The findings in the literature confirm the idea that (perceptual and motor) simulation guide comprehension during the exposure to linguistic materials. This simulation appears to be rather specific, as it depends on different kinds of modality (e.g., Pecher et al., 2004) and on the time of the exposition to the linguistic cues (e.g., Boulenger et al., 2006). Furthermore, linguistic triggered simulation appear to be sensitive to a number of objectual properties presented in the real world by their referents, like shape (e.g., Tucker & Ellis, 2004), size (e.g., Gentilucci et al., 2000; Glover et al., 2004), orientation (e.g., Stanfield &

Zwaan, 2001) and also color (e.g., Connell & Lynott, 2009). The simulation is sensitive to different perspectives on the linguistic described scenes (e.g., Borghi et al., 2004), to motor characteristics of the described events (e.g., Zwaan et al., 2004), and reproduces the typical location in which referents are experienced in every-day life (e.g, Estes et al., 2008; Šetic´ & Domijan, 2007). Simulation processes also iconically represent the referents' affordances (e.g., Myung et al., 2006) and the spatial configuration of the referents' parts (e.g., Zwaan & Yaxley, 2003; Borghi et al., 2004). Also eye-tracking data support these findings, showing that an oculomotor search of external space accompanies the mental search in internal memory of object-related knowledge (Spivey & Geng, 2001). Finally, the sensory-motor simulation triggered by linguistic contents re-enacts action-relevant features at different fine-grained levels, that can include both action goals as well as effector-specificity (e.g., Glenberg & Kaschak, 2002; Scorolli & Borghi, 2007; Borghi & Scorolli, 2009). Overall, the behavioural evidence reviewed in the first part of chapter 3 clearly indicate that the simulation is tied to linguistic content in a very precise manner, showing a high degree of specificity and flexibility in dependence of the linguistic content and the described context.

The neuroscientific literature on both words and sentences is subsequently reviewed, to confirm that the simulation occurs in a very precise manner. This evidence shows that a resonance mechanism is at work during language processing, and that it is constrained in a fine way by language contents. Indeed, as the hypothesis of embodiment is that language recruits the same (or contiguous) sensory-motor areas as for real action and interaction with objects, it behooves that brain activation should display the same topography for processing language as for processing visual objects, affordances, actions, etc. (e.g., Jirak et al., 2010). A number of studies on action/object-related words (e.g., action verbs, names of tools) are reviewed to confirm this somatotopic organization for the brain's linguistic system (e.g., Aziz-Zadeh et al., 2006; Buccino et al., 2005; Hauk et al., 2004; Pulvermüller, 1999, 2005;

Simmons et al., 2007; Tettamanti et al., 2005; Willems et al., 2010). Along with somatotopy also evidence of the involvement in language processes of the mirror neuron system is reviewed (e.g., Gallese, 2008; Rizzolatti & Craighero, 2004), as in the literature those topics are connected because they both rely on the theoretical principle that language is sustained by the perception-action neural circuits. On the whole, the evidence reviewed support the idea that action/object-related language comprehension elicits a resonance directly connected to somatotopy, and possibly to the mirror neuron system. In general it can be affirmed that somatotopy of language contents can be consistently identified in premotor regions, conclusion that bolsters embodied accounts. The fact that the same cannot be affirmed for motor cortices' involvement is not a problem for embodiment. Indeed, it is possible that this indicate a different role played by motor cortices in the direct execution and observation of actions, in comparison to actions conceptualization and linguistic description (Jirak et al, 2010). This is predicted by the neural reuse principle, according to which language flexibly recruits some characteristics of the motor system to modify and building on them in function of the current context.

The discussion then focuses on a very important point about the relation of language comprehension with sensory-motor processes. Indeed, the largest part of the literature on simulation and resonance reviewed so far in the chapter was aimed at demonstrating the similarity of the simulation triggered by language in respect to the simulation triggered by objects/actions processing. Instead, the idea that I propose is that there might be important differences, with the support of recent approaches and evidence, either reviewed in this chapter (e.g., Borghi & Riggio, 2009; Costantini et al., 2011). Interestingly, these findings indicate that in linguistic tasks functional affordances are more likely to be activated, in contrast with the results collected in non-linguistic settings reported in chapter 2 (e.g., Jax & Buxbaum, 2010). This suggests that linguistic content simulations put the subject in the most frequently experienced situation, which is the use context, activating a motor prototype of the objects/actions linguistically denoted that is mostly based on stable affordances and function-related features. This points out that the simulation triggered by language comprehension differs from the activation during observation and interaction with objects at least in regards of the kind of predictions it produces (Borghi, 2012).

Another fundamental issue afterwards outlined in chapter 3 concerns abstract language. Indeed, the literature on simulation and resonance discussed so far built its view of language grounding mostly relying on investigations about action/object-related language. However, theoretically it should be possible to extend what said so far for concrete concepts/words to abstract ones, as embodiment assumes the existence of a single representational format for both kinds, either grounded in perception-action and introspection systems and characterized by a (multi)modal nature. But, as pointed out by Barsalou (2008, p. 634), "how can theories that focus on modal simulations explain concepts that do not appear modal?". Thus, the chapter follows with the review of the three principal embodied attempts of solving this question, which sums up the symbol grounding problem (Harnad, 1990). The first explication is based on conceptual metaphors (e.g., Lakoff & Johnson, 1980, 1999; Santiago et al., 2011). The second, on the re-enactment of action-features (e.g., Glenberg & Robertson, 2000; Glenberg et al., 2008). The third - and more promising in my point of view - takes into the discussion the possibility that multiple representations underlies concrete and abstract concepts. Among the proposal within this third account, the Words As social Tools theory (WAT) is detailed in particular. WAT suggests that a different modality of acquisition (MoA, i.e., perceptual, linguistic, mixed, see Scorolli et al., 2011) is what lead to the contrasting results obtained for each of the two kinds of words/concepts in the literature. Several findings that support this idea are reviewed, confirming that concrete words/concepts rely (more) on the sensory-motor system, while abstract words/concepts rely (more) on the linguistic system in function of their MoA.

In the final part of chapter 3 language grounding, which have been considered so far exclusively from an ontogenetic point of view, is discussed also at the phylogenetic level. The starting point of the discussion is the basic assumption, common to many embodied perspectives on language evolution, of a gestural origin for contemporary verbal languages (e.g., Arbib, 2005; Corballis, 2002; Flumini, 2011; Gentilucci & Corballis, 2006; Ramachandran & Hubbard, 2001). Human language could have existed in an original, entirely gestural form analogous to contemporary signed languages, with a gradual, phylogenetic blending of vocal components in this kind of linguistic system (Corballis, 2009). As the principal grounding property that actually observable gestural languages, signed languages, exhibit is *iconicity* (e.g., Corballis, 2009; Emmorey, 2002; Pietrandea, 2002; Pizzuto & Volterra, 2000), intended as the spatiotemporal resemblance in the mapping of signs to meanings, the discussion is focused on the analogue of iconicity detected in speech, a process called *sound-symbolism* (e.g., Hinton et al., 1994). Several scholars already suggesting that it played an important role in the evolution of vocal languages. Sound-symbolism is indeed considered as a sort of phylogenetic grounding for verbal language (e.g. Kita, 2008; Kita, Kantartzis, & Imai, 2010; Kovic et al., 2009; Ramachandran & Hubbard, 2001). The hypothesis discussed is that if visual iconicity is the system for agreeing upon referents in gestural languages, sound-symbolism - intended as the natural, non-arbitrary link between the sound and the meaning of a word – might have been this system for speech in its initial pahses. The evidence reviewed in the final part of the chapter supports this idea, showing a reality of sound-symbolism in modern lexicon, at both developmental and cross-linguistic level (e.g., Imai et al., 2005; Imai et al., 2008; Iwasaki, Vinson, & Vigliocco, 2007; Kantartzis et al., 2011; Maurer et al., 2006; Nielsen & Rendall, 2011; Nygaard et al., 2009). This boosts recent hypotheses (e.g., Perniss et al., 2010) that indicate iconicity as playing a decisive role at different levels for both spoken and signed language. Indeed, it is very likely that symbolic resemblance relations might be crucial for more than one fundamental aspect of language. For example, in phylogenesis iconicity may facilitate displacement, that is the ability to refer about things distant in time and space. In ontogenesis, it might provide the mechanism for establishing referentiality, linking linguistic form and meaning to facilitate word learning and following meaning generalization and lexical access. Finally, in language processing iconicity might facilitate grounding of words in sensory-motor and emotional systems, thus determining embodiment. This discussion closes the overview on verbal language. On the whole, chapter 3 indicates that perceptual and motor simulation guide linguistic content comprehension. Simulation processes, that seem to be automatic at the word level, at the sentence level occur when all available information is integrated (e.g., Fischer & Zwaan, 2008). Furthermore, simulation and resonance are associated with linguistic content in a very precise manner, and this shows an high degree of specificity and flexibility in dependence of the current context (even if only implied by language).

Chapter 4 opens the second part of the thesis, that is the experimental section. The behavioural study reported in this chapter (Borghi, Flumini, Natraj, & Wheaton, 2012) has a twofold aim. First, this study is an attempt to extend the evidence on affordance-based compatibility effects, mostly based on objects in isolation, to a more ecological situation, corresponding to complex visual scene including a tool, another object and a hand interacting with them. Second, the effects of different kinds of contexts on the activation of affordances are investigated (see par. 2.5). Indeed the presence of a second object along with a tool allowed the creation of different visual contexts defined by the tool-object relation, as the two items could be either spatially, functionally or not related. Moreover, the active object and the second object were presented either alone, with a hand in potential interaction with the tool, or with a hand in effective interaction with the tool, in both a manipulative or functional posture. In this setting, the study tried to address the issues pertaining the relationship between different affordances (i.e., manipulative/functional) evoked by the same active object displayed in different contexts (i.e, move/use). The results of Experiment 1 revealed a clear action-context compatibility effect: manipulative postures are favored by move contexts and inhibited by use contexts, while the reverse was true for functional postures. This confirms the hypothesis that the emergence of object affordances is selectively modulated by the visual and social context displayed by the pictures. Additionally, as in the first experiment responses were manual key-presses, a second experiment controlled if foot responses could be able to replicate the observed effects. The results of Experiment 2 did not show the compatibility of hand actions and contexts, suggesting that the findings of Experiment 1 are due to an effector-specific motor simulation, rather than to simple associations between contexts and hand-postures. On the whole, this study shows that the emergence of affordance effects is tied to the visual scene in which objects are embedded, thus confirming the hypothesis that affordance activation is flexible guided by the context.

The study reported in chapter 5 further investigates the influence of different contexts on affordance activation (Kalénine, Shapiro, Flumini, Borghi, & Buxbaum, 2013). In this case, the study is focused on a particular class of tools, called conflict objects, that are objects endowed with multiple affordances associated to different actions (see par. 2.5.). For example, a kitchen timer may be clenched with a power grip to move it, while it requires to be pinched with a precision grip on its active part in order to use it. In the present experiment, to test the hypothesis that the simulation during conceptual object processing would follow the guide of the context displayed in naturalistic situations where the objects are embedded, participants were asked to categorize (i.e., natural/artifact) conflict object pictures presented in different complex visual scenes that evoked either move- or use-related actions (i.e., move/use context). Categorization judgments were performed on a manipulandum ad-hoc by executing a move- or use-related action (i.e., clench/pinch). The results showed shorter latencies when the motor response and the visual context were compatible, with the effect being particularly marked for pinch responses in the use-context compared to the move-context. This confirms the action-context compatibility effect observed in chapter 4 with conflict objects too. This evidence demonstrate that the emergence of different affordance-based compatibility effect related to the same object is flexibly biased by properties of the current context that are relevant in selecting appropriate actions.

The study presented in chapter 6 compares the simulation triggered by the processing of visual objects with the simulation triggered by language (Flumini, Barca, Borghi, & Pezzulo, under review). While the largest part of previous research has been aimed at demonstrating their similarity, recent proposals highlighted the possibility of finding also important differences in the simulation related to objects and words (e.g., Borghi & Riggio, 2009; see par 3.4). In order to clarify this issue, a categorization task (i.e. artificial/natural) tested the emergence of affordance-based compatibility effects presenting pictures of manipulable objects in Experiment 1, and the names of those objects in Experiment 2. Participants reported their choice using either a big mouse, requiring a power grip, or a small mouse, requiring a precision grip. The results of Experiment 1 showed a compatibility effect between the grip required by the mouse and the grip elicited by the pictures in function of the depicted objects size. The same effect was not found in Experiment 2 results, however the compatibility of hand postures and size interacted with the category of the target-names (i.e., artificial/natural). This finding is predicted by theories of reuse, which suggest that when language recruits structures and mechanisms characteristic of the sensory-motor system, the simulation triggered modifies and builds on them. On the whole, the hypothesis that visual and linguistic stimuli evoke a simulation that is characterized not only by similarities, but also by differences, is confirmed by this study.

Chapter 7 reports a study aimed at testing the predictions advanced by the Words As social Tools theory (WAT) about the grounding of concrete and abstract words (Borghi, Flumini, Cimatti, Marocco, & Scorolli, 2011). According to WAT, the differences between abstract and concrete words have to be referred to their different modalities of acquisition (MoA, e.g., Wauters et al., 2003; see par. 3.5). The hypothesis that abstract words meaning is more strongly grounded in socio-linguistic experiences, while concrete words meaning rely more on sensory-motor experiences, was tested in four experiments. In order to mimic the acquisition of novel concrete and abstract concepts, participants were submitted to a training in which they manipulated invented objects (i.e., concrete entity) and observed groups of invented objects interacting in novel ways (i.e., abstract relations) on a screen. A subsequent recalling task measured participants learning requiring them to judge whether the two elements presented on the screen belonged to the same category or not. Then, a second training presented each item with a category label, that could be or not accompanied by an extended explanation of the category meaning. The recalling task was then repeated and followed by a category-label verification task, where participants were presented with the previously seen exemplars along with other novel elements and a label and required to decide which of the items corresponded to the label. The results showed that, across the experiments, it was more difficult to deal with abstract than with concrete categories, even when the labels were added. A third task differed across the experiments. In Experiment 1 participants performed a feature production task, aimed at controlling if the properties produced with the novel categories matched the patterns evoked by real existing concrete and abstract words. The result confirmed the typical pattern found in the literature (i.e., concrete words evoked more perceptual properties than abstract ones). In Experiments 2, 3 and 4, the third task consisted of a color verification task with manual/verbal responses (i.e., keyboard/microphone). The results showed that, in function of the different MoA, novel concrete words evoked more manual information, while novel abstract words more verbal information. On the whole, these finding support WAT predictions and confirm the idea of multiple representations approaches that indicate the grounding of abstract words relying mostly on socio-linguistic interactions (e.g., Vigliocco et al., 2009).

Chapter 8 reports a further investigation about the grounding of abstract concepts (Flumini & Santiago, in preparation). The study in this case is focused on the metaphoric mapping of an abstract dimension on a more concrete one (see par. 3.5). In particular the TIME IS SPACE metaphor was investigated, with the aim to clarify the issue of automaticity in the activation of conceptual mappings. In the literature, compelling evidence showed that the TIME IS SPACE metaphor consists in the use of a mental timeline (i.e., a concrete schema) to organize and spatially represent the abstract contents of temporal expressions. Prior investigations aimed at testing the activation of the mental timeline in time-irrelevant task have produced mixed, unclear evidence. In two experiments, we improve the design of typical space-time compatibility effect studies, presenting participants with both a time-relevant and a time-irrelevant task (i.e., temporal reference judgment and lexical decision, respectively) on an identical set of tensed Spanish verbs and pseudo-verbs (i.e., both categories of stimuli were endowed with past and future tense markers). This allowed to perform a trustworthy statistical comparison of the two tasks results. Furthermore, Experiment 1 measured reaction times and accuracy as in previous research, while in Experiment 2 also the kinematics of the mouse trajectories performed by participants to respond were recorded. The principal finding in both experiment indicated that the left-right mental timeline was activated only when the task required explicit temporal discriminations. On the whole, the results stays in favor of a flexible account for the activation of conceptual mappings, what fit well with the evidence reported for affordance-based compatibility effects in the previous chapters.

Chapter 9 reports the last experimental investigation on verbal language, a study focused on sound-symbolism (Flumini, Ranzini, & Borghi, under review). Following the assumption that speech evolved from gestures, several scholars have already suggested that sound-symbolism could have played an important role in the phylogenetic evolution of contemporary vocal languages (e.g., Corballis, 2009). Recent research has also highlighted the possible role of iconicity in language as a mean to ground the linguistic-communicative medium in the sensory-motor characteristics of words referents (Perniss et al., 2010; see par. 3.6). These hypotheses are thought to be supported by a number of findings, among them the evidence of iconic correspondences between the sound of invented words (i.e., strident, sonorant) and information from various modality, especially vision (e.g., Maurer et al., 2006). The empirical literature on sound-to-meaning mappings, however, exhibit some methodological problems. In particular, it is striking that the research has been limited to ad-hoc figures, created especially for such experiments. To assume sound-symbolism as playing a role in natural language evolution and grounding implies the necessity of observing it in the most possible ecological settings. Thus, the hypothesis of the reported study was that sound-shape correspondences would be observed when participants had to choose, between two invented words, the name which better suited an image representing an every-day, common entity. As stimuli reproduced every-day entities, a following hypothesis was that this effect would be modulated by the entity category (i.e., artificial/natural). The results confirmed the "classic" sound-shape correspondence with this more ecological images, showing the effect both in Experiment 1, when participants chose a name for figures of natural objects (e.g. leaf) and artifacts (e.g. fork), and in Experiment 2, when participants chose a name for figures of natural (e.g. animals) and artificial agents (e.g. robots). Furthermore, a modulation of the

category emerged when participants had to name agents. On the whole, these results extend sound-symbolic correspondences to known entities, demonstrating the reality of sound-symbolism and validating the idea that sound-to-meaning correspondences could have played a crucial role in language evolution.

Finally, a general discussion of the six studies with their implications is presented. This discussion is aimed at summarizing the results observed, and at highlighting their original contribution to the hypothesis that embodied simulation processes related to both affordances and verbal language are flexibly tied to the context, as it is defined at the experimental level by the materials used, the kind of mappings evaluated, the training performed, the responses executed and the goals of the task.

Keywords

Cognition, concept, perception, action, embodiment, grounding, situatedness, flexibility, context, experience, embodied simulation, sensory-motor system, motor resonance, symbol, categorization, modal, intention, object, affordance, hand posture, effector, compatibility effect, manipulation, function, artificial, natural, visual, linguistic, social, word, sentence, verbal language, concrete, abstract, modality of acquisition, metaphor, space, time, iconicity, sound-symbolism, shape, sound, gestural, vocal, language evolution.

Notes

All human studies reported in this thesis have been approved by the appropriate ethics committee and have been performed in accordance with the ethical standards of the Declaration of Helsinki (1964).

Some references along the thesis may appear to be redundant, but they were repeated in order to allow an autonomous reading of each chapter.

The studies presented in chapter 4, 5 and 7 are published in international journals. The studies presented in chapter 6 and 9 are actually submitted to the editorial board of two international journals for peer reviewing. The study presented in chapter 8 is actually in preparation and will be submitted as soon as possible to an international journal.

PART I

Theoretical framework and literature review

1. Embodied and grounded view of cognition

In the first chapter I sketch the classic cognitive science view of cognition, then I contrast it with the new perspective brought in by embodied and grounded cognition theories. Describing the embodiment framework, I refer about the notions of concepts, simulation and situatedness, basic elements to understand how cognition works in this account. Finally, I briefly refer about the neural mechanism that might sustain embodied simulation.

1.1. Classic cognitive science view of cognition

For many years, cognitive scientists have described the human brain as an information processing device, receiving sensory input, transforming it through some algorithmic process (that manipulate that input like a software) and producing a meaningful output. This idea, called the Human Information Processing metaphor, is described in many seminal works, for instance in Fodor (1983) and Marr (1982). Classic applications of this framework have resulted in accounts of cognitive processing that are largely serial and modular, with different processes specialized in particular types of information and transformations.

The traditional view of concepts was built by many scientists (e.g., in the area of Cognitivism) on this idea. The mind was considered no more than a processing software endowed with the ability of syntactic manipulation symbols. Concepts would be generated in the human cognitive system by a syntactical combination of symbols which are abstract, arbitrary and amodal (e.g., Collins & Loftus, 1975; Fodor, 1983; Pylyshyn, 1973, 1986). Following this classical perspective, perception and action are low level, fixed processes (Fodor, 1975), peripheral with respect to the higher level processes which are characteristic of the human behavior (the “sandwich model”, Hurley, 2002). This independency pertains not only low and high level properties, also perception and action were described as totally independent (e.g., Sternberg, 1969). Thus, no penetrability of perception is allowed, neither for action nor for cognition (Pylyshyn, 1999): all perceptual process have always to happen in the same, constant manner. So, as the most simple or the more complex categorization process is fixed and identical for each individual, every human being shares the very same world. No situation, no context in which cognition could take place may be influential on its processes, properties and products. Life was totally neglected in this account of human cognition.

In the classical framework, sensory-motor processes were strictly separated from cognitive processes. Perception and action were no more than an input and an output devices: perception, which occurs at each time in the same way, always precedes action. What is perceived is independent from both the movements and the knowledge that produce goal-directed action, as the motor system carries out only executive functions. In this way, concepts and meanings were static psychological elements, that must be posited as autonomous from both the body of the human subject and his environment. Thus, the structure of concepts in the mind would simply consist of lists of properties or statements that are represented in a propositional way (Fodor, 1998). In the cognitive science literature many authors have proposed accounts of the nature of this propositionality, which was thought not only in the form of pure statements, but also as semantic networks (Collins & Quillian, 1969; Collins & Loftus, 1975), propositional networks (Anderson, 1983), scripts (Schank & Abelson, 1977), frames (Minsky, 1975). A transduction process is always necessary in order to transform the information collected by the lower level processes in symbols available and usable for the mind. As cognition serves for knowing, not for acting, the transduction final products would arise as still representations, with no linkage with the human organism properties.

It is clear why such a perspective has been named *AAA view*: concepts and meanings are represented in the mind as abstract symbols arbitrarily linked to the world, with no need of being distinguishable for any modality property. Neither the influence of the bodily existence, nor of the ambient in which such behaviors take place, are considered as aspects that could shape our perception of entities and relations in the external world. When the mind is conceived as the particular (human) software necessary for the manipulation and combination of abstract symbols organized in static linguistic schemes, it has necessarily to be independent from the hard-wired structure in which the software itself is implemented (as any software is). This is why the hardware – our body and brain, with their natural functioning – fall completely out of this description: it is simply an unnecessary part for understanding human cognition.

A consequence of the fact that any knowledge of reality would be organized in a linguistic-type form has been that some scientists, to operationalize the classical view, have proposed mathematical models for the extraction and evaluation of words' meanings, considering it as the preferential way to understand and study concepts. In fact, assuming that knowledge is propositionally organized, the representation of a concept was reduced to the lexical co-occurrences presented by its name in the lexicon: it is the semantic relatedness between each element and the whole system what can define the meaning of a concept. In the last years at least two well-known models have approached the investigation of concepts and meanings in this way: they are the Latent Semantic Analysis (LSA, Landauer & Dumais, 1997) and the Hyperspace Analogue to Language (HAL, Burgess & Lund, 1997). The outcomes of the two statistical models are based on mathematical computations performed over selected corpus of texts in order to determine word-to-word or global co-occurrence values for each of the word/concept examined. In both systems the meaning of a word is a vector, hence a point in a virtual space organized in matrices. It is possible to calculate the distance between words, the value of which represents their degree of semantic relatedness. The values obtained by these mathematical calculation simply index if the words appear in similar or different contexts, but the authors, assuming that each concept/meaning would be always the same for every subject, affirm that an appropriate set of these dimensions may be considered as a unified representation of human knowledge (Landauer & Dumais, 1997).

Even if LSA and HAL have received some demonstrations of their predictive validity (e.g Deerwester et al., 1990; Landauer et al., 1998), it is doubtful if they really can be considered theories of human cognition, as claimed by their authors. Indeed, they propose that the meaning of an abstract symbol (i.e., a word) arises from its relations to other undefined abstract symbols. Well-known philosophical arguments – of the type of the Chinese Room (Searle, 1980) – are enough to reveal that such abstract and arbitrary symbols need to be built on something more than relations to other similar symbols, if somewhere they must became meaningful for a real subject. With a simple thought experiment (from Harnad, 1990) it is possible to show that cognition is not what they assume, introducing at the same time a core question in the embodied theoretic framework. Imagine that a man lands at an airport of a certain country without speaking the local language. He notices an advise printed in that language, which is only a collection of abstract and arbitrary symbols for him. To understand it, he can only use a dictionary of that language (i.e., the relations of LSA and HAL), which obviously – like the advise – consists for him of nothing more than meaningless symbols. He searches for the first word in the dictionary, but he does not understand any of the words in the definition or their meaning. So he goes on looking at the first word found in the definition, but as before he does not know the meaning of the words of that definition, and so on. At the end, for many definition he could have read, he would never figure out the meaning of those words: for all the structural relations he could determine among the arbitrary abstract symbols cannot help him in catching anything meaningful. The point is that he cannot anchor such relations to anything known and understandable to him. It implies that in order to know the meaning of an abstract symbol, as it is each of LSA vectors or any word of a certain language, the symbol need to be *grounded* in something more than other abstract symbols. Only such a grounding could make meanings really accessible for a subject. The issue that here emerges has been called the *Symbol grounding problem* (Harnad, 1990), and the questions it sum up about how human cognition create meaningful representation are enough to discard the classic propositional view of cognition, concepts and meanings. This kind of issues in the last two decades made the embodied perspective on cognition gain support from a wide variety of disciplines, ranging from neuroscience and psychology to philosophy, cognitive linguistics and artificial intelligence (e.g., Barsalou, 1999; Berthoz, 1997; Cangelosi et al., 2007; Clark, 1997, 1999; Prinz, 1997; Lakoff & Johnson, 1999; O'Regan & Noe, 2001; Pulvermüller, 2005; Rizzolatti & Craighero, 2004; Talmy, 2000).

1.2. Embodied and grounded view of cognition and concepts

Embodiment is the theoretical framework ascribable to those perspectives of cognition based on the idea that the particular body in which a mind exist shapes its properties, contents and possibilities, as already proposed for example by Phenomenology (Merleau-Ponty, 1965; for a theoretical discussion see Flumini, 2009) and Gestalt psychology (Kohler, 1947). In this view, the meaningfulness of our psychological representations of the world rests on the capabilities of the human body for sensing it and acting in it. In this direction, concepts are thought to be directly grounded in our bodily experiences (e.g., Barsalou, 1999; Gallese & Lakoff, 2005; Gallese, 2008; Gibbs, 2006; Glenberg, 1997; Lakoff & Johnson, 1999; Zwaan, 2004). In this view conceptualization correspond to the storage in long-term memory of those multi-modal states that arise across the brain's systems during perception, action or introspection (intended as the internal state which could include affect, intentions, metacognition etc.). Concepts can be considered as *simulators* capable of the re-enactment of the neural activation patterns that were active when we have perceived or interacted with certain entities in the world (Barsalou, 2008, 2009). External objects/entities are rebuilt in the mind by way of perceptual symbols combinations, which are analogically (not arbitrarily) connected to their mundane referents (Barsalou, 1999). Indeed, perceptual symbols properties depend on the different type of modal information (visual, auditive, tactile, propioceptive, motoric etc.) they convey, which is directly related to the particular interactions in/with the environment that the subject has performed in everyday experience. In this way, concepts and categories become learning products, capable to trace in the mind the real format of their referents. With the partial re-enactment of the multi-modal states they are grounded in, embodied concepts allow a meaningful representational use in the mental workspace which is called *embodied simulation* (see par. 1.3) and can include also prediction about novel situations (Barsalou, 2009).

The embodied view of concepts shiftes the focus on the contextuality and flexibility of cognition, highlighting that concepts are neither fixed nor static. Indeed, the cognitive system needs to be able to build on experience to acquire new concepts as they become relevant in everyday life, thus a concept can be developed for any salient component of experience which attention has repeatedly selected (Barsalou, 1999). For example, when attention focuses repeatedly on a certain type of object in experience such as cars, a concept is developed for it. Analogously, if attention focuses repeatedly on an action (e.g., driving), a concept is developed as well. Because selective attention is flexible and open ended, a concept could develop for any element repeatedly experienced over time, but even the grounding of the most simple concepts involves a very specific set of experiences depending from the individual (physical) skills (Anderson, 2003). It is why a concept as TABLE would not be the same for a toddler and an adult. The possibilities conveyed by the object (its *affordances*, see chapter 2 for extensive description of this topic) for an adult are of putting things on it, even of sitting on it, while a toddler would probably see other possibilities, as walking or standing on it. Thus, the meaning of a situation depends from the bodily experiences collected in the motor repertoire (see par. 1.4), and fundamental for embodied concepts is the role of development, including the personal history of the subject as whole, for example the social norms present in his culture. Furthermore, the resulting set of available responses depends from the individual goals and how they can apply to the actual situation (all these aspects are discussed in detail in chapter 2). Indeed if our motor representations of external objects takes into account the constraints imposed by the body along with the constraints present in the environment, in order to render our behavior really adaptive it needs to follow the guide of our current purposes too. For example, when is time to paint the walls, in absence of a ladder a table could become a good place to stand on for an adult too.

The characterization of cognition as *grounded* relies on the assumption that human knowledge is the outcome of a close coupling between bodies and environment. For understanding cognition (and all mental activity) is foundational the *context* where it happens, not only the (local) context determined by the mind's dependence from the particular properties of the human body, but also to the (global) context of its being always situated (by the bodily existence) in the physical and social ambient which constitute the external world. In this perspective, action is no more the mere execution of responses that sequentially follows perceptual processes (as said in par. 1.1). The subject, with his goals, is no more a simple observer detached from the world, but he builds his own world by acting in it and on it. It is why both the state of the motor system and the planning of responses can influence the perception of the present situation (e.g., Hommel et al., 2001; Prinz, 1997; see par. 2.2). For example, preparing to grasp an object results in facilitated processing for similar shaped/oriented objects (e.g., Craighero et al., 1999); and also the classic mental rotation task (e.g., Shepard & Metzler, 1971) is executed faster and more accurately, if people concurrently perform a manual rotation in the same direction as the required mental rotation (Wexler, Kosslyn & Berthoz, 1998). Thus, as the grounding of conceptual knowledge is the results of repeated interaction with objects, entities and events in the external ambient, in the embodied view there is permeability and circularity between perception, action and cognition (e.g., Glenberg, 1997; Zwaan, 2004).

1.3. Simulation and situatedness: no context, no cognition

The functions recognized in embodiment to *simulation* are many, as much evidence indicates it as a basic neural computational mechanism (e.g., Barsalou, 2008, 2009; Gallese, 2008; Gibbs, 2006; Jeannerod, 2001; Martin, 2007; Pecher & Zwaan, 2005; Pulvermüller, 2005; Semin & Smith, 2008). What here is intended for simulation is that when a concept is stored, and represents all the instances of a certain category, aspects and properties of its content can be re-enacted by an emdodied simulation. It does not mean that simulation is a representational process that necessarily reaches consciousness. If it does, and some simulated aspects become conscious, they emerge as elements of aware mental imagery (Barsalou, 2009). The whole content of a concept could not become simulated simultaneously, only small subsets can be activated at a time to represent particular instances of the concept in particular scenes and occasions. When a car is simulated, for example, it might be a big suv, a two seater sport car or an electric car, as all the experienced contents for CAR reside implicitly in the car concept/simulator: any subsets can be re-enacted on different occasions, tailored to the specific context which elicited it. As concepts can be developed in semantic memory for objects, events, relations, actions etc., an important issue of embodiment concerns in which way simulations of abstract concepts (those concepts which do not have a direct perceptual referent, e.g., DEMOCRACY) are grounded (see par. 3.5 and chapter 7 and 8). Here it would be enough noticing that it has been proposed that simulators for abstract concepts generally capture complex multi-modal simulations of temporally extended events, with simulation of introspections being central in this process (Barsalou, 1999, 2009). According to this, simulators develop to represent categories of internal experience just as they develop to represent categories of external experience. In support of this account stay empirical studies which confirmed that both concrete and abstract concepts contain extensive amounts of situational information, but that abstract concepts tend to contain a greater amount of information on introspections, individual states and events (e.g., Barsalou & Wiemer-Hastings, 2005; Borghi, 2004; Borghi et al., 2011; Wiemer-Hastings et al., 2001; Wu & Barsalou, 2009).

Simulation is what sustains the representation in the mind of any categories' instance not only in its actual perception, but also in its absence, for example during memory or thought (e.g., Gallese, 2008; Barsalou, 2008). It is also what represent all linguistic contents, from single word to complex propositional structures (e.g., Barsalou, 2009; Glenberg & Gallese, 2012). It can produce inferences or predictions based on the previous experience about a category instance or a specific situation (Barsalou, 2008; Pezzulo & Castelfranchi, 2009). It is also what fulfills and give meaning to abstract concepts (further discussion of this point in par. 3.5). Thus, in the embodied framework simulation is able to implement all the functions of the classic AAA symbolic systems, but its functionality emerges implicitly from modal and associative brain systems, rather than being directly represented in the symbolic structures (Barsalou, 2009).

One of the fundamental aspect of embodied simulation, important to highlight for the aims of this thesis, is its *situatedness* (Barsalou, 2008; 2009). Indeed, concepts are processed out of the typical isolation indicated by AAA theories, as they are always situated in background settings like scenes, or events. When representing a car, for example, people do not represent it in isolation, but in relevant situations. Subjects situate concepts at least because when the brain simulates a perceptual experience while representing a concept instance, it should simulate it in a certain situation because situations are intrinsic in real-world perceptions (Yeh & Barsalou, 2006; Barsalou, 2009). Indeed, perceptual activity depends from the embeddedness of the subject in a physical (and social) context: a human being always perceives the space around the body, which includes objects, events, agents etc.; in everyday life, if we focus the attention only on a particular entity/event in the perceived scene, the background situation does not disappear, but continues to be perceived, because total isolation in the real world is impossible. Thus, as the situation never disappears, when we built a simulation to represent a category instance we tend to simulate it in a relevant perceptual scene: the simulation is always situated because in real life cognition is always situated, as a direct consequence of its correlation to a living body.

In different occasions, the simulation process would give place to different situated representations; conceptualizations that are more specialized are constructed in place of single general representations, as the latter would be indefinite, thus not able to support the relevant inferences which are necessary for adaptive behavior in specific situations. This aspect is also compatible with well-known principles of cognitive load economy (e.g., Zipf, 1949). If we consider for example the representation of a chair, according to the traditional view it would be represented as a generic list of amodal propositions which become simultaneously active every time an instance of the category is processed. According to the embodied view, the cognitive system would produce many different situated representations of the concept CHAIR, each of them based on individual previous experiences, but also anchored to the present context. It means that the grounded concept works in a way that would help the human agent in the interaction with the peculiar chair present in that specific situation. Thus, one situated simulation for chair could support the goal of sitting on it, while others could support moving it, or repairing a broken leg etc. In the embodied and grounded cognition view, a concept is not the single, general and propositional representation of the category, but rather the concept is a subject skill, that is the ability to produce a wide variety of situated representation in order to allow the achievement of the goals in the specific contexts in which the simulation was triggered. To notice here that a multi-modal simulation is something built out of many different concepts activation. For example, the situated simulation of driving a car would include the necessary different sub-simulations for any involved object, action, agent, setting and also internal state (e.g., being afraid of a weat road). Actions, entities and introspections are re-enacted from the conceptualizer particular perspective, so during simulation the subject is put in that situation with the first-person experience of being there (Barsalou, 2009). This means that the situation is no more represented in separation from the knowing subject, because cognition is not possible out of any context.

To summarize what said about simulation so far, our cognitive processes are constrained by the body and the modal sensory systems. Activations and re-activations in these systems by simulation are the basis of cognition (Shapiro, 2010). Simulation identifies objects and agents, sustains thinking about events or places, predicts behavioral consequences, makes situated plans for action, and so on. Thus, according to this view, to know what entity is in front of a subject, he must simulate the previous experiences (visual, auditory, sensory-motor etc.) with it. Additionally, when reasoning about or imagining that entity, similar simulations serve as the cognitive operations' basis: all thoughts on that entity rely in some ways on these modality-specific simulations, because even abstract thoughts rest on the possibility offered by embodied simulation. Finally, as it relies on reactivation of real-life experiences, simulation is always situated, as it puts again the subject with his/her first-person perspective in the same context in which those experience were initially produced.

1.4. Neural substrate of embodied simulation

In this paragraph I briefly describe the neural substrate where situated simulation processes rest. As it is fundamental for some of the results of my investigations on the two empirical topics of this thesis, that are affordances (see chapter 2) and verbal language (see chapter 3), it will be discussed further in the next introductory chapters.

The neural substrate for simulation is offered by the mechanism of *motor resonance* (review in Rizzolatti & Craighero, 2004). It was revealed by well-known neurophysiological studies on the macaque brain, which confirmed the idea that concepts are structures which can reactivate past interactional experiences in the mind. These studies found in the monkey's premotor cortex, especially in the area F5, two kinds of visuomotor neurons, named canonical and mirror neurons (e.g., Di Pellegrino et al., 1992). What was observed during the imaging of this brain areas is that the canonical neurons system is active in the direct execution of actions (e.g., the action of grasping a banana), with that same set of neurons being active during the merely observation of objects (e.g., a banana). The mirror system instead is active during goal-directed action execution (e.g., the action of grasping a banana) and during the observation of actions performed by other agents (e.g., other monkeys or experimenters grasping a banana), while it seems that no neural activity is detected in this system during the observation of objects presented alone (see par. 2.4).

In the last years, much research has been aimed at extending the evidence collected on animal populations to human subjects, and the results suggesting that in the human brain there are cortical circuits endowed with functions similar to monkey's F5 area. Indeed, as predicted by the supposed existence of a canonical neuron system, much evidence showed that our brain is able to resonate to the action afforded by external objects. For example, in an experiment in which participants were asked to mentally manipulate the objects they were presented with, brain activity was shown in the prefrontal, parietal and motor cortex (Grèzes & Decety, 2001). Similar circuits were active during execution of the hand grips evoked by manipulable objects (Grèzes et al., 2003). At the same time, a great number of results collected with fMRI during passive viewing of objects showed an analogous fronto-parietal activity (e.g., Binkofski et al., 1999; Chao & Martin, 2000), which was also found in PET studies (e.g., Grafton et al., 1997).

The motor resonance to others' actions, supposed to be mediated by a human mirror neurons system, has been intensively studied too. For example, a fMRI study (Buccino et al., 2004) investigated functional activations in human subjects during the observation of videos of biting and communicative gestures, performed not only by humans, but also by monkeys and dogs. Independently from the species, viewing biting actions activated in the participants regions (thought to be part) of the human mirror system, while viewing communicative gestures elicited activation of the human mirror systems only for actions performed by humans (lip-reading) and monkeys (lip smacking), not by dogs (barking). In the same direction, reduced mirror activity was found as the consequence of observation of actions performed by a robotic hand, compared to human hand actions (e.g., Perani et al., 2001). These results implies for the activation of the mirror system, and so for the emerging of a motor resonance, a fine-graded structure, which is modulated (at least) by similarities in the body structure of the observer and the agent observed (I further discuss this point in par. 2.6).

Furthermore, another fMRI study (Buccino et al., 2001) found that when a subject observes different type of actions, involving respectively as effector the mouth, the hand or the foot, selected neural areas – the same that map the first-person execution of those actions - are actived in dependence of the effector used for the action. This selective activation of different cortical areas in dependence of the observation of different kinds of action observed revealed a *somatotopic organization* for the human brain. In the literature, this was confirmed also in empirical studies on verbal language (e.g., Pulvermüller, 2005; Hauk et al., 2004). For example, Tettamanti et al. (2005) found differences in the brain activity that were triggered by the exposition to sentences referring to actions performed with different effectors: the mouth, the hand or the leg. The three categories activated selectively the brain regions involved in the real performance of the action they referred to; additionally, activity in Broca's area was detected independently from any effector-specificity, implying a special role in language processing. Similar evidence was observed also with TMS (Buccino et al., 2005) in a study where the recording of motor evoked potentials from hand and foot muscles found a decreasing amplitude during the exposition to sentences respectively describing hand and foot actions. The authors performed a behavioral study too, which extended the TMS results showing that participants responded faster to hand action sentences with a pedal and to foot action sentences with a keyboard. The inhibition effect observed in both neurophysiological and behavioral measures by Buccino et al. (2005) showed a modulation of the motor system activity during the comprehension of action language which is symmetrical to the effect emerged with the observation of actions in video of Buccino et al. (2001). Similar behavioral findings were also observed in a lexical decision task performed on words that referred to mouth and hand actions (Scorolli & Borghi, 2007, I further discuss this topic in par. 3.2 and 3.3).

Thus, the reviewed results on motor resonance support of the simulation hypothesis, as the simple exposition to visual and linguistic stimuli referring to objects/actions implies the activation of the same neural patterns that sustain the real, first-person interaction with objects and performance of that actions. Furthermore, the behavioural evidence clearly indicates the bidirectional link between perception, action (low level process) and language (high level process) which is predicted - as described in par 1.2 and 1.3 - by the embodied and grounded view of cognition.

2. The Affordance Effect

The first topic of my thesis concerns the experimental investigation of the affordance effect in different context. In this chapter I discuss the literature on affordances both at a theoretical and empirical level, discussing first the original notion of affordance by J.J. Gibson, then the differentiation of this notion in contemporary literature. I review important empirical evidence on affordances based on research at both the behavioral and the neural level, describing the functioning of the compatibility paradigm, the same used in my experiments to investigate the affordance effect in different (visual and social) contexts. Finally, I refer about the distinction between manipulative and functional affordances (my empirical investigation on these topics is reported in chapter 4 and 5), which reveals the high contextuality and flexibility of affordances, and about the mechanism of motor resonance, thought to be the substrate which sustains their emergence at a neural level.

2.1. The origins of the notion of affordance

J.J. Gibson was the first psychologist to use, in his works on visual perception, the notion of affordance. His theory, condensed after three decades of research on vision in the well-known volume *The Ecological Approach to Visual Perception* (Gibson, 1979), defined affordances as something that the environment offers to living organisms for tuning their actions to the surrounding physical reality: “The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill” (Gibson, 1979, p. 127). Thus, an affordance can be defined as an *invitation* to act, which presuppose both an action context as an agent. Gibson wrote that “an affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behavior. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer" (Gibson, 1979, p. 128). The concept of affordance has its origins within the Gestalt psychology research. The Gestaltists argued that we perceive the function of a thing as quickly as the color. Gibson, quoting Koffka (1935), said that "each thing says what it is […] a fruit says eat me, water says drink me" (Gibson, 1979, p. 137). It means that we are able to see directly what these objects/entities are for, and how to use them too.

Affordances were described as phenomenal in their nature, and not as physical properties of objects. Indeed an affordance is, rather than an object or a part of it, the outcome of a dialectic process which involves both acting organisms and external ambient. Such an idea reveals that the ecological approach to cognition is a view that strongly emphasizes the relation between knowledge and action, assuming the reciprocity of the perceiving/acting/knowing organism and the surrounding environment as its basic structure. We already indicated in the precedent chapter (par 1.3) that a similar mutual relation, which can be identified as *situatedness*, is a basic idea also for embodied and grounded views of cognition.

The ecological approach was entirely constructed by Gibson on his research on visual perception, which marked an important innovation with respect to the traditional approach reserved to perception by Cognitivism. Here it can be useful to identify the differences between the two theorizations, as the cognitivist theory of perception was based on the traditional perspective on cognition (whose theoretic view was discussed in par 1.1). Cognitivism considered the perceiving subject as an entity detached from the surrounding environment, a pure observer. Visual perception was reduced to the mechanical impression of images in the eye produced by the stimulation of the retinal receptors by light (Merleau-Ponty, 2003). In the ecological perspective of Gibson, this situation simply does not exist, because in real visual perception the external world is not composed of geometrical solids on a plane, as in a painting. Gibson rejected the retinal image principle, and claimed that on the contrary, what sustains our possibility to see is the existence of an *ambient optic array*: what we perceive is the whole array of the refrained light rays that reach us. The idea is that structure existing in the surface of the environment structures in turn the light, and the angles created in the refrained light directly provide the observer with visual information on objects, entities, textures, substances, events in the environment. Thus, it is the *structure* of light, rather than the *stimulation* of light, that furnishes information in visual perception: stimulation *per se* does not lead to visual perception (Bruce, 2003), as revealed by the perceptual experiences experimentally induced in a Ganzfeld (e.g., Avant, 1965; Gibson & Dibble, 1952; Gibson & Waddell, 1952). It is why the ambient optic array has been shortly defined as "how things look from here in these conditions" (Noë, 2004). As a consequence of the ecological approach starting point, in this view perception can be defined as the bottom-up process of direct registration of the *invariants* present in the ambient optic array. The optical flow invariants are obviously what Gibson identified with the term affordance, and this is why he claimed that the right level for understanding visual perception is the one set by ecology: as these invariants are relative to animals, they cannot be understood in the terms of optic or physics, nor reduced to a geometrical description, like the traditional theory of perception tried to do.

Another fundamental point of difference from Cognitivism concerned its assumption that perception is a mediated process. Indeed, ignoring the perceiver natural surroundings, the traditional theory considered the structure of light as not adequate for specifying the properties of the environment, so it invoked the supplement of inference processes, or stored representation etc., in order to explain perception. In the ecological view, perception is a direct process not in the need to be sustained by memory, conceptual inferences or any other high level process: "You do not have to classify and label things in order to perceive what they afford" (Gibson, 1979, p. 134). As said above, an organism can directly perceive objects in the world just picking up the invariants in the optic array, this is enough to specify all the entities and events present in the environment, and no identification is needed to benefit of their invitation to act. The ecological psychology assumed with the direct perception of affordances an ability of the individual to *resonate* to the ambient like a tuning fork, which responds to sound frequencies without doing any sort of information processing. Gibson stated overtly that organisms have evolved to develop the possibility of accessing relevant information about the environment directly from the invariant structures in the array, without any mediation of higher level processes.

A fundamental aspect, that is connected to the latter and was originally highlighted by the ecological theory and later recognized also in embodied theory, is the assumption that perception is an active process. Indeed "invariants of structure do not exist except in relation to variants" (Gibson, 1979, p. 87). The variants are identified both with the movement of the observer and entities in the ambient, where the possibility for the perceiver to actively explore the world is considered a very powerful way for collecting information. When an organism moves the entire optic array is transformed: the transformation contains information about the layout and shape of objects and surfaces, and about the perceiver's same movements in the ambient too. It is evident that here perception itself is a form of action. If traditional cognitive theories described visual perception as the passive detection of retinal sensations, Gibson always underlined the mutuality of perception and action.

A final point important to notice here, as it follows from the active view of perception and it is also fundamental for the embodiment view, is the complementarity of the perceiving organism and the environment. In the traditional theory visual perception was the processing of the retinal reflections of light stimulation without any context. It implied a fixed pathway, that goes from the outside to the inside of the observer. Against the theory of visual impressions, the ecological approach stated that perception of the environment is simultaneously perception of the self: “Perception of the world and of the self go together and only occur over time” (Gibson, 1979, p. 87). As affordances are neither objective properties nor subjective ones, and at the same time they are both, they can cut across the classic dichotomy of subjective-objective. They are the conjoining of subjective and objective, which fulfilling the netherland between the perceiver and the world creates the contact point necessary for adaptive behavior. It is why, as said, invariants make sense only for variants: affordances only take meaning from a specific system’s point of view as anchored to the environmental context. This makes organisms and environment linked and mutually interdependent: not only an organism needs an environment, but also the environment implies organisms acting in it.

To briefly summarize, affordances are invariants in the ambient (optic array) that can be registered directly by our perceptual system, with no more effort that just picking them up. Affordances are holistic in nature, as when we look at the external objects/entities we do not perceive geometric perimeters, dimensions, physical properties etc.: we perceive their affordances. Affordances cannot be considered as something exclusively objective or subjective; reciprocally, they cannot be considered as something purely physical or psychical. Indeed they are equally “a fact of the environment and a fact of behavior. […] An affordance points in both ways, to the environment and to the observer” (Gibson, 1979, p. 129). So affordances are structures whose nature depends both from the specific bodily properties of the acting organism and from its being in a certain specific context (the surrounding environment): affordances are intrinsically relational properties, which exist and operate only in a perception-action loop (Turner, 2005). If the conditions sustaining the emergence of affordances pertain to both the agent and the ambient, or - as better said above - to their complementarity, this assumption is fundamental for the perspective on the contextuality of cognition that guides this thesis. Indeed since affordances depend from both bodies and ambient, as from both perception and action, they could be intended not only as relational, but also as situated and contextual structures, even if it could mark a discontinuity with their original notion.

2.2. Contemporary perspectives on affordances

The notion of affordances have been employed so far in many areas of cognitive science in the last decades. In ergonomy, for example, the notion has been extensively used in well-known exploration of the properties of cognitive artifacts; for example Norman (1988) discussed the roles of affordances and natural mappings for usability; Gaver (1991) applied and extended the notion investigating the design of computer displays; Robertson (1997) used it for studying the situational features that support remote cooperation. Other examples of the use of the concept can be commonly found in the literature on anthropology (e.g., Cole, 1996; Wenger, 1998; Holland et al., 2001), and also in the literature on education (e.g., Hammond, 2010; Roth et al., 1996). In cognitive psychology, the contribution given by Gibson challenged as said many long-accepted ideas, thus at the beginning many authors (e.g., Neisser, 1977) have refused his proposal. However, then started an attempt (e.g., Neisser, 1994, 1995; Shepard, 1984) to bring together the classic approach, which highlights high level functions and top-down processes, with the gibsonian ecological approach, in which central is the information present in the ambient collected in a bottom-up way. Gibson's notion of affordance converged with cognitive representationalism in a view which affirmed that not only perception, but also imagination, thought and even dream are similarly guided by internalizations of long-enduring constraints in the external world (Shepard, 1984). This is a foundational step in the development of a useful notion of affordance for describing the relations between perception, action and cognition in simple and complex behaviours. The contemporary views of affordances that I discuss below can be differentiated exactly for their resort (or not) to conceptual representations.

Nowadays the value of the ecological approach for cognitive psychology has been widely recognized, even if the term affordance is sometimes used in slightly different ways with respect to Gibson notion, in function of specific attempts to unfold what an affordance really is. The largest part of the post-gibsonian attempts have described affordances as dispositional properties in the environment that are in the need to be supplemented by certain dispositional properties pertaining to the individual organism. These approaches explicitly reject the notion of mental representation (e.g., Michaels, 1988, 1993; Michaels et al., 2001; Shaw et al., 1982; Turvey, 1992). A recent, radical perspective has even refused the definition of affordances as dispositional environmental properties, because although they are considered as relative to an agent, they cannot be identified with properties. In this view, they are better described as the relations existing between the features of a situation and the abilities of an individual (Chemero, 2003, 2009). Far from being a minor point, there is an important epistemological reason to mark the difference between properties and features (Costantini & Sinigaglia, 2012). Indeed, in order to perceive a certain property of an object a subject needs first of all to identify the object itself, and then to recognize that such an object really has the aforementioned property. But in the original gibsonian view, as said above, to perceive an affordance a subject does not have to identify the particular entity in question (Gibson, 1979): what is needed is simply the ability to resonate with the ambient and perceive immediately that the specific situation *as a whole* supports - in a certain measure imposes - a certain kind of action. Thus, only the situation as a whole is what can exhibit that certain feature (Chemero, 2009). In the same vein, neither the notion of individual abilities are explained in terms of dispositional properties, as the concept of ability would show a certain degree of normativity: individuals endowed with particular abilities are supposed to have the possibilities of behave in certain specific manners, but without having any guarantee of doing it successfully. On the other hand dispositions are not something that can succeed, nor fail: simply, they are or not in the adequate situation to make their emergence and manifestation possible. Furthermore, individual abilities are not just reciprocally interconnected, but hierarchically organized: all higher abilities are grounded on the more basic ones, which are obviously the primary motor abilities. It implies that if in a certain context the basic motor abilities cannot be exercised, no one of the higher could be (Chemero, 2003). And according to the motor possibilities defined by having a certain body, basic affordances could be distinguished in two main categories: micro-affordances and macro-affordances (Costantini & Sinigaglia, 2012). The notion of micro-affordances (which I discuss further in the following part of this paragraph, to show that actually they have a different connotation in the work in which they were originally proposed) pertains transitive, object-centered interactions: micro-affordances are active if the situation suggest behaviors like grasping, pulling, biting, kicking etc.. Macro-affordances (or walking like affordances) instead indicate motor abilities as locomotion or navigation, thus are elicited when the situation suggests behaviors like walking, jumping, climbing etc.

From the aforementioned perspectives, which explicitly reject the notion of mental representation, seems that if affordances directly support us in interacting with objects, concepts are not in the need to come into play for behaving successfully. Indeed, the possible behaviors are entirely specified by the pattern induced by the environment in the perceiver, or by the features of the situation as imposed by the perception-action loop. It is very likely that such views, even if they are more in line with the original proposal of the ecological approach, could may be applied to simple affordances (i.e., the ones related only to very basic motor skills), not to those affordances that elsewhere have been defined as *complex* (e.g., Turner, 2005). Indeed, scholars that have focused the discussion on the possible relations between affordances and conceptual knowledge (e.g., Borghi, 2005) have already noticed that there are some problems within the above described views, especially when it is taken in consideration more than simple actions in a merely physical world. In this alternative perspective on affordances, our ability to interact appropriately with objects depends also on the capacity of categorizing objects storing conceptual information about them, and on the capacity, fundamental for homo sapiens, to link concepts with names (Borghi, 2005). Concepts, intended as the cognitive and internal aspects of categories (Barsalou et al., 2003; Borghi, 2005), are the mental glue (Murphy, 2002) necessary to connect our previous experiences to the actual interactions with the world. Indeed, if we consider the affordances derived from a slope blocking our way, it is possible to quickly pickup from the situation the affordance of going back, or climbing the obstacle. But if we consider the affordances derived for example from a bicycle, many behaviors are simultaneously afforded: the handles may afford grasping, the seat sitting, etc.; to coordinate action successfully, catching the global and conventionalized affordance of the bike, that is to ride it, we need to access some form of previous knowledge related to experiences with bikes. We need this conceptual information in order to know how to react properly to each affordance. Thus, it is in particular the ability to use tools appropriately what reveals as fundamental the possibility of combining actual affordances with the previous experience of that object (or a similar one) and with any previous experience of its conventional function (Borghi, 2005), even if this experience is not direct: the information regarding object affordances can be acquired socially, as human are able from very early stages of development to extract the meaning of a situation simply observing others, by mean of imitation (e.g., Marschall & Johnson, 2003). In this view, another fundamental element to take in account is the goal that subject has in interacting with the external objects/ambient: the intentions of the actor are in the need to have an effect on action (as said in the previous chapter, see par 1.2 and 1.3) if human behavior would exhibit fitness in adapting to the real demands of the physical and social world. Indeed, if for example the action of grasp and move a blender might be elicited automatically by perceiving it visually, the act of using the blender for beating eggs is something more, resulting not only from the basic motor abilities of the subject, but also from stored knowledge and goal-directed intentions, and the act to wash it after cooking implies further representations and capabilities, the selection of which is again driven by the agent's specific goals. Thus, the link between the perceived properties of the object and the subject possible actions might be based on a stored representation of the particular object (Humphreys, 2001) and, furthermore, to the mediation of conceptual knowledge and conventionality it is to add also the one of action intentions (Borghi, 2005). This perspective highlights not only the relational character of affordances, but also the high grade of flexibility in their emergence, their being driven by the context and the situation as a whole.

In a similar direction, an influential views of affordances indicated them as "states of the observer" (Ellis & Tucker, 2000, p. 453). In contrast to the notion of dispositional properties of objects and events, or features of a situation, as discussed above, the definition of micro-affordances originally proposed by Ellis & Tucker (2000) look at them as dispositional properties of a viewer's brain, arising as the result of adaptations of the nervous system not only during phylogenetic evolution, but also in the life span of an individual[1]. This

---

[1] To support this description, Ellis & Tucker (2000) indicated the kind of neural architecture which can explain both the developmental nature and the biological sense of micro-affordances. It is the one defined in the theory of Neural Darwinism (Edelman, 1978, 1987). According to the theory, an individual brain is scaffold by a pressure similar to natural selection, called somatic selection. Adaptive behaviour development requires an integration of sensory and motor processes, and learning coordinated actions (e.g., grasping) is the direct outcome of the gradual adaptation of neuronal groups, which are the morphological unit somatic selection operates on: those groups involved step by step in a successful action (such as in grasping one leading to contact with an object) become the preferential neural mapping for that purpose. The mappings resulting from this process not only sustain adaptive behaviour, they are also seen as the basis for categorizing external objects. Micro-affordances are intended to be in these mappings motor components, which have come to represent visual objects during development.

definition implies that, even if the term micro-affordances was chosen to underline the similarities with the gibsonian tradition, as micro-affordances are brain structures capable of representing possibilities of action they have to be the outcome of previously stored perception-action experiences. Furthermore, as the notion of micro-affordance directly refers to potential elements of an action, it means that, if for example we refer to the case of a manipulable object affording hand grasping actions (e.g., a wine-glass), what is elicited is not grasping "in general": only specific components of grasping, which are the more appropriate to suite the action on that particular object, are selected. It implies that the activation of micro-affordances is mediated by a visual identification of the external objects, thus their emergence depends from selective attention (Vaino et al., 2007). However, in line with the gibsonian view, micro-affordances facilitate simple and specific kinds of interaction with objects and they do not pertain to complex actions, which in order to be successfully executed would need - as said before - the mediation also of a particular intention to act, a certain agent's specific goal to be reached by the use of the object (e.g., drinking). In this direction, as they come into play independently of the goal of the agent, micro-affordances have been explicitly defined as structures which are automatically activated (see par 2.5 and my empirical study presented in chapter 4 for further discussions on this topic). This idea seems to be confirmed by compatibility studies as the one by Tucker & Ellis (1998) in which participant judged with bimanual lateralized key-presses if handled objects in figures were lying on their habitual plane or not (upright oriented/inverted). The position of the handle of the object, totally irrelevant for the task, could match or not the side of the manual response (left/right). Results showed that the left-right orientation of the handle played instead a role, facilitating compatible responses (those that were carried out by the congruent hand, e.g., handle on the right and response given with the right hand) and interfering with those same responses if incompatible (those carried out by the incongruent hand, e.g., handle on the right and response given with the left hand). This kind of evidence is considered as a support for the existence in the brain of a direct, non-semantic route from perception to action (e.g., Milner & Goodale, 1995; I discuss further this topic in paragraph 2.4 and 2.5) and also for the idea that the psychological effect between perception and action is unidirectional, that is only from perception to action (see Girardi et al., 2010). This position in anyway problematic (see par. 4.4 for a detailed discussion).

It is worth to notice, before closing this theoretical paragraph on the contemporary notions of affordance, that in the literature there is another very influential approach to the coupling of perception and action, provided by theories of *common coding*, which assume that representations of perception and action are based on the same cognitive codes (e.g., Hommel et al., 2001; Prinz, 1990, 1997), and operate in the same representational space. With respect to the relation of perception and action, this account assumes an ideomotor link (Greenwald, 1970) that would predict a bidirectionality and reciprocity between their effects. The central assumption here is that motor actions and perceptual effects are highly interrelated and mutually dependent: movements are exhaustively coded in terms of their perceptual effects (Girardi et al., 2010). Evidence for this idea has been provided by studies showing effects of motor-visual priming, for example that the perception of a manual movement facilitates the following execution of congruent motor actions (e.g., Brass et al. 2000; 2001). Sturmer et al. (2000) required participants to open or close their hand in response to the color of a picture of an opening or closing hand: the results showed faster responses in case of a match of the action depicted in the picture and the actual action that was executed. Moving from such a theoretical framework a recent paper (Haazebroek et al., 2011), that proposes a computational model of cognition as grounded in perception and action processes, offered a very comprehensive definition of affordance (applicable also to complex ones), in which the action planning can modulate the perception of objects (see par. 2.3). Indeed, the authors affirm that an affordance is effectively realized by allowing an automatic translation of perceptual object features (e.g., object shape) to action, but that this happens in overlap with the prediction of the action effect features (e.g., hand shape). In this way, by focusing the attention on certain features in the action planning, these dimensions become enhanced in object perception, and as a consequence these sensory features are processed more strongly than the others. Therefore, "affordances can be defined in terms of the intrinsic overlap between stimuli features and action effects as encountered in sensorimotor experience […] By having common codes, perceiving objects fundamentally implies anticipating intrinsically related action plan features" (Haazebroek et al., 2011). Because of the ideomotor link they show, the activation of those features shared by both objects and actions easily biases the cognitive system in programming and executing the appropriate motor responses. Crucially for the aims of our research, the final element that plays a role in complex behavior is the task: the task with is *goal* fully determines the context, driving the monitoring of selected effect features in the situation. It gives the hint that indicates among the object and the action features, the ones which are really relevant for the agent's goals, specifying which of the available affordances apply in the situation.

To close this discussion about the contemporary theoretical views on affordances, I want to underline that all the approaches sketched in this paragraph have in common the following assumption: that at least some compatibility effects should arise and be observed in dependence of the relations between external (visual) objects and the possible real-world actions that can be performed on them (by a human agent). Furthermore, the most interesting accounts of how it can happen diverge from the ecological view of affordance, as in order to explain adaptive behavior they explicitly recover the notion of conceptual representation to the discussion of the relation between perception and action. Finally, the role of concepts for action results to be strictly connected with the current context, intended as the whole situation which is determined by taking in account the acting subject goals too.

2.3. Behavioural evidence on affordances

In the literature, evidence on the existence of an affordance effect comes from behavioural research as well as from a variety of neural techniques - neuroimaging, neuropsychological, neurophysiological, single cell recording studies. These investigations are aimed at addressing the question of how perceptual and action representations are related in embodied representations of affordances.

Much behavioural evidence has been provided in support of the idea that visual representation of objects includes the partial activation and potentiation of the motor patterns associated with the actions they afford. It is now well established that the visual presentation of an object can affect subsequent action. For example, in a semantic categorization task (artificial/natural) in which participants responded by grasping a manipulandum ad-hoc, Tucker and Ellis (2001) reported that responses given by mean of precision grip (or pinch, that is grasping with the index finger opposed to the thumb) were executed faster when small objects were presented, and power grips (or clench, that is grasping opposing the fingers to the palm) were executed faster when big objects were presented. This evidence suggests that the visual presentation of an object can automatically activate the specific action it affords. In the same direction, Rumiati & Humphreys (1998) asked participants to pantomime the correct action for an object presented in figure, showing that there was a greater amount of errors of visual nature (e.g., mimicking the use of a hammer for a visually-related razor) than errors of semantic nature (e.g., mimicking the use of a hammer for a semantically-related saw): the authors – that in a second experiment observed on the contrary more semantic errors when participants mimicked to words - suggested a direct link between visual object representations and action. There are also a number of behavioral results on the affordance effect which confirm that a motor simulation is activated, but that it is not necessarily so specific. In the literature, stimulus-response compatibility effects, that show facilitated responses to visual elements appearing in the left/right space with the left/right hand - often considered as simple Simon effects (e.g., Proctor & Cho, 2006) - have been reported using, rather than limb movements, lateralized key-press responses. As seen above, Tucker & Ellis (1998) showed that judging if an object was upright or not was faster when the handle of the object (e.g., a coffee pot) was spatially congruent in respect to the side of the manual response. Analogously, Tipper et al. (2006) presented door handles in passive or active states to participants which have to discriminate their shape or color. They showed that simple button-press responses in making these decisions were faster when the handle was aligned with the responding hand, but only in the shape judgments, with the effect being greater for the active state (for similar findings, i.e., absence of an affordance effect in color judgments, see Pellicano et al., 2010). These findings are very interesting also because they seem to suggest that judgments which are usually considered to be “implicit” in respect to object action-features (e.g., upright/inverted, artificial/natural) probably need to successfully perform the task at least an access to object shape features, that is an access to previous action-related knowledge about the object (for further discussion, see par. 4.4). Also Phillips & Ward (2002) reported related evidence priming participants with handled objects in figure, such as a picture of a frying pan. The handle could be on the left, right or middle, and be placed near or far from the participant. The prime was followed after a varying stimulus onset asynchrony (SOA) by a target-stimulus which required a single button press by left/right hand or foot. The data showed a correspondence effect between handle orientation and side of response which increased with SOA regardless of the response modality (which could be hands, hands crossed, or feet). These results supports the idea that participants form a simulation of their interaction with the objects, being sensitive to the relationship between the (irrelevant) object property, the handle location, the current position of their limbs in space and the force of the effectors (Borghi, 2005). They may also indicate that the affordances of an object do not always potentiate a specific response for the effector more suited to respond, but that are activated more global, abstract spatial codes, which may potentiate a wide variety of responses in the congruent side of space. Indeed, additional evidence indicated that visual object processing is able to activate action representations even when the task does not involve an overt motor response (e.g., Kalénine et al., 2012; Lee et al., 2013; Myung et al., 2010).

On the whole, the reviewed behavioral results confirm that the activation of affordances corresponds to a (more or less specific) motor potentiation depending on visual objects presentation. A greater number of behavioral experiments than the few reviewed, which mainly employed compatibility paradigms (e.g., Bub et al., 2003; Creem & Proffitt, 2001; Edwards et al., 2003; Fischer et al., 2008; Girardi et al., 2010; Tuker & Ellis, 2004), supported the idea that the simple observation of objects can activate an embodied simulation of appropriate actions – it was confirmed also in modeling works (e.g., Caligiore et al., 2010). However, the visual presentation of an object is not the only factor able to potentiate the associated affordances, as also the intention to perform an action can modulates visual processing by favoring those perceptual features which are action related and eliciting overt limb movements. In Craighero et al. (1999) the preparation to grasp an object facilitated the detection and discrimination of visual shapes congruent with it. Participants were trained to prepare a grasping movement towards a bar oriented clockwise/counterclockwise; in the experimental block, they had to really execute that movement after the presentation of a go-signal which was the picture of a bar oriented clockwise/counterclockwise (or of a circle, in catch trials). Grasping latencies were shorter when the orientation of the go-signal and of the bar to grasp were congruent. The compatibility effect was present with different effectors and even when the response was not of the type affected by kinematic relations with the grasping movement (i.e., eye blinking), but extinguished if the visual properties of the presented target were others than those for which the grasping movement was trained. Thus, as noticed by Borghi (2005, p. 17), “the effect was not due to orientation effects per se, but to the matching of the motor affordances of the visual object with those of the real object”. In the same direction Bekkering & Neggers (2002) found that the first eye movement performed to select an oriented target-object situated among distractors was more accurate when the object was individuated to be then grasped than to be pointed. As orientation is a relevant feature for grasping only, not for pointing, these results suggest again that action planning can influence visual processing. Further evidence of motor-visual attentional effects is offered by a number of hand-primes studies (a review in Borghi & Cimatti, 2010). For example, Borghi et al. (2007) in a categorization task asked participants to classify (with bimanual lateralized key-presses) as artificial or natural big/small manipulable objects, that afford a power/precision grip, while the objects were preceded by hand-primes pictures depicting power/precision grips. The results showed a compatibility between the posture depicted in the prime (power/precision) and the grip required by the target-objects (power/precision), even if the effect was present only in the case in which the experimental block was preceded by a motor training, in which participants reproduced with their hands the prime postures in order to strengthen the potentiation of the action elements associated with the real interaction with the showed objects. To further investigate the nature of this motor priming effect, Vainio et al. (2008) replicated the study using video-clips in place of still-hand pictures. The results showed a compatibility effect even if in this experiment participants did not perform any preceding motor training.

Overall, the reviewed body of behavioural research shows tight (bidirectional) relationships for the visual presentation of an object and the execution of the actions that object affords (which can also be very specific). As the behavioural effects occur in the context of cognitive tasks (e.g., deciding whether an object is upright/inverted, artificial/natural etc.), this evidence implies a fundamental role for motor simulation in cognitive processing, which - as seen before - is predicted by embodied accounts of cognition. Finally, as the reviewed behavioral studies showed compatibility effects in dependence of different object properties (orientation, size etc.), it is worth mentioning here the proposal by Borghi & Riggio (2009) to distinguish between stable affordances (e.g., shape), based on information stored in memory, and temporary affordances (e.g., orientation), variable across situation because based on actual object properties (further discussed in par. 2.5). This is a proposal that clearly underlines a high contextuality in the emergence of the affordance effect.

2.4. Neuroscientific evidence on affordances

Neuroscientific evidence has shown that the visual presentation of manipulable objects activates regions known to be associated with motor activity both in monkeys and humans. The discovery of mirror neurons in the area F5 of the monkey frontal lobe (i.e., those neurons firing both during the execution of a certain action and the observation of that same action), individuated directly by mean of single cell recording, has prompted much speculation about the brain's capability to simulate other people's actions, and it is often argued that mirror neurons are the basis of action understanding (e.g., Rizzolatti & Craighero, 2004). Indeed, the response properties of such neurons appear to validate many assumptions made by embodied theorists and serve as a solid basis of simulation accounts of cognition (e.g., Gallese & Sinigaglia, 2011), even if the role of mirror neurons in representing concepts remains debated. In macaque area F5 there is a small subclass of visuo-motor neurons called canonical neurons, that – unlike “pure” mirror neurons - selectively respond both to the presentation of an object as to the performance of an action on that object (e.g., Murata et al., 1997; Jeannerod et al., 1995; Raos et al., 2006). For example, for some of these neurons the size of an object affects their firing such that those cells firing during precision grips execution favour objects that afford precision grips actions; similarly, those that fire during types of grasps such as power grips and full hand grips favour larger objects (Sakata et al., 1995). This suggests that these neurons are coding complex sequences of movements that pertain to interaction with specific objects (Rizzolatti et al., 1988), thus that they are resonating to their features. This type of neuronal response, called *motor resonance*, has been taken as evidence of the automatic activation of motor-representations by the visual presentation of an object that may constitute the neural basis for object affordances (e.g., Garbarini & Adenzato, 2004). Indeed, the identification of affordances implies that the same neurons are active to encode the motor acts they control and to respond to the visual situation that elicit these motor acts. Recently, Ferrari et al. (2005) discovered a unique class of mirror neurons that selectively respond to the observation of actions performed with tools, called tool-responding mirror neurons. These neurons are triggered by actions that are made with a tool, compared to similar goal-directed actions without tools. They do not respond to the visual presentation of a tool alone, so this suggest that such neurons might be representing the global action-vision relationship for tools. On the whole, the discovery of these different types of neurons in monkeys’ brain support the idea of the occurrence of a simulation of motor actions that is triggered by the visual perception of objects.

Interesting evidence of the occurrence of motor resonance mechanisms are present also in the findings of the neural literature on humans. For example, a number of fMRI results have revealed fronto-parietal activity during passive viewing of manipulable objects (e.g., Binkofski et al., 1999; Chao & Martin, 2000). Also in PET studies an activation in the premotor cortex in response to tools have been observed during passive viewing (without any requested response), and also during naming and description of their use (Grafton et al., 1997), with the results interpreted as evidence that the premotor cortex is involved in representing object semantics. In an experiment in which participants were asked to imagine themselves manipulating objects they were visually presented with, brain activity was shown in the prefrontal, parietal and motor cortex, as in supplementary motor area (Grèzes & Decety, 2001). Similar findings have been reported also in an orientation judgment task (upright/inverted) performed on pictures of manipulable objects (Grèzes & Decety, 2002). A similar fronto-parietal circuit was active also during execution of the specific hand grips afforded by visually presented objects (Grèzes et al., 2003). Additional evidence also suggested that this motor activation evoked by visual objects might be context-dependent. For example, Gerlach et al. (2002) showed premotor cortex involvement in a semantic categorization task (natural/manmade), but not in object decisions (real/non-real), thus the emergence of the resonance was mediated by the different goal of the tasks. This is a very interesting result for the hypotheses that guide the research of this thesis, as it highlights a certain flexibility in the occurrence of the motor resonance, in direct analogy to what observed above for the embodied simulation processes (which are thought to be based on that neural mechanism).

Additionally, a number of brain imaging studies have shown that the human brain is differently activated by tools in respect to objects which do not afford action (a review in Martin, 2007). A PET study showed that, although naming images of animals and tools activated overlapping visual cortical structures, there were also different category-specific brain activations: in the occipital lobe in response to animals, in motor cortical areas in response to tools (Martin et al., 1996). Furthermore, in a TMS study the observation of tools (e.g., a cup) that could be or not broken showed significantly larger MEPs in cortical motor areas for entire items, indicating that these regions sustain a graded identification of affordances in dependence of the perceived visual context (Buccino et al., 2009). Additionally, a meta-analysis over several PET studies (Devlin et al., 2002) that examined specific activation patterns for man-made objects/tools with respect to other object categories (e.g., fruits, vegetables) found evidence of activations elicited by tools in the left posterior temporal regions especially when subjects were engaged in semantic tasks (e.g., naming), not during simple passive viewing. This result implies an influential role for the subjects' orientation to visual objects in function of the task on the kind of motor representations evoked.

The embodied account of object representation is consistent with the reviewed brain imaging patterns. They have often been interpreted in light of a very influential architectural and functional model, proposed by Milner & Goodale (1995). In this model, there are two streams where visual information is processed. Information about object shape is processed in the ventral stream, extending from V1 into the inferior temporal lobes, while information about movement, location and correct metrics to grasp it are processed dorsally, in the posterior middle temporal gyrus and regions of the anterior inferior parietal lobe. The two routes have been respectively identified as a "what" (or "how") pathway, which contains stored representation that mediate object recognition, and a "where" pathway, which contains on-line information that sustain actual interaction with the objects. Affordances were thought to be processed only by the dorsal stream (e.g., Tucker & Ellis, 2004). However, recent proposals seem to be able to re-interpret this (strong) distinction of a semantic and a perceptual route. A discussion about this topic, which is fundamental for the empirical research presented in chapter 4 and 5 is sketched in the following paragraph.

Overall, this body of neural research supports the idea of a motor resonance as the base of embodied simulation and object representation. As the behavioral findings discussed in the previous paragraph are consistent with the patterns of brain activity reviewed here, the presence of resonance mechanisms in the brain seem to be the neural counterpart able to sustain the (reciprocal) effects of object perception and action preparation/execution indicated previously.

2.5. Manipulative and functional affordances

With respect of the two-routes model of visual processing (Milner & Goodale, 1995) we referred to in the previous paragraph, affordances were initially seen to be allied exclusively to a dorsal stream processing (see Borghi, 2005; Young, 2006). However recently some researchers, starting from both neurological and behavioural findings, have proposed an important distinction between affordance classes that may help clarify the possible contributions of the dorsal and ventral regions. For example, Creem-Regehr & Lee (2005) proposed to distinguish between actions associated with the *structure* of the object and actions associated with the *function* of the object, moving from the data they collected in an imaging study. In the experiment, they presented participants with 3-D images of graspable objects in two tasks: passive viewing and imagined grasping. Half of the presented items were everyday tools (e.g., hair brush), the other half were objects not associated with a conventional use (e.g., elongated block). In the passive viewing task, only known tools activated posterior parietal and premotor regions along with the middle temporal gyrus. In the imagined condition, a similar result was observed with both kind of items, but there was anyway evidence that frontal regions were active more for familiar tools (e.g., hair brush). The results indicated that the previous experience with a tool function constrains the way in which functional knowledge is represented in comparison to structural knowledge. This suggests that the dorsal areas might be more important in controlling structural grasping (e.g., processing both natural kinds and tools), while the ventral areas might be more related to functional aspects (e.g., processing objects affording functional use).

The specificity of affordances activation has also been investigated in the study by Bub et al. (2008). The authors trained participants to produce a particular hand action (e.g., poke) in response to the colour of a visually presented object (e.g., red); any item in the set of objects used afforded a structurally and/or a functionally action, that was compatible/incompatible in respect to the manual action response previously trained to its colour. The results shown that responses which were congruent with moving/using the objects (e.g., poking-calculator; clenching-spray bottle) were executed faster than incongruent actions, suggesting that object processing may recruit both of the action types. Furthermore, Masson et al. (2011) showed that manual movements in response to objects are facilitated with those objects whose handle is congruent with the response action to execute (e.g., a power grip executed after the prime of a beer mug), and suggested that the presentation of objects automatically specifies the action the object affords. But the issue of automaticity is controversial (as discussed in chapter 4 and 5). For example, Bub & Masson (2010) trained participants to respond to colour cues by grasping a manipulable apparatus with their left/right hand, or by pressing a button with their left/right hand; in both cases participants were primed by pictures depicting either an object congruent (e.g., a beer mug with the handle facing the cued hand) or incongruent (e.g., a beer mug with the handle facing away from the cued hand) in respect of the response performed. They found that when participants responded by a grasping movement, showing a compatible object shortened response latencies, while no effect was not present for press responses. The authors concluded that affordances are elicited by objects under the guidance of the specific motor intentions that characterized the action executed, and not by simple ballistic movements. This goes in support of the possibility that objects do not (always) automatically elicit (general) motor simulations, but that motor simulations critically follow the guide of the action intention of the subject, having the affordance effect an highly contextual characterization.

In this vein, Jax and Buxbaum (2010) investigated the distinction between structural/volumetric grasping (i.e., grasp-to-move) and functional grasping (i.e., grasp-to-use) presenting participants with (real) objects that afforded the same grasping to be moved and to be used (e.g., drinking glass) and with objects that afforded different structural and functional grasps (e.g., a calculator, affording clench for moving it and poke for using it). They showed that initiation movement for use actions was slower for objects associated with distinct move-related actions (that they called conflict objects, discussed further in chapter 5) as compared to objects for which use- and move-related actions are similar (non-conflict objects), while initiation for move-related actions was not different in both conflict and non-conflict objects. These results suggests that move-related activations may be relatively rapid, and so produce an interference with the planning of use-related actions, which is longer because characterized by retrieval of motor information.

If only functional responses require activation of long-term conceptual representations, whereas structural responses can be activated more quickly than functional responses, the existence of two classes of actions associated with a given object raises questions about the factors that may influence the strength and time course of their activation. One possibility is that both types of action are invariably activated during object recognition, but it is more likely that affordances activation may be mediated by task goals and context. It is why Jax & Buxbaum (2010) proposed that the intention to act on an object triggers a race-like competition between structural/volumetric and functional responses during action planning (see also Cisek, 2007). In support of this idea, there is already evidence in the literature, at both the behavioural and neural level, that the visual context in which objects are embedded exert an influence on their processing (e.g., Bar, 2004; Borghi et al., 2012; Gronau et al., 2008; Kalénine et al., 2013; Mudrik et al., 2010; Natraj et al., 2013), and it is very likely that for the objects associated with more than one action (such as the conflict objects of the latter study) is the visual context what selectively amplifies one of the actions associated with it (e.g., Wurm et al., 2011). In an iterative manner, this bottom-up facilitation of an object-related action by the context may resonate with the intention-driven facilitation of action by the action planning (for similar proposal see Chambon et al., 2011; Shen & Paré, 2011).

Finally, Buxbaum & Kalenine (2010) have explicitly proposed that different brain circuits might sustain the two kind of affordances: one dorso-dorsal stream (a “pure” dorsal pathway) is mostly involved in the processing of object structure, whereas a dorso-ventral stream is more related to the processing of object function. The overlap point they propose in the activation of the two streams (at least in the condition of functional activations) smoothens the original, very dichotomist distinction of the ventral and dorsal streams. An analogous proposal has been presented by Borghi & Riggio (2009), that focused on the distinction of stable and variable affordances. Stable affordances would emerge primarily from properties which are rather constant across contexts like shape, while variable affordances would emerge from incidental properties such as orientation. A special case is represented by canonical affordance. Considering, for example, a cup, it is obvious that even if we interact with cups in many different orientations depending on specific situations and goals (e.g., move, wash, drink, fill etc.) there is a tendency to find this object in a typical orientation that is upright (especially in contexts of use). Due to its higher frequency, it might be useful to store additional information on this canonical orientation. At a neural level, stable affordances are thought to be represented more ventrally, whereas variable affordances more dorsally. Thus, the indication of the authors is also in this case for a partial overlap of the two kinds of affordances, as the first would activate a dorso-dorsal stream, the second a dorso-ventral pathway (see also Rizzolatti & Matelli, 2003).

Overall, the results discussed here are in line with the embodied literature on object-related action representation, reviewed in par. 2.3 and 2.4. Indeed, they confirm again the evidence that the perception of manipulable objects lead to the activation of appropriate motor representations. Furthermore, these results extend the reviewed evidence on affordances indicating that the motor simulation can be (selectively) triggered in respect of structural and functional properties of the presented objects. The elicited simulation, in line with the description made in par. 1.3, can vary even for the very same object if it is found across different contexts, or perceived under different action intentions.

2.6. Mirror mechanism in motor resonance

On the whole, the evidence reviewed so far clearly suggest that when a subject is exposed to manipulable objects, its affordances elicit the activation of appropriate motor programs. The effects have clearly confirmed that the possibility of evoking motor behaviours rests "on the possibility of a sensory-motor coupling allowing any onlooker to map such a feature onto the motor possibilities belonging to his or her motor repertoire" (Costantini & Sinigaglia, 2012, p.436). The reviewed findings have been interpreted as evidence of a motor simulation, idea that holds at both the neural and the behavioral level. This last introductory paragraph is dedicated to a brief overview of the findings on the motor resonance which triggered by the observation of action execution, as it is one of the basis for the hypotheses of the experiments described in chapter 4 and 5.

In par. 2.3 results of behavioural hand-primes studies were reviewed as evidence for the existence of a motor-visual attentional effect. In light of the description in par 2.4 of the motor resonance, it is possible to argue that at a neural level the effect of the hand-prime was based on this same mechanism, and that its architectural basis might be the mirror neuron system. Indeed, it was observed in primates that - differently from the canonical neuron system - mirror system is activated during the performance of an action (e.g., grasping, moving an object) and the observation of other actors performing the same action (e.g., Di Pellegrino et al., 1992; Gallese et al., 1996). Mirror neurons in humans have not been ultimately identified, as it is very hard to observe directly the activity from single neurons as done many and many times in monkeys. It happens that, for a number of (valid) reasons, single cells recording in humans is almost impossible, except in rare occasions (e.g., Mukamel et al., 2010). However, even if single neurons have not been directly identified, in the literature of neural research several results stay in support of the existence of a human mirror system (review in Rizzolatti & Craighero, 2004; Gallese, 2011).

Much evidence shows that a resonance phenomenon is triggered when we observe others acting, interacting with objects, and even performing complex movements such as playing basketball or dancing (e.g., Cross et al., 2006). The neural data indicate that the mirror system activation is higher and the resonance is stronger if actors and observer share a similar motor repertoire. For example, in an fMRI study by Calvo-Merino et al. (2005), when capoeira dancers saw other people dancing capoeira, their mirror neuron areas were more activated than when they observed classical ballet dancers. In the same vein, Aglioti et al. (2008) demonstrated that, while observing free throws videos, elite basketball players (from the Italian first basketball league) predicted the outcome of the shot earlier and better than both novice players and expert observers (e.g., coaches). The authors related this advantage in predicting the outcome to the elite players' higher capability in reading the actor's body kinematics in the early phases of the movement. The results reveal a more detailed motor resonance in function of motor practices and experiences: motor expertise specifically contributes to anticipating and predicting the ongoing of other people's actions. Furthermore a study of Calvo-Merino et al. (2006), aimed at controlling if these kind of effects could be simply due to the higher familiarity of the observers with certain types of movement, showed a larger motor resonance effect when classical ballet dancers observed movements performed by other classical ballet dancers of the same gender. As the sequences of movements performed by the two genders were equally familiar to the observers, this evidence supports the explanation based on the occurrence of a motor simulation. In the same direction, evidence of higher mirror system activity was found during the observation of actions performed by a human hand in comparison to hand actions performed by an artificial agent (Perani et al., 2001; Tai et al., 2004), whereas observing actions impossible to execute with respect of the observer's body structure does not elicit a motor resonance (Stevens et al., 2000). An fMRI study by Buccino et al. (2004) presented human participants with videos of biting and communicative gestures performed by a human, a monkey or a dog. Independently from the species, viewing biting actions activated in the participants (among other areas) regions thought to be part of the mirror system, while viewing communicative gestures elicited activation of the human mirror systems only for actions performed by humans (lip-reading) and monkeys (lip smacking), not by dogs (barking). Thus, only the actions belonging to the observer's motor repertoire were mapped by the motor system. But motor resonance is not only modulated in function of body similarity and shared motor experience: other inter-individual differences, such as cultural factors, can affect this mechanism, as it was suggested by the TMS study by Molnar-Szakacs et al. (2007). In order to investigate the neural bases of cross-cultural social communication, the authors measured participants' corticospinal excitability during observation of culture-specific emblems (i.e., autonomous communicative gestures). Foreign Nicaraguan and familiar American emblems (as well as meaningless control gestures) were performed by Euro-American and Nicaraguan actors, and the data indicated that in Euro-American participants were found a higher corticospinal activity during the observation of the Euro-American actor. On the whole, these neural results clearly suggest that the resonance emergence is very flexible, as it is sensitive to a multiplicity of factors that can vary across different contexts.

Among the contextual factors that modulate motor simulation in function of action observation, behavioral evidence indicates also perspective. For example, Vogt et al. (2003) demonstrated it with a simple response procedure. Participants, given the instruction “clockwise” or “counterclockwise” and a prime, were asked to grasp a bar in the indicated orientation. By manipulating the perspective of the hand presented as a prime, which could either be congruent with the hand posture of the participant’s own hand (egocentric perspective) or the hand posture of a person in front of the observer (allocentric perspective), they found a compatibility effect when a hand-stimulus was given as a prime for the egocentric perspective. Similarly, Bruzzo et al. (2008) presented hand-primes and target-objects in egocentric or allocentric perspective asking participants to decide whether the action displayed by the hand was suitable (or not) for interacting with the object depicted. The data revealed an advantage of the egocentric over non-egocentric targets, and moreover an interaction that indicated that the fastest responses were performed when egocentric primes were followed by egocentric targets. These results show that it is easier to put ourselves in others’ shoes and to resonate to their actions when we share with the actor action-relevant characteristics, as the point of view is. Additional behavioural evidence on the graded nature of motor resonance and the role of individual expertise is provided by the study of Pezzulo et al. (2010) on climbers. They divided for expertise participants in two groups (novice/expert), and presented them with three routes distinguished in function of their difficulty (easy/difficult/impossible). A following recalling test showed no differences between participants for easy or impossible routes, but for the difficult route the performance were significantly better in experts than novices.

To close, much evidence clearly demonstrated that our brain is able to resonate, as to the affordances of external objects, with the actions performed by other agents. This resonance mechanism is not fixed and serial, but has a graded structure: its emergence is the result of the mediation (at least) of body structures similarity, shared motor skills, cultural factors and perceptual features of the current situation. Thus, the findings reviewed confirm the context-dependence of the motor simulation, as they demonstrate a high flexibility in the activation of the neural circuits that sustain conceptual representations of action possibilities and action meaning.

3. Language and embodiment

The second topic of my thesis concerns the experimental investigation of verbal language. In this chapter, I sketch a discussion at both a theoretical and empirical level on the embodied view of language, and describe results in the literature that indicate the occurrence of simulation processes triggered by language. Subsequently, as much research has the aim of showing the similarity of the simulation evoked by language and by objects/actions processing, the possible differences in these processes are discussed, introducing an issue that is experimentally investigated in my study reported in chapter 6. Then, as the embodied literature is prevalently focused on concrete language, while the grounding of abstract concepts is a test bed for any theory of cognition, I will describe theoretical proposals inside the embodiment framework for explaining abstract words meanings, that I also investigated empirically in two of my following reported studies (chapters 7 and 8). Finally, following this description of language grounding at the ontogenetic level, I refer also about iconicity and sound-symbolism (my investigation on this topic is reported in chapter 9) as the possible phylogenetic grounding that supported the emergence of contemporary human languages.

3.1. Embodied and grounded view of language

Many widely accepted theories in cognitive science have been assuming for decades that cognition is divided in lower and higher level processes (e.g., action and language, respectively), and that any mental process would be handled in separate systems. Following the assumption of a modular and encapsuled organization, each system can work independently from the others, according only to its own principles (e.g., Fodor, 1975, 1983; Marr, 1982; Pylyshyn, 1986). Under this traditional view the mind is considered no more than a symbols processing unit (see par. 1.1). These symbols are posited as amodal, abstract and arbitrary, because their content is totally detached from perception and action processes, so they do not exhibit any systematic relation to the original sensory patterns they emerged from (Barsalou, 1999, 2008; Anderson & Spivey, 2009). These kind of theories have dominated cognitive science for decades because of their power to account the relations of types and tokens, to combine symbols productively, as well as to represent abstract concepts' content (Fodor, 1975; Pylyshyn, 1986). But much evidence against such a modular view has been accumulated in more recent years, and a new understanding of the mind emerged by relating cognition directly to the mechanisms which govern the perceptual processes feeding into cognition, as well as the actions selected by and guided through cognition. The physical instantiation of the cognitive system as a brain inside a body which exist in the world provides all its principles, and this is the core idea that defines *embodiment* (see par. 1.2). Thus, what embodiment proposes in deep contrast with the traditional view is that perception and action are not peripheral with respect to cognition, and that our cognitive activity is embodied, environmentally embedded and perceptually grounded (e.g., Barsalou, 2008; Glenberg, 1997). Far from the idea of the subject as an isolated observer, cognition is a matter of *situatedness* (see par. 1.2 and 1.3).

According to embodied theories, from the continuity of perception, action and cognition (Berthoz, 1997; Gibson, 1979; Hommel et al., 2001; Prinz, 1990, 1997) follows that also language is grounded in sensory-motor activity. While in the traditional view words would be ungrounded, thus understood only in terms of other words associated to them (e.g., Landauer & Dumais, 1997, see par. 1.1), the embodied view of language comprehension proposes that the understanding of a word/sentence occurs by mean of mental simulation (see par. 1.3 and 1.4) of the entities, events and situations that are linguistically described. This implies that the neural areas recruited by language are the same involved during perception and action (e.g., Anderson, 2010; Barsalou, 1999; Barsalou et al., 2003; Borghi, 2012;

Gallese, 2008; Gallese & Lakoff, 2005; Gibbs, 2006; Glenberg, 1997; Glenberg & Gallese, 2012; Glenberg & Robertson, 2000; Jirak et al., 2010; Pecher & Zwaan, 2005; Pulvermüller, 2005; Zwaan, 2004). The idea proposed by embodiment is that words are something able to sustain and guide action in the world, as language activates online simulations that help us to interact appropriately with objects, entities, agents and events in the environment (Barsalou, 2009; Borghi & Cimatti, 2010; Gallese, 2008; Gallese & Goldman, 1998; Glenberg & Robertson, 2000; Zwaan, 2004). Therefore, this (re)activation of modal contents associated with particular described scenes is what serves as the "engine of meaning" (Bergen et al., 2007, p. 76).

A growing body of research has revealed that words, far from being linked in an arbitrary way to their referents, are grounded in perception, action and bodily processes. Much evidence has been obtained recently on action-related linguistic stimuli by studies performed with several techniques (e.g., behavioural, kinematic, eye tracking, brain imaging). This research has demonstrated that continuous visual availability is not necessary to create a bias in the sensorial and motor system: concept activations seem to be sufficient to do so. Indeed, the emergence of affordance effects (see chapter 2) have been observed also when participants were only briefly exposed to a perceptually degraded object as when they simply see (e.g., Tucker & Ellis, 2004) or hear (e.g., Myung et al., 2006) a word denoting the object. Thus, analogously to visual stimuli (as described in chapter 2), objects names would evoke the affordances of their real-world referents; similarly, motion verbs would evoke the action they refer to, eliciting the corresponding motor programs, and so on.

As object information is stored in the brain in terms of affordances (e.g., Ellis & Tucker, 2000; Glenberg & Robertson, 2000), it is plausible that a word which refers to a certain object would activate the same affordances as the object itself (this assumption is directly investigated in a study I performed on affordances using both visual and verbal stimuli referring to the same objects, reported in chapter 6). However, as the same object could possess simultaneously concurrent affordances (e.g., tools, see par. 2.4 and 2.5), or be linked to more than one action across contexts, one issue in the literature concerns how the simulation activated during language comprehension is built up and detailed (Bergen et al., 2007). If simulation processes should sustain comprehension as well as prediction and inference (Barsalou, 2008, 2009; Pezzulo & Castelfranchi, 2009), it is very likely that what would emerge during language-triggered simulation are only those elements relevant for the current goals and situations, or the more canonical in respect to the state of things linguistically described (Borghi, 2012). The evidence reviewed in the next paragraphs about both single words and sentences points exactly in this direction, revealing first a certain degree of sensitivity to a number of real-world properties in the simulations of linguistic contents, and second highlighting their flexibility in dependence on the available context-related information.

3.2. Behavioural evidence of embodied simulation during word and sentence comprehension

The embodied cognition view applied to language claims that when we understand words, following the neural exploitation (or reuse) principle (e.g., Anderson, 2010; Gallese, 2008), the same sensory-motor areas are recruited as when interacting with the objects/entities the words refer to. Similarly, when we comprehend sentences we internally re-enact, by mean of multimodal simulations, the state of the world those sentences describe. In this paragraph, the sensory-motor representations evoked by language are examined reviewing the results of behavioural and kinematics research.

In the literature, semantic effects on grip aperture by the exposition to single words have been observed. Indeed, the meaning of words that prime objects of fixed size can affect the movement directed towards those objects. For example, Glover et al. (2004) investigated the influence of object names on the grasping actions kinematics, asking participant to reach and grasp wooden blocks of different sizes after having silently read a word that could refer to small (e.g., GRAPE) or large objects (e.g., APPLE). The results observed larger maximum grip apertures after reading words referring to larger objects (e.g., APPLE) than after reading words referring to smaller objects (e.g., GRAPE). This suggest that the names - and obviously in a broader sense any element of grammar – are able to trigger a mental simulation of the referent real-world properties which overlapped with the planning of the action for responding, thus influenced online motor control of the required movement (see also Glover, 2004). This interpretation is confirmed by the early presence of the effect in the grasping kinematics (i.e., it was observed in the reach phase). Also in the study of Gentilucci et al. (2000) an analogous affordance-based effect on grasping kinematics by word reading has been observed. In order to investigate the influence of automatic word reading on visuo-motor transformations, the authors asked participants to reach and grasp big and small objects printed on with the adjectives SMALL or LARGE (in their Italian versions, see Experiment 4). The results showed again that the kinematics of the movement initial phase were affected by the meaning of the read words, as the word LARGE evoked grips with a larger maximum aperture, while the contrary was true with SMALL. Thus, subjects simulated the meaning of the word (activating the corresponding real-world property) and associated it to the response motor program, which was built up under the influence of the linguistic stimulus. The results of this study showed that also this class of words (i.e., adjectives) is able to influence motor control. This was further confirmed by a very similar study by Glover & Dixon (2002), which reported that maximum grip aperture was enlarged when subjects grasped an object with the word LARGE printed on top, as compared to the grasping of an object labeled with the word SMALL. This implies that word reading is able to activate automatically motor representations that constrain subsequent or simultaneous action. This finding is obviously consistent with the embodied account of object/action representations (see chapter 2). Other scholars have examined if qualitatively different types of hand postures corresponding to the real interaction with objects - rather than simple grip aperture - are activated in dependence of a simulation triggered by the exposure to their name. For example, Tucker & Ellis (2004) asked participants to use a manipulandum ad-hoc, that required either a power or a precision grip, to judge whether names of manipulable objects referred to natural or man-made kinds (see Experiment 3). The results showed faster power grip responses to words denoting objects requiring a power grip than to words denoting objects requiring a precision grip, whereas the reverse was true for precision grip responses. Taken together, these findings provided conclusive evidence that on-line visual processing of an object is not necessary to generate an affordance- based compatibility effect for properties like shape and size.

The simulation has been shown to be also sensitive to the spatial configuration and the order that parts of objects present in the real world. For example, Zwaan & Yaxley (2003) asked participants to judge whether the members of a word pair were semantically related or not. The names, which in the critical trials denoted parts of objects (e.g., ATTIC - BASEMENT), were presented vertically (one above the other), and could bear with their referents either an iconic relation (e.g., ATTIC presented above BASEMENT) or a reversed one (BASEMENT above ATTIC). The results showed that the iconic condition yielded significantly faster semantic-relatedness judgments than the reversed condition. The authors ruled out the explication of the effect as due to the order in which the words are usually coupled in every-day language with further experiments with horizontal presentation, where no effects were observed. Furthermore, and importantly for the embodied hypotheses, each participant saw a list that included along with the critical word pairs (that matched or not the vertical orientation of their referents) also semantically related/unrelated filler pairs that were not constrained with respect to vertical orientation. All the pairs were controlled for semantic-relatedness in Latent Semantic Analysis (LSA, Landauer & Dumais, 1997, see par. 1.1), and no differences were found between critical and filler pairs. Thus, the factor responsible for these findings - consistent with the notion of iconicity (extensively discussed in par. 3.6) - was the similarity between the word pairs' spatial arrangement on the screen with the spatial arrangement of their referents. In a similar direction, other studies reported compatibility effects between the word meaning and the location where the word were presented (i.e., spatial position on a screen). For example, Šetic´ & Domijan (2007) presented words referring to flying or non-flying animals at the top/bottom of a screen, asking participants to decide with bimanual key-presses whether the word referred to flying/non-flying animals. The results showed faster and more accurate decisions for words presented in a congruent than incongruent position (e.g., shorter latencies and less errors for STORK at the top than at the bottom of the screen). This may be due to a congruency between word meaning and stimulus position: as readers understand the meaning of a category by mean of situated simulations (Barsalou, 2008, 2009, see par. 1.3), the representation of the meaning of words referring to flying animal involves simulating the first-person experience of looking up at the sky to see the animal fly (see par. 1.3).

It is worth to notice here that in the literature is present an alternative explanation for this kind of results, which argues against the simulation account (not only of language comprehension) that these compatibility effects are simply due to markedness (e.g., Greenberg, 1963; Clark, 1969). Some authors pointed out that in many binary decision tasks the speed of response selection might be affected by a polarity correspondence (e.g., Proctor & Cho, 2006; Lakens, 2012), that would occur when a stimulus dimension with binary values is encoded as having a plus or minus polarity (e.g., flying/non-flying animal) and simultaneously the response alternatives are also encoded as plus or minus (e.g., yes/no), with response selection being faster when stimulus and response polarities correspond than when they do not (i.e., yes/flying). The basic idea is that the polarity coding of dimensions and responses is not entirely random, as some dimensions are always coded in the same way: a yes-response is always plus while a no-response is always minus, and in the same way up and right are always represented as plus while down and left are always represented as minus; so, accordingly, right key-presses are coded as plus and left key-presses are coded as minus. Furthermore, the opposite ends of these dimensions are thought to consist of a positive, unmarked member (e.g., more, big, tall), the name of which is extended to the dimension in its entirety, and a negative, marked member (e.g., less, small, short). Polarity correspondence has been thought to explain results on verticality similar to the ones reviewed, among a number of other effects of spatial congruency observed with many concepts - also more abstract, such as valence or number magnitude (e.g., Fischer et al., 2004; Meier & Robinson, 2004; Santens & Gevers, 2008). For example in Šetic´ & Domijan (2007) the flying animals always required a yes-response and the non-flying animals always required a no-response. Therefore, the polarities of position and response (up/yes and down/no) matched. To rule out the markedness-based account, Pecher et al. (2010) used two semantic judgment tasks for which the same linguistic stimuli required reversed responses. One task was an ocean judgment (i.e., Is it usually found in the ocean?) and the other task was a sky judgment (i.e., Is it usually found in the sky?). These tasks were chosen as sky and ocean have clear spatial positions, but are not the linguistically marked/unmarked end of a conceptual dimension. The authors used in both tasks the same set of names referring to objects typically found in one of the two locations (e.g., WHALE, EAGLE), and presented the target-word at the top/bottom of the screen. Subjects responded with lateralized bimanual key-presses counterbalanced across participants, with the same name in two different tasks required both yes and no-responses. The markedness account would predict that yes-response, higher location of the target word and right-hand response would be aligned (i.e., all coded as plus), but the results showed only that in the sky decision task responses were faster to words at the top than at the bottom of the screen, while the reverse was found in the ocean decision. Thus, this finding support the simulation account: spatial attention was directed by the situated simulation of the task-relevant conceptual dimension up or down. The asymmetry found in this study, as in many others (e.g., Casasanto, 2009; Meier & Robinson, 2004; Van Dantzig, 2009; Boot & Pecher, 2010), is very problematic for maintaining a polarity principle, which - as noted elsewhere - is a of generality that should be observed consistently (Landy et al., 2008).

A further confirmation of the simulation account derives from the literature that show, in place of facilitations, interference effects. Indeed, if during linguistic materials processing occurs a re-enactment of (multi)modal information, this process has to occupy sensory-motor systems (Barsalou, 1999, 2008). Following this principle, it is equally possible to find facilitation as interference effects, and it has been already pointed out that both effects fit well in the simulation account because different findings are probably due to differences in timing (e.g., Bergen et al., 2007; Borregine & Kaschak, 2006; De Vega et al., 2004; for modeling work in this direction, see Chersi et al., 2010). Much evidence in the literature confirms this idea. For example, it has been clearly confirmed by the kinematic study of Boulenger et al. (2006) that the timing between the planning of a movement and word processing is able to modulate the direction of a motor simulation effect. Participants were asked to perform a reaching movement in two experiments. In the first, the reaching movement was performed concurrently to a visual lexical decision task on action verbs, action nouns or pseudo-words, as the target-letter string appeared when the hand released a touch-pad where was placed before each trial. In the second experiment, the lexical decision was made prior to movement onset, as the letter string represented in this case the go-signal to start the execution of the movement. The results showed that processing of action verbs, compared to nouns, interfered with the simultaneous execution of the reaching movement. By contrast, processing action verbs prior to movement onset favored the subsequent reaching movement execution. The interference result could be explained also referring to common coding theories (e.g., TEC, Hommel et al., 2001, see par. 2.2): comprehending an action verb (e.g., PAINT), the features that code the execution and outcome of the action are activated, so they cannot be used in the planning of the motor response, and this results in an interference effect.

In the literature, evidence of perceptual interferences has been found as well. For example, in an early study (Craver Lemley & Reeves, 1992) interference was found when people simultaneously performed mental visual imagery and a visual perception task, with similar effects being observed also without explicit imagery instructions. Interference effects were found also in the study by Estes et al. (2008), where the authors presented in the center of the screen adjective-name pairs that referred to an object with a typical vertical location (e.g., COWBOY HAT/COWBOY BOOT). After the presentation of this prime, a target-letter appeared at the top or the bottom of the screen, and participants were asked to identify the letter as quickly as possible. The results showed that the identification of the letter was slower when the position of the letter matched the typical location of the prime's referent.

In the literature, findings supporting the idea of a simulation during word processing has been collected, as well as by visual presentation, with auditory stimuli. For example, a priming effect between names of objects which afford similar motor programs to be used has been observed by Myung et al. (2006). Participants were acoustically presented with a series of prime-target pairs, with the task of making a lexical decision on the target-word pressing two buttons (correct/incorrect) on a response box. The results showed significantly faster decisions if the target-word (e.g., TYPEWRITER) followed a prime related with the target in terms of the manner of manipulation (e.g., PIANO) than an unrelated one (e.g., BLANKET). As objects affording similar actions tend to be similar in shape such that priming might be explained on the base of visual similarity, to rule out this possible account the authors selected words denoting objects that were rated as having high similarity in terms of manipulation, but low visual similarity. Thus, they assumed that the words automatically evoked object knowledge, as it is represented in pattern of potentials action (Glenberg & Robertson, 2000), which mediated the effect from the prime to the target. This finding clearly indicates that sensory-based functional knowledge is an intrinsic part of the semantic and lexical representation of objects, and that this action knowledge can be activated by simulating manipulation features also in implicit task (e.g., Tucker & Ellis, 1998, see par. 2.3).

To close with the behavioural evidence on simulation during language processing, it is worth reporting that Pecher et al. (2004) have demonstrated that verbal language is able to select different modal information related to the same concept. Indeed, according to the perceptual symbols theory (Barsalou, 1999), the sensory-motor simulations which underlie the representation of concepts are componential, so they have to vary with the context in which the concept is presented. In this study, concept names (e.g., APPLE) were presented in a property verification task with two different property in different trials. The two properties were either from a same perceptual modality (e.g., GREEN - SHINY) or from different ones (e.g., TART - SHINY). The results showed that response times and error rates for the second presentation of the name were higher if the properties were from different modalities than if they were from the same modality (even if there was a gap of several trials between the first and second presentation). This finding demonstrates that a simulation of the word's referent is elicited automatically and that this mental representation is affected by recent experiences with the concept. Indeed, in direct dependence of the context imposed by the priming adjective, an activation of a neural pattern corresponding to the modality specific domain occurred. This earlier activation could be compatible with the following, so the same circuits were used, but in the case of incompatibility the switching from a brain system to another explains the modality switching cost in terms of response latencies.

On the whole, the reviewed evidence confirms the idea that (perceptual and motor) simulation occurs automatically to guide comprehension during the exposure to words. The simulation appears to be rather specific: it depends from different kinds of modality as on the time of the exposition to the linguistic cue; it is sensitive to a number of objectual properties presented in the real world by their referents and to their affordances; it represents spatial configuration of the referent's parts; it reproduces the typical location in which referents are experienced in everyday life.

In the literature similar behavioural findings have been reported also for sentence comprehension. For example, it has already been addressed that the simulation during sentence comprehension occurs in a figurative way which is highly sensitive to subtle modulations of objectual properties, like shape or orientation. According to the assumption that a symbol is analogically, not arbitrary, linked to its referent (see par. 1.1, 1.2 and 1.3), perceptual symbol theories (e.g., Barsalou, 1999, 2008) predict that "the complete representation of an object, called a simulation, should reflect physical characteristics of the object" (Stanfield & Zwaan, 2001, p. 153). In order to test this hypothesis, Zwaan et al. (2002) presented participants with a prime-sentence describing a same object/animal (e.g., EAGLE) in two different locations that implied a change of shape state that was only implied by the sentence (e.g., THE RANGER SAW THE EAGLE IN THE NEST vs. IN THE SKY). The sentence was followed by a target-picture (i.e., a line drawing) that participants had to judge as displaying or not the item mentioned in the sentence. The results of the recognition task showed that response latencies were shorter in the case in which the item shape in the picture was compatible with the state of the shape implied by the sentence (i.e., EAGLE IN THE SKY / eagle drawn with open wings). To notice that an amodal account of cognition could not explain why recognition latencies were different between the conditions. In order to control for the task-dependency of the effect, in a following experiment the authors simply required participants to name the item in the target-picture. Indeed, the naming task arguably provides a stronger test of the perceptual simulation hypothesis because, unlike a recognition task, it does not require an explicit comparison between the sentence and the picture. The results showed the same advantage in term of mean latencies for the congruent condition, confirming that during the comprehension of a sentence a perceptual representation of its meaning is automatically activated, even when the task demands does not request it. The results also suggested that the context imposed to the simulation by the situation described in the sentence has a guiding role in building these representations. Thus, as the linguistic simulation is context-driven it is characterized by a certain flexibility. It's worth noting here that Spivey & Geng (2001), using the eye-tracking technique, have probably unfolded oculomotor patterns subtended to this kind of results. The eye movements of participants were recorded while they were instructed to simply listen to visual descriptions of objects, or when they had to retrieve properties of objects previously shown to them. The data showed that during both tasks participants spontaneously looked at particular regions (in a blank space) in a systematic iconic fashion in relation to the object's spatial configuration that was implied in the description/explicitly retrieved. The authors explained the results pointing out that "a mental search of internal memory is accompanied by an oculomotor search of external space" (Spivey & Geng, 2001, p. 235). Moreover, the results were obtained in the first task using only auditory linguistic stimuli with no execution of a response, what seems to reproduce an everyday situation of language comprehension more ecologically than other experimental settings and paradigms.

The dynamicity and flexibility of language-triggered simulations has been highlighted also in a study of Stanfield & Zwaan (2001) testing the hypothesis that, while simulating, people do not only mentally represent shape and object spatial configuration, but also the orientation of an object as it is implied in verbal descriptions. In this case, participants read a prime-sentence which implicitly suggested a horizontal/vertical orientation for an object (e.g., HE HAMMERED THE NAIL INTO THE WALL vs. INTO THE FLOOR), followed by target-pictures of the mentioned object that could be in both horizontal or vertical orientation. Similarly to the previous experiment, recognition times to pictures matching the orientation of the object implied by the sentence (i.e., HE HAMMERED THE NAIL INTO THE WALL / horizontal nail) were shorter than in the orientation mismatch condition (i.e., HE HAMMERED THE NAIL INTO THE WALL / vertical nail). In a similar setting, Connell & Lynott (2009) examined how people represents implied perceptual information about color during language comprehension, performing a semantic Stroop task that tested both object-typical and context-implied color information activation during sentence reading. Participants were presented with a color-associated target-word (such as TOMATO in typical red, atypical green, or unrelated brown ink) after having read a prime-sentence implying either typical or atypical color for the object (e.g., JANE TASTED THE TOMATO WHEN IT WAS READY TO EAT or BEFORE IT WAS READY TO EAT). Results observed an advantage in naming the target-word color both when its ink color was typical for that object (e.g., TOMATO in red ink) and when it matched the color implied by the previous sentence (e.g., TOMATO in green ink following JANE TASTED THE TOMATO BEFORE IT WAS READY TO EAT). These findings suggest that the simulation tends to follow the default configuration of a certain scenario (e.g., seeing red tomatoes) if it is the most usual in everyday experience. However, when subject are pushed in unusual contexts (in this case by the prime-sentence description) it can be elicited the representation of scenario-specific perceptual information that is compatible with the current situation.

Other scholar demonstrated that also the representation of motor events is affected by language comprehension. For example, Zwaan et al. (2004) presented acoustically to participants sentences suggesting a movement close to or away from the body (e.g., THE

SHORTSTOP HURLED THE SOFTBALL TO YOU). Then two pictures of a ball differing in size were presented on the screen, such as they suggested a movement in direction from the observer (smaller ball before bigger ball) or not (bigger ball before smaller ball). The task was deciding with bimanual responses if the two pictures displayed the same object or not. The results showed that when the movement implied by the sentence and the movement suggested by the picture of the ball were compatible response latencies were shorter. This indicated that participants activated a motor simulation of the sentence contents that influenced the visual representation of the motor event. Moreover, these results were obtained in a task were the meaning of the sentences was totally irrelevant.

In the same direction, the results of Borghi et al. (2004) demonstrated the sensitiveness of the linguistic content simulation to different perspectives on related scenes. Participants read a prime-sentence describing a location/object from inside (e.g., YOU ARE EATING IN A RESTAURANT), outside (e.g., YOU ARE WAITING OUTSIDE A RESTAURANT) or mixed (e.g., YOU ARE WALKING TOWARD AND ENTERING A RESTAURANT) point of views. The prime-sentence was followed by a word (e.g., TABLE, usually found inside a restaurant, or SIGN, usually found outside a restaurant) that participants had to judge as referring or not to a part of the location/object. Response latencies were faster when the word referred to element more typically available in the perspective implied by the sentence (e.g., YOU ARE EATING IN A RESTAURANT / TABLE). The results of the part verification task also showed an effect of location: participants were faster if the target-word referred to a part that was near than far within the perspective (e.g., YOU ARE EATING IN A RESTAURANT / TABLE faster than KITCHEN). Importantly, the authors controlled the semantic-relatedness of primes and targets in Latent Semantic Analysis (LSA, Landauer & Dumais, 1997, see par. 1.1), ruling out a propositional account for their data. Thus, the results are consistent with the previously reviewed findings, suggesting that the perspectives implied by the sentences guided the organization of the information feeding in a mental simulation. It is again the context what selected the necessary conceptual knowledge to build an appropriate representation of the linguistic meaning. In a further experiment, the authors presented prime-sentences describing an orientated object (e.g., THERE IS A DOLL UPRIGHT IN FRONT OF YOU). Participants was asked to judge if the target-word presented after the sentence referred or not to part of the object mentioned by the sentence, giving their response manually on a vertical response box. The results showed shorter latencies in case of a compatibility between the response movement (yes-up/down) and the part location (upper/lower), extending to action the findings about the iconicity effect found by Zwaan & Yaxley (2003) on perception.

In the literature, many scholars have reported similar results also for sentences explicitly describing actions. For example, Glenberg & Kaschak (2002) asked participants to judge the sensibility of imperative sentences (e.g., OPEN/CLOSE THE DRAWER) moving their hand from a starting button to response buttons that could be either close or far away in respect to their body. So, in the case of a sentence as OPEN THE DRAWER, the movement performed by the subject to respond would be compatible with the described action if the hand goes close to the body, while responding away from the body would serve as incompatible condition. The finding showed the predicted action-sentence compatibility effect (ACE): action-compatible responses were faster than action-incompatible ones. The effect was observed not only for imperative sentences, but also for descriptive sentences (e.g., MIKE HANDED YOU THE PIZZA) and, moreover, for abstract transfer sentences (e.g., LIZ TOLD YOU THE STORY, see par. 3.5 for further discussion). In the same direction, Scorolli & Borghi (2007) confirmed the effector-specificity of the motor simulation. They presented sentences that referred to actions performed using hands (e.g., UNWRAP THE CANDY), mouth (e.g., SUCK THE SWEET) or feet (e.g., KICK THE BALL) in a lexical decision task where participants were instructed to respond vocally/pressing a pedal in case of correct responses. Response latencies were shorter for mouth-sentences than for hand-sentences with vocal responses, while an advantage of foot sentences over hand-sentences was found in pedal responses. Thus, the results suggested again that the simulation of the sentence meaning is quite detailed, as it is sensible to the effector implied by the action.

On the whole, the results on sentences provide evidence of the generality of the simulation triggered by action/object-related language comprehension, showing simulations as arising during the whole sentence comprehension, rather than merely during lexical access. Thus, following the evidence reviewed so far, it is plausible that there is an action (or object) specific sensory-motor simulation elicited by individual words (or by words combinations), as well as a more general simulation evoked by the whole linguistic construction. Presumably, the ACE reported in Glenberg & Kaschak (2002) reflects a combination of both these aspects - at least in the concrete sentences, there is probably only the latter in the more abstract sentences (see also Glenberg et al., 2008, reviewed in par. 3.5). Indeed, if we take for example a word such as the verb "open", by itself it does not necessarily imply moving a hand horizontally toward the body: "one can open one's eyes, open a sunroof, open a bank account, or open a lead in a race" (Fischer & Zwaan, 2008, p. 842), none of which necessarily involves such a movement. In the same direction, a name as "drawer" does not necessary imply such action without further context, and may imply many other actions across different situations (as in for example in "put the shirts in the drawer"). In other words, these findings suggest that perceptual and motor simulation might occur when all available information is integrated (Fischer & Zwaan, 2008). Furthermore, they show that simulation (and so the resonance by which simulation is carried on at the neural level, see par. 2.4 and par. 2.6) is associated with linguistic content in a very precise manner, showing its high degree of specificity and flexibility in dependence of the current context (even if only implied).

3.3. Neuroscientific evidence of motor resonance during word and sentence comprehension

In the previous paragraph, the evidence reviewed suggested that the simulation processes are tied to language in a very precise manner. The literature of neuroscientific research on both words and sentences showed that a resonance mechanism (see par. 1.4., 2.4 and 2.6) is at work during language processing supporting this conclusion as well. In the last years compelling evidence has been produced on the role of motor resonance in language comprehension, for example during the exposure to action verbs and names of tools. It has been demonstrated that naming/describing tools in comparison to animals selectively activates the left middle temporal gyrus and the left premotor cortex (regions associated to action), which are active also during imagery of interaction with the same items (e.g., Martin et al., 1996). In the same direction, exposition to tool words and action verbs elicits a fronto-central cortical activation that is not found during exposition to words not associated with action (Preissl, Pulvermüller, Lutzenberger & Birbaumer, 1995; Pulvermüller, Lutzenberger & Preissl, 1999). However, considering that the hypothesis held as true here is that language recruits the same sensory-motor areas as for interaction with objects and real action, it behooves that brain activation should display the very same topography for processing language as for processing objects, affordances, actions etc. (Jirak et al., 2010). To give an example, neuro-imaging research indicated that regions of the fusiform gyrus are active during color perception, and that in the retrieval of color knowledge this same color perception system is involved in the posterior temporal cortex. Thus, it is expected that an analogous neural circuit would be activated by color knowledge retrieval triggered exclusively by linguistic stimuli. And it is exactly what has been observed in the literature. For example, Simmons et al. (2007) asked participants to perform two tasks while undergoing fMRI. First, participants performed a property verification task in which they had to judge whether a named color/motor property was true for a named object (e.g., TAXI-yellow, HAIR-combed). Then, they performed a color perception task, and a region of the left fusiform gyrus - found as said as highly responsive during color perception - showed a greater activity during color retrieving than motor property knowledge retrieving. These results confirmed that conceptual knowledge is grounded in modality-specific systems (e.g., Barsalou, 1999, see par. 1.3 and 3.2), and provided direct evidence for an overlap in the neural bases of color perception and stored information about object-associated color (similar results in Hsu et al., 2011; Yee et al., 2012).

In the same direction, a selective activation of different cortical areas in dependence of the observation to different kinds of actions has been observed, revealing a somatotopic organization for the human brain. For example, an fMRI study by Buccino et al. (2001) found that during exposition to videos of actions involving as effector the mouth, the hand or the foot, the same neural areas that map the first-person execution of those actions were activated. Under the embodiment view, it is expected to observe similar brain patterns also in the resonance mechanism triggered by action-language exposition, and this is exactly what has been found recently by studies on action words and verbs. Indeed, much evidence of a somatotopic activation of the premotor cortices has been collected, and with different techniques (e.g., TMS, fMRI, MEG, EEG). For example, Pulvermüller (2005) recorded with TMS neurophysiological (and behavioural) responses in a lexical decision task to three kinds of verbs, referring respectively to actions performed with the face (e.g., LICK), arms (e.g., PICK) or legs (e.g., KICK). The results showed the predicted somatotopic organization in the corticospinal activity patterns: different neural sites in topographical patterns were observed during the exposition to the different kinds of verb (starting at 250 ms after stimulus onset). In addition, EEG-recordings confirmed activations of the effector-specific motor resonance as occurring quite early, that is before than 200 ms from word onset (Pulvermüller, 1999).

Similar findings on the somatotopic organization of the motor cortex were observed by Hauk et al. (2004) using action words in a magnetic resonance setting. Again, processing language semantically related to hand and foot actions produced different brain patterns, specifically the middle frontal and precentral gyrus for words referring to hand actions, and areas in the dorsal premotor cortex for words referring to foot actions. Several other studies have identified body-part specificity in the topography of premotor cortices (but not in other motor areas). For example, Willems et al. (2010) using fMRI found in a lexical decision task a selective activation of the left premotor cortex for right-handers and of the right premotor cortex for left-handers during the processing of verbs denoting manual actions. This result suggests that the simulation evoked during language processing is not only sensitive to somatotopy, but also to inter-individual body differences. However, if language comprehension seemed to be able to activate only the premotor cortex, while the authors investigated the neural substrate of the same manual verbs during mental imagery, the activation of both the motor and the premotor cortex were observed.

Further evidence on the activation of the premotor areas during the exposure to action verbs have been collected using high-density MEG. For example, Pulvermüller et al. (2005) presented acoustically words referring to actions involving the face or the leg to participants engaged in a distractor task. Different neural patterns in the premotor cortex were observed in relation to the different kinds of stimuli: listening to face-related verbs resulted in stronger activation of inferior fronto-central areas in respect to leg-related verbs, whereas listening to leg-related verbs resulted in stronger activation of superior central areas in respect to face-related verbs. Furthermore, these activations occurred within 170 ms after stimulus onset, so it is difficult that other strategic factors could contribute to the results. As noted elsewhere, the results of this study clearly showed that “meaning access in action word recognition is an early automatic process reflected by spatiotemporal signatures of word-evoked activity”

(Fischer & Zwaan, 2008, p. 839). A related study from the same lab (Pulvermüller, Hauk et al., 2005) provided support for the idea that as motor resonance is automatically evoked by words, also the reverse process should run as well. The guiding hypothesis was that the identification of these action words might be favored when the neural regions subtended to the actions they denote are directly stimulated. TMS was applied over left hemisphere motor areas while right-handed subjects performed a lexical decision on action-related words, hence mean latencies to leg-action words (e.g., KICK) were compared to hand-action words (e.g., PICK). Participants responded to meaningful word with an energetic lip movement, chosen to minimize interference between semantic and motor planning processes. The results showed the expected interaction between stimulation sites and word type. Indeed, TMS of hand/leg areas modulated the processing of hand/leg related words selectively: stimulation on hand areas led to faster hand than leg-word responses, whereas when leg areas were stimulated a reverse effect showed shorter lexical decisions latencies for leg than hand-words. Such effects related to word types disappeared during sham or right hemisphere TMS stimulation. Thus, the influence of activations in motor and premotor areas over the processing of different kinds of words semantically related to hand or leg actions confirmed a somatotopic mapping for linguistic stimuli, providing clear evidence that language and motor functions interact in the processing of meaningful information about language and action. As the effects emerged only for left hemisphere stimulations, these results fitted well in the typical language dominance observed in the literature for this hemisphere. Furthermore, they demonstrated that if accessing action-related words evokes motor resonance, this process is also bidirectional: lexical access of this kind of words is favored by motor resonance. This bidirectionality is a fundamental assumption for several theories inside the embodied framework (e.g., TEC, Hommel et al., 2001, discussed in par. 2.2).

The neuroscientific literature reports also several findings on sentence processing which confirmed that a resonance mechanism is at play during language exposition, and that the resonance connects simulations and linguistic contents in a very fine way. For example, topographic differences in the brain activity patterns have been observed by Tettamanti et al. (2005) during the exposition to sentences denoting mouth, hand or leg actions. Abstract sentences with analogous syntactic structure were used as control stimuli. The action sentences, confirming the somatotopic organization observed in the previously reviewed studies on words, activated selectively the same regions involved in the real performance of the action they referred to. Specifically, hand action activated the left precentral gyrus, the posterior intraparietal sulcus and the left posterior inferior temporal area, while leg activity was identified in the left dorsal premotor and left intraparietal sulcus. The detection of a bilateral pattern in the posterior cingulate showed a clearly different activation when processing abstract sentences. Thus, the results displayed activity in a fronto-parietal circuit with a temporal participation in the left hemisphere. In addition, activity in Broca's area was detected independently from any effector-specificity, implying a special role in language processing that seemed to go beyond the one usually attributed to it. Similar findings were observed by Buccino et al. (2005) using TMS and a behavioural paradigm. By means of single-pulse TMS, either the hand or the foot/leg motor area in the left hemisphere were stimulated in distinct sessions, while participants were acoustically presented with sentences describing hand or foot actions (abstract content sentences were used as control stimuli). Motor evoked potentials were recorded from both hand and foot muscles. The results indicated that MEPs from the hand muscles were selectively modulated by hand-action sentences, while MEPs from foot muscles by foot-action sentences. This modulation took the form of a decreasing amplitude in the recorded MEPs. In the behavioural task, participants responded with the hand or the foot while listening the same pool of sentences on hand/foot actions and abstract situations. In line with the psychophysical results, manual responses were inibhited by hand-related sentences and foot responses were slower for foot-related sentences, in comparison to abstract sentences processing. This evidence is in line with a number of results obtained with behavioural tasks (Scorolli & Borghi, 2007; Borghi & Scorolli, 2009) that showed the simulation activated during combinations of nouns and verbs as sensitive to the compatibility of the effector implied by the sentence and the one used to execute the response (see par. 3.2).

In a similar direction, another fMRI study observed that the same brain areas that were activated by action observation also mapped sentence comprehension in an effector-specific way (Aziz-Zadeh et al., 2006). Indeed, regions in the premotor areas that were active during the observation of hand actions were also active during the comprehension of sentences denoting actions performed with the hands. In the literature there is also clear evidence that the referents' affordances have a direct influence on the comprehension of sentences describing object in space. For example, in a visual-world paradigm study (Chambers et al., 2004) participants were presented acoustically with sentences describing spatial scenes during the observation of a display composed by elements connected to the sentences, while holding or not a tool in their hand. The fixations on the visual scene were monitored, and the results showed it was modulated by the linguistic input in interaction with the holding of the tool. Indeed, holding the tool affected the syntactic parsing in a way that changed the amount of time spent looking at (possible) goal locations that is likely/unlikely under that specific parse in respect to an alternative one. This research showed that motor resonance occurs very rapidly during comprehension, substantially before the associated linguistic constituent has been fully processed, even if it did not allow us/researchers to affirm directly that motor resonance occurs during every-day language comprehension.

There are also studies, however, the conclusions of which seem to be at odd with those of the behavioural and neural studies discussed so far. For example Postle et al. (2008) used fMRI on primary and premotor cortices while presenting verbs denoting hand, foot or mouth actions, concrete nouns not related to action/body parts, non-words, and characters strings (with verbs and nouns matched for imageability). In the data there was no evidence of a somatotopic organization for action words, but the pre-supplementary motor area presented a different pattern for foot verbs in respect to nouns, so it is possible that this region had both a cognitive and motor role. As the main novelty of this study, in comparison to the literature reviewed above, was the use of cytoarchitectonical probability maps, these results suggested that "studies on somatotopy connected to word meaning extraction should be also related to cytoarchitectural information and functional criteria, in order to correctly interpret activation distibution as somatotopy" (Jirak et al., 2010, p. 713).

In the neuroscientific literature, somatotopy has been connected also to the involvement in language processes of the mirror neuron system, as they both rely on the theoretical principle that assumes language as sustained by the same neural circuits of perception and action. Indeed, canonical and mirror neurons (see par. 1.4, 2.4 and 2.6) are thought to be the neural basis of the simulation activated during language referring to objects and actions (e.g., Gallese, 2008; Rizzolatti & Craighero, 2004). Originally identified in macaque's area F5 (ventral premotor cortex), mirror cells are a class of visuo-motor neurons that map first-person execution of action, as well as the observation of others performing actions. In almost a total absence of investigations on humans at the single-neuron level, the area which is regarded as a homolog of monkeys' F5 is the Broca's region (inferior frontal region of the human cortex). The Broca's area, originally thought to be a pure speech processing area, is now recognized by many scholars as containing mirror neurons, and regarded to as the action-language linking area (e.g., Binkofski & Buccino, 2004;

Pulvermüller, 2005) at the core of the human mirror neurons system. As it suggest additional circuits linking motor and speech processing, the whole mirror neurons system possibly has an impact on language comprehension. Supporting this idea, for example Glenberg et al. (2008) using TMS found comparable mirror activation patterns by either presenting the typical sound of actions or their verbal description. But, as noticed by Jirak et al. (2010), the activation of the Broca's area seems to differ in typically mirror-activating tasks, like action observation compared to language processing. For example, a series of fMRI studies (Aziz-Zadeh et al., 2006; Aziz-Zadeh, Wilson et al., 2006; Aziz-Zadeh & Damasio, 2008) revealed that the pattern of activation found for actions observation and action-sentences reading did not result in a total overlap. The authors suggested that mirror neurons, as also pointed out elsewhere (e.g., Rizzolatti & Arbib, 1998), are probably not directly involved in the process of language understanding, but have possibly played a fundamental role in the phylogenetic development of language (see also par. 3.6). However, this is not in contradiction with the integrating role attributed to Broca's region in sounds and actions processing. For example, D'Ausilio et al. (2009) reported facilitated perception of a given speech sound when the motor articulator mapping that sound was stimulated with TMS. Thus, it is possible that mirror neurons integrate sounds and actions or even sustain simulation in order to understand linguistic stimuli. However, the conclusion that motor regions closely linked to regions of the mirror neuron system are indispensable in language processing will need further evidence.

The suggestion of being careful in discussing the relation of action, perception and language can also be drawn from of a lesion study which tested both visual and linguistic action understanding of both aphasic patients and control subjects (Saygin et al., 2004). Participants in the experiment would see, for example, a line drawing of a boy licking an ice cream but with no an ice-cream cone depicted in the picture, or sentences fragment like HE LICKED THE with no object as patient (i.e., followed by a blank space). Then, they were presented with pictures that could display the missing target object, or an affordance-related object which invite to the same manual interactions (i.e., bouquet of flowers), or a semantically related distractor (i.e., cake), or a semantically unrelated distractor (i.e., rooster). Participants decided which among these pictures fitted the visual scene/sentence fragment by pressing the button corresponding to the chosen picture, while response latencies and accuracy were measured. The results showed no overall correlation between patients' deficits in visual and linguistic comprehension, suggesting that different brain regions were associated (at least in this task) with the deficits in visual representation than the deficits in linguistic representation. However, correlations inside specific subsets of participants were observed, suggesting that action and language understanding share overlapping neural substrates. Thus, this evidence implies here the emergence of a critical question about the real extent of similarity ascribable to action observation and language comprehension (see the following paragraph). Indeed, it seems that these two processes are not identical, at least with respect to the timing and the manner involved: action observation is much closer to the execution of action than how could be linguistic action description (Fischer & Zwaan, 2008).

To summarize what have been seen so far, several studies showed a great amount of evidence confirming that - although critical views have also been proposed (e.g., Postle et al., 2008) - action/object-related language comprehension elicits a resonance directly connected to somatotopy, and possibly to the mirror neuron system. In general (but also considering the critical views), it can be affirmed that somatotopy of language contents can be consistently identified in premotor regions, conclusion that bolsters embodied accounts even if it cannot be affirmed the same for motor cortices' involvement. Indeed, this might be simply due to the tasks used (very little results have been collected in really ecological settings), or it is possible that the evidence highlights a different role played by motor cortices in the direct execution and observation of actions, in comparison to actions conceptualization and linguistic description (Jirak et al, 2010).

3.4. Objects/actions processing and language comprehension: are there identical simulations?

The largest part of the literature on simulation and resonance reviewed so far has investigated embodied language mostly in order to demonstrate the similarity of the simulation triggered by language in respect to the simulation triggered by objects/actions observation, action preparation or action execution. The evidence is compelling, and has been very valuable for the spreading of embodied and grounded views of cognition and language. However, as it has been highlighted in the last part of the previous paragraph, there might be important differences in the cognitive processes (e.g., understanding the meaning of a certain action) elicited during execution of action, observation of objects/actions, preparation to action, than during language exposition. And it is probably time to better investigate and understand these possible differences. From an epistemological point of view, it is obvious that language cannot be considered identical to execution or observation, at least because "observation and real interaction with objects is always situated in a specific context, whereas during language comprehension we typically have to mentally construct a situation" (Borghi, 2012, p. 127). Thus, assessing to what extent the simulation during language comprehension resembles or differs from the one during action/object observation results to be an important test bed for the explicative power of embodied cognition accounts (my empirical investigation on the differences of affordance and language, the results of which confirmed the supposed differences, is reported in chapter 6).

In order to better understand the differences of linguistic triggered simulation in comparison to simulation during action execution or observation, it is useful to refer to the literature on affordances and language (Borghi, 2012). A starting point to try a clarification of what happens during real interaction with objects and what happens during language comprehension could be the recent proposal of Borghi & Riggio (2009) to make a distinction about specific affordances kinds (see par. 2.3 and 2.5). In their view, stable affordances are affordances related to properties rather constant across contexts, like size, while variable affordances are affordances related to mutable properties, like orientation. A special case are represented by canonical affordances: a canonical affordance is the more frequent configuration of a variable affordance usually depending on object use, such as typical object orientation (e.g., for a right-hander a beer mug upright with the handle on the right, ready to be filled and grabbed), the information of which might be useful to store along with stable affordances. Given this situation, there are several possibilities about language.

The first possibility is that words are not grounded, not linked to objects' perceptual and motor characteristics; in this case the existence of different kind of affordances related to the same object/action would be not problematic. Indeed words would be simply understood in terms of semantic relatedness to other words (e.g., Landauer & Dumais, 1997, see par. 1.1). I hope I have given enough examples of the fallacy of this position in the course of this dissertation, at least enough to avoid discussing it again now (see chapter 1 and par. 3.2). A second possibility is about an identical activation of sensory-motor systems in the situation in which we are interacting with a beer mug, as in the situation in which the words "beer mug" are read on a menu, listened from the radio or said to a friend. In this view, words and sentences are directly linked to their referents (as in the Indexical hypothesis, see Glenberg & Robertson, 2000) or to their analogical representations (as in Perceptual symbols theory, see Barsalou, 1999, 2008, 2009). In this view, words denoting objects would evoke perceptual and motor information relative to such objects. Much evidence supporting this idea has been reviewed in this thesis (see par 3.2 and 3.3), confirming that the way we understand language is constrained by the affordances elicited by the objects/actions denoted by word and sentences. However, as recently advanced elsewhere (Borghi, 2012), there is a third possible view about what happens during language processing. Indeed, if it is held as true that language is tied to affordances and simulation is constrained by them, assuming that language comprehension is under the influence of object affordances implies here to be able to indicate what *specific* affordances are referred to. For example, it is obvious that interacting successfully in the real world with beer mugs imply the capability to use previous knowledge about object-based interaction, especially experiences with beer mugs. In the case of linguistic information, to understand what it is meant by the words "beer mug", in absence of a general and still representation of the concept, we might at least retrieve perceptual information about the typical shape of beer mugs and the related information about the grip that is usually necessary to interact with an object of that shape and parts configuration. Indeed, a rather canonical property of our interaction with beer mugs is to use a power grip on the handle to manipulate it and to start any kind of interaction with it.

In the direction to clarify which affordances are activated during the exposition to action sentences, Borghi & Riggio (2009) used the recognition task described above for Stanfield & Zwaan (2001). In their setting, participants were visually presented with prime-sentences denoting observation of an object or action on the same object (e.g., LOOK AT vs. GRASP THE BRUSH). After the cut off at 400 ms, the sentence was replaced by a picture of an object that participants were required to judge as named or not in the prime-sentence; responses were produced by manual key-presses. The target-objects elicited power/precision grips (e.g., brush vs. pencil) and could be represented, in respect of the participants' perspective, in their canonical orientation or not (i.e., the affording part in the lower vs. the higher part of the screen). The results showed that action verbs (in imperative forms) were processed faster than observation verbs even if the verb frequency was controlled, indicating that action sentences induced a motor preparation. Additionally, the objects in canonical orientation produced shorter response latencies in respect of the reversed presentation, underlying - rather than a visual - a motor nature for the effect, relative to objects' orientation for use (upright vs. reversed) and not to object visual orientation (vertical vs. horizontal). However, the most important results concerned mismatching trials, when the object in the prime-sentence and the target-picture did not correspond; in this case, when the grip of the sentence's referent and the grip elicited by the picture were compatible (e.g., GRASP THE BRUSH followed by a picture of a pencil), mean latencies were longer for action than observation verbs. This interference was probably due to the unavailability of the motor system, still occupied in the production of the linguistically elicited (mismatching) simulation. The inhibitory effect observed can be explained in terms of common coding (e.g., Hommel et al., 2001): if an event code is activated from two different sources (i.e., the linguistic task and the motor response), an inhibitory process may have place. Thus, Borghi & Riggio's results suggested first, that when we read an action sentence a motor preparation of proper actions with the described object take place, and second, that object-names favor the appropriate grasping of the referent. This means that the representation of the concept of an object follows the guide of its referent size and shape, along with the grip they evoke.

On the whole, these findings permit a better characterization of the simulation occurring during language understanding, indicating which kinds of affordances are evoked by action-related sentences reading. Indeed, analogously to the simulation during object observation, linguistic triggered simulation re-enacts previous experiences preparing us to future action, thus sustaining prediction. Specifically, linguistic simulation of action language seems to build a motor prototype based primarily on stable affordances along with affordances related to objects' grip and canonical use orientation. This confirms that language understanding is directly tied to object affordances, rather than being merely implemented by associative relations between words. However, the simulation during sentence reading seemed to be slightly different in comparison to the simulation triggered by objects observation. Indeed, linguistic simulation seems to consider only stable and canonical affordances, favoring the most frequently experienced object properties, which have the higher probability to be encountered. Thus, language evokes a rather detailed simulation, but for this simulation is not possible to contain variable properties of objects. On the contrary, variable properties are taken into account online in object observation (e.g., Tucker & Ellis, 1998), similarly to what happens during real interaction with objects.

Other scholars have pointed out that, as we can interact with objects in specific ways in order to reach for distinct goals, different affordances might be also related to objects manipulation and objects use (see par. 2.5). Tools are the object kind with perfectly represent this situation, as they usually can evoke distinct skilled actions that are consistent to their use as well as grasping actions that are adapted to their structure (e.g., Bub et al., 2008; Creem & Lee, 2005; Jax & Buxbaum, 2010; Natraj et al., 2013; see also chapters 4 and 5 of this thesis). For example, the appropriate way to move a knife is to grasp it with a precision grip, whereas to cut something with a knife a power grip is required. At the neural level, it has been proposed that affordances related to object use are represented more ventrally, with those related to manipulation represented dorsally (e.g., Young, 2006; Buxbaum & Kalénine, 2010). Interestingly, Jax & Buxbaum (2010) showed a behavioural difference for conflict objects (i.e., require different grips for manipulation and use) and not-conflict objects (i.e., manipulated and used with the same grip), and proposed that the intention to act on an object triggers a race-like affordances competition. Considering that functional behaviours require activation of previously stored information, the activation of manipulation responses is faster. Moreover, when a manipulative grasping is performed before using the object there is an interference effect which is not present when use is performed before manipulation. In light of this further evidence, and of the distinction between functional and volumetric affordances, an interesting proposal has been very recently advanced by Borghi (2012) about affordance simulation in language comprehension. Borghi hypothesized that when a word/sentence denoting a manipulable object as a tool is processed, functional information is more likely to be activated than volumetric information. This is what might be observed because “functional actions are more frequently associated with tools than other kinds of actions” (Borghi, 2012, p. 131). Evidence supporting this idea have been already collected. For example, Costantini et al. (2011) presented participants with 3D-pictures of a room which displayed an everyday tool (e.g., bottle) that was located in peripersonal/extrapersonal space. Then, participants were presented with a function/manipulation/observation verb (e.g., DRINK, GRASP, LOOK AT), being instructed to judge if the verb was congruent with the previously presented object in the 3D-scene. Faster responses were observed for verbs denoting function and manipulation than observation, with latencies to manipulation and function verbs being shorter if objects were in the peripersonal than extrapersonal space. Further, in the peripersonal space the fastest responses were observed for function verbs. Thus, these results confirmed that objects are represented in terms of affordances and that the linguistic context flexibly modulates the simulation. Additionally, they show that in the situation in which first-person action usually happens (i.e., peripersonal space), function is the content more likely to be activated. Thus, the activation of affordances might differ for linguistic and non linguistic task, as in the latter online information (e.g., variable affordances), along with manipulation to access the object, is activated firstly, as it has been shown that in real interactions information related to structural characteristics is activated earlier than functional information (e.g., Jax & Buxbaum, 2010). Instead, during action-sentence reading a motor prototype based on function was activated because, especially with tools, functional actions are the more frequently performed (e.g., Borghi & Riggio, 2009; Costantini et al., 2011). Thus, in the case of linguistic tasks the competition of manipulation and use had a different winner, as language processing would activate primarily function information.

On the whole, the findings reviewed in this paragraph indicated that the simulation of linguistic contents re-enacts previous experiences to produce comprehension in light of specific predictions aimed at action planning and execution. Indeed, the simulation put the subject in the most frequently experienced situation, activating a motor prototype of the objects/actions linguistically described. This prototype contains stable and canonical affordances, along with the sensory-motor representation of the object function. This suggest that there is a difference in the activation triggered by language comprehension from both observation and interaction with objects, and that this difference regards in particular the kind of predictions implicitly produced (Borghi, 2012).

3.5. Abstract concepts in embodied language

The literature on simulation and resonance reviewed so far has built an idea of language embodiment and grounding which is mostly based on investigation about words and sentences denoting concrete concepts. The primary reason of this is that distinguishing between abstract and concrete concepts/words is not an easy work. The standard assumption that concrete words index perceivable entities, whereas abstract words refer to entities more detached from sensible experience (e.g., Paivio, 1991; Barsalou et al., 2003), does not solve the problem. For example, if it is agreeable that TO GRASP is a concrete verb, judging noun-verb combination as for example GRASP THE MEANING seem to be not so easy (Aziz-Zadeh, Wilson et al., 2006). Similarly, words denoting social roles as TEACHER might be considered more abstract than words referring to single objects as GLASS, but anyway less abstract than a purely definitional word as NUMBER (Keil, 1989); similarly, basic or subordinate words, like DOG or COCKER, may be considered more concrete than a superordinate word like

ANIMAL (e.g., Borghi et al., 2005). Moreover, words and sentences meaning is determined by the context they are embedded in, which comprehend at its most general level the language/culture in which they are produced, with evidence in the literature suggesting that abstract concepts/words are strongly influenced by sociolinguistic and cultural factors (e.g., Boroditsky, 2000, 2001). These reasons are starting to led scholars to look at the abstract-concrete dimension as a continuum, rather than a clear-cut dichotomy (e.g., Wiemer-Hastings et al., 2001).

If we consider the general explanation proposed by the propositional account and embodiment, it is evident that they either suppose the existence of a single representational format for both concrete and abstract concepts. Indeed, according to propositional theories (see chapter 1) the representation of both concrete and abstract concepts/words is symbolic and amodal (e.g., Fodor, 1975, 1998). On the contrary, according to embodied accounts of cognition (e.g., Barsalou, 1999; Gallese, 2008; Glenberg, 1997; Lakoff & Johnson, 1999), both concrete and abstract concepts/words are grounded in perception, action and introspection systems, being characterized by a (multi)modal nature. However, given this starting point for embodiment the grounding of abstract language remains something unclear, as the consequent question about abstractness that follows from it is: “how can theories that focus on modal simulations explain concepts that do not appear modal?” (Barsalou, 2008, p. 634). Thus, abstract concepts constitute a very problematic issue for the embodied view of cognition, if not the most problematic one.

As embodied theories propose, at a very general theoretic level, that abstract concepts/words are grounded in sensory-motor systems in the same way that concrete ones are, starting from this all-embracing assumption at least three more specific explanations have been advanced (Glenberg et al., 2008; see also Flumini, 2011). The first embodied account of abstract concepts, and probably the most well-known and influential, explains them referring to *metaphors*. Indeed, Conceptual Metaphor Theory (CMT, Lakoff & Johnson, 1980, 1999) has pointed out that our vocabulary about time has metaphoric, spatial roots, with this idea being supported from several investigations in linguistics and psychology (e.g., Boroditsky, 2000; Clark, 1973; Lakoff & Johnson, 1980, 1999; Núñez & Sweetser, 2006; Talmy, 2000). For example, Lakoff & Johnson (1980, 1999) showed the presence of dickers of metaphorical mapping schemata in every-day English language (e.g., DEATH FINALLY CAUGHT HIM could be an example from the personification schema). The authors suggested that these metaphoric linguistic mappings directly observable in speech are something that renders explicit the (pre-linguistic) conceptual mappings subtended to them. Thus, metaphors would be at work at a deep cognitive level, and in the case in which abstract concepts would be referred to, the elicited embodied simulation should follow the guide of stored mappings between abstract and concrete concepts. This basic principle of CMT is supported especially by those empirical findings on conceptual metaphors that discovered the existence of metaphoric mappings which are not explicitly attested in language (for further discussion see Flumini & Santiago, 2013; Santiago et al., 2011). In the last years, one of the most studied example of this kind of mapping has been the TIME IS SPACE metaphor which maps temporal reference onto a horizontal spatial axis (e.g., left-to-right in Western cultures), in place of the explicitly attested in many languages front-back axis (e.g., Sell & Kaschak, 2011; Torralbo et al., 2006; Ulrich et al., 2012). Indeed, in his review of cross-linguistic space-time metaphors, Radden (2004) observed for all the languages considered a total lack of linguistic conventions in speech that referred to time as organized in an horizontal space. However, in every-day experience we are all used to conventional associations of time as flowing from left to right in a horizontal axis, i.e., in written language as in many types of sequences and graphic devices (e.g., comic strips, calendars, etc.); thus, the establishment of a left-right mapping has been referred to the cultural exposure to an habitual reading-writing direction (review in Bonato et al., 2012). Along with this observational evidence, a wide set of response time studies reported interactions between the processing of the temporal reference of words/sentences (and/or the spatial presentation of the stimuli) and a variety of response mappings: lateralized key presses, forward-backward manual movements, vocal responses (e.g., Boroditsky, 2000, 2001; Flumini & Santiago, 2013; Ouellet et al., 2010b; Sell & Kaschack, 2011; Torralbo et al., 2006; Ulrich & Maienborn, 2010; Ulrich et al., 2012). Space-time congruency effects are part of a wider family of compatibility studies which manipulate concrete and abstract dimensions in tasks that require responding only to aspects of the abstract one. In this setting, modulations due to (task-irrelevant) concrete dimensions are often observed on the processing of the (task-relevant) abstract dimension. The resulting metaphoric congruency effect is interpreted as the index of the use of underlying concrete representations to organize the abstract dimension, with many examples of this being furnished in the literature on the so-called SNARC effect (Dehaene et al., 1993) literature. As the experimental results showed an unexpected degree of flexibility in conceptual mappings, nowadays there seems to be a well-motivated support to the idea that conceptual congruency effects could have contextual nature (e.g., Flumini & Santiago, 2013; Torralbo et al., 2006; Santiago et al., 2008; Santiago et al., 2011, 2012). This cognitive flexibility clearly contradicts the classic interpretation of the effects as indexes of stable, automatic long-term memory mappings, opening a question about how and why specific mappings are selected from the semantic memory to be used in the working memory mental space (my empirical investigation on this topic is reported in chapter 8). In the literature there is evidence of different degrees of flexibility depending on the abstract dimension studied, the task and materials used, the kind of mappings which are evaluated. A case considered as the example of strong automatic activation (usually related to its salience for basic human beahviours) concerns the mapping of valence evaluation to approach-avoidance manual responses. Indeed, several results showed faster responses to positive and negative items respectively by pulling and pushing a lever (e.g., Chen & Bargh, 1999). It happened not only when the decision was based on the valence of the stimuli: also when performing a lexical decision task (Wentura et al., 2000) and even a stimulus detection task (Chen & Bargh, 1999), which really minimized the task-relevance of the evaluative dimension. Given this evidence, Freina et al. (2009) investigated if the content of the simulation triggered in such settings would be automatically determined or not. Participants judged words as positive/negative by pressing two buttons, one close to and one far away from their body. Results showed that in the case they had to press the buttons with the open palm of their hands, shorter latencies were found when responding to positive words with the button away from their body, like if they were simulating the reaching of something positive, while the reverse was true for negative words, like if they avoided something negative. The opposite pattern was instead found in the case in which participants responded holding a tennis ball in their hand. The authors suggested that in this case participants simulated keeping the "good" for themselves and throwing the "bad" away from them. This study, in line with the literature, confirmed that processing words with different valence evokes oriented arm/hand movements based on previous approach-avoidance experiences, but it revealed also that an identical flexion/extension movement could be dynamically associated to positive or negative valence in dependence of the overall situation (i.e., hand-posture, action goal). This evidence, describing the simulation as sensitive both to the overall action goal and to fine-grained kinematics aspects (see also Hommel et al., 2001), showed for the metaphoric relation between valence language (abstract dimension) and motor schemata (concrete dimension) a higher flexibility to the one usually attributed to it, indicating that also in this case the context, able to elicit the assignment of different meanings to the very same action, guided the emergence of the effects.

On the whole, flexibility seems to be the commanding rule in abstract metaphors. Furthermore, this view of metaphoric mappings as contributing to the access to abstract meaning by appealing to a conceptualization depending from real (action) experiences have given in the last decades an important boost to the spreading of the embodiment's perspective about language comprehension.

The second explanation of abstract concepts/words grounding follows a strongest version of embodiment, suggesting that both abstract and concrete language would be represented by overlapping activations in the same sensory-motor neural units. The idea is that the same representation of the verb GRASP is activated for YOU GRASP THE HAMMER and for YOU GRASP THE IDEA (e.g., Aziz-Zadeh & Damasio, 2008): literal and abstract meanings would elicit activations of the same neural patterns. This notion that a same action schema is subtended to the understanding of both concrete and abstract contents is widely accepted among linguists that work on signed languages. For example, signs analysis of Al-Sayyid Bedouin Language (ABSL, Sandler et al., 2005) has showed evidence that a hand movement, starting near the body to finish far from it, is used either in the mapping of concrete (e.g., GIVE) and abstract (e.g., INFORM) transfer verbs. In the psychological literature, there is both behavioural and neuroscientific evidence supporting this idea. For example, Glenberg & Kaschak (2002) instructed participants to judge the lexicality of imperative sentences (e.g., OPEN/CLOSE THE DRAWER) moving their hand from a central starting button to response buttons located close to or far away from their body, and reported an action-sentence compatibility effect (ACE, see par. 3.2). The effect was found for concrete sentences as well as for descriptive sentences (e.g., MIKE HANDED YOU THE PIZZA) and abstract transfer sentences (e.g., LIZ TOLD YOU THE STORY). Glenberg et al. (2008) have confirmed this behavioural evidence in a following study which also extended it to the neural level in a TMS setting. The neuropshysiological results confirmed that, relative to no-transfer sentences reading, there was greater modulation of the hand muscles during the reading of abstract (e.g., ARTHUR PRESENTS THE ARGUMENT TO YOU) and concrete transfer sentences (e.g., ANDREA CARRIES THE PIZZA TO YOU). The processing of abstract and concrete words has been detailed at a finer level in the fMRI study of Rüschemeyer et al. (2007). The authors performed a go-no go lexical decision task requiring participants to respond with a key-press to pseudo-words whereas no response was required for the correct items, which were verbs with a motor meaning (e.g., BEAT) or abstract verbs (e.g., GUESS). The authors found first a greater activation (bilateral but higher in the left hemisphere) in areas associated to action as the motor cortices and the somatosensory cortex for simple motor verbs compared to abstract verbs; no difference during processing of motor verbs in respect to abstract verbs was instead observed in the activation of frontal mirror neurons areas, ventral premotor cortex and inferior parietal lobule. Second, while the authors investigated the neural correlates of processing morphologically complex verbs with abstract meanings built on stems with motor meaning (e.g., BEGREIFEN/COMPREHEND) or with abstract meaning (e.g., BEDENKEN/CONSIDER), no evidence for the hypothesized residual effects of the motor stem meaning in the activation pattern was found, whereas complex verbs built on motor stems resulted in increased activation patterns of the right posterior temporal cortex compared with complex verbs built on abstract stems. However, in the fMRI study by Tettamanti et al. (2005, see par. 3.3 above) premotor cortex activations were reported during the exposition to action-related sentences, with the data indicating activations in the motor regions only for concrete sentences denoting action on manipulable objects in comparison to sentences denoting abstract objects. Thus, from this results the representation of verbs seems to be highly dependent on the interactions with the semantic context, if it is present. In this direction, a more recent fMRI study (Raposo et al., 2009) investigated this supposed flexibility in a set of fMRI studies in which participants passively listened to arm/leg-related action verbs that were presented in isolation (e.g., KICK), in literal sentences (e.g KICK THE BALL) and idiomatic sentences (e.g., KICK THE BUCKET). The results showed significant activations in motor regions when action verbs were presented in isolation and, to a smaller extent, in sentences describing literal contexts. In the case in which the very same verbs were presented in idiomatic contexts activation was observed in fronto-temporal regions, associated with language processing, but not in motor and premotor cortices. Thus, these findings clarify the previously reviewed by showing that the nature of the semantic context determines the degree to which alternative senses and particularly relevant features are processed when a word is heard. This suggests that motor representations are context-dependent, rather than automatic and invariable.

On the whole, from this evidence clearly follows that a sensory-motor grounding for abstract concepts/words has been confirmed only in rather specific domains, while much and more extensive evidence is still needed before considering its demonstration conclusive.

The third account of the grounding of abstract concepts/words has been advanced by recent proposals which suggest that multiple kinds of representations underlie abstract and concrete concepts (review in Borghi & Pecher, 2011). In this view, abstract language is characterized by specific type of properties, as it re-enacts first-person and introspective situations (or *qualia*, e.g., Edelman & Tononi, 2000), at least more frequently than what happens with concrete language (e.g., Barsalou, 1999, 2008; Barsalou & Wiemer-Hastings, 2005; Pecher et al., 2011). For example, Wiemer-Hastings & Xu (2005) used a feature generation task to systematically compare the content of abstract and concrete concepts. Participants were required to list characteristics of the indicated concepts (i.e., item properties) or of their usual context (i.e., context properties). Abstract concepts elicited a greater number of properties expressing subjective experiences than concrete concepts, which mostly evoked intrinsic item properties. Similarly, the situation components generated for abstract and concrete concepts differed not in number, but in kind, as abstract concepts were mainly related to social aspects of situations. In general, the properties produced were less specific for abstract concepts in respect to concrete ones. Hence, in this study abstractness seems to emerge as a function of multiple factors, both qualitative and quantitative. This pattern of results might fit in the Language and Situated Simulation theory (LASS, Barsalou et al., 2008). According to LASS, linguistic material interact continuously and in many different ways with situated simulation processes, so that mixed contributions of the two systems are selectively subtended to a wide variety of mappings and tasks. The theory suggests that left-hemisphere language areas by which the brain's linguistic system is composed, in particular the inferior frontal gyrus (or Broca's area), are mostly involved during superficial linguistic processing, whereas during deeper conceptual processing also the simulation system is engaged, especially the bilateral posterior areas that are associated to episodic memory and mental imagery (Scorolli et al., 2011). In contrast to the LASS theory, the very recent Words As social Tools proposal (WAT, e.g., Borghi & Cimatti, 2010) suggests that the linguistic system does not merely sustains a superficial processing. Indeed, words cannot be described as simple indexes of something else, but - in line with well-known philosophical views of language (e.g., Austin, 1962; Merleau-Ponty, 1962; Wittgenstein, 1953) - they are better conceived as tools that allow us specific behaviours and operations in the world. In addition, the WAT theory formulates predictions about the grounding of concrete and abstract concepts/words that are more detailed than the ones advanced by LASS. According to WAT, abstract words' meaning is more strongly related to the every-day experience of being exposed to language and immersed in socio-cultural contexts than concrete words' meaning. Thus, the difference between abstract and concrete words is referred to different modalities of acquisition (MoA, Wauters et al., 2003): perceptual, linguistic and mixed. The basic idea of WAT is that MoA rates, in correlation to age of acquisition, concreteness and imageability, gradually change during ontogenetic development. In the first steps of acquisition the grounding is mostly perceptual, whereas going on with the development become more and more linguistic, until learning presents mainly linguistic characteristics. This is why abstract words are typically learnt after the first steps of development: the acquisition of abstract words, due to their complexity, require elaborate linguistic explanations in enduring and repeated social interactions. In contrast, the acquisition of concrete words in young children seems to be effortless, often requiring only a single episode of exposition to the word in the context of simultaneously pointing to its referent (e.g., Pulvermüller, 2012). WAT does not neglect that for the acquisition and representation of both concrete and abstract words meanings either sensory-motor and linguistic experience are crucial. The theory simply points out that people rely more on language to understand and build abstract words' meaning, as they do not have a specific entity as referent, while sensory-motor experience is usually enough to fulfill and give sense to concrete words. This is true also for the different degree of complexity of abstract and concrete words' meaning (as confirmed by the results of my empirical investigation on WAT predictions, reported in chapter 7). This was confirmed in the study on WAT by Scorolli et al. (2011) in which, starting from the assumption that the concrete-abstract dimension represents a continuum, a concrete/abstract verb was combined with both a concrete and an abstract noun (e.g., GRASP THE FLOWER vs. THE CONCEPT; DESCRIBE THE FLOWER vs. THE CONCEPT). In order to disentangle semantic meaning from the words' grammatical class, the authors focused on two syntactically different languages, German and Italian. Results showed that compatible noun-verb combinations (i.e., concrete verb and noun, abstract verb and noun) were processed faster than mixed noun-verb combinations. Additionally, if in the mixed combination a concrete word preceded an abstract one, participants were faster regardless of the grammatical class and the spoken language. This is probably due to the peculiar MoA of abstract words, as they are acquired more in a linguistic than perceptual manner. Overall, these findings support the idea that abstract and concrete words are processed selectively in parallel systems, that is abstract words (more) in the linguistic system and concrete ones (more) in the sensory-motor system, as indicated by the lower processing costs when the combination was confined within one system. Thus, results confirmed WAT predictions: if for abstract words is crucial the role of both sensory-motor and linguistic experience, in respect to concrete ones there is a more marked contribution of language. Scorolli et al. (2012) confirmed this behavioural evidence and extended it to the psychophysical level using TMS. The role of the motor cortex was investigated in a sensibility task on sentences again constructed by coupling concrete (i.e., action-related) and abstract verbs with nouns of manipulable/non-manipulable objects. Single-TMS pulses were applied to the left primary motor cortex 250 ms after verb or noun presentation in each of the four possible verb-noun, with registration of the cortico-spinal excitability of the right first dorsal inter-osseous muscle. The MEPs pattern observed after TMS pulse during noun presentation revealed greater peak-to-peak amplitude in phrases containing abstract rather than concrete verbs. The response times analysis confirmed that compatible verb-noun combinations (i.e., concrete verb and noun, abstract verb and noun) were processed faster than mixed combinations. In addition, combinations containing concrete verbs were responded faster when the pulse was delivered on the first word (i.e., verb) than on the second one (i.e., noun). Thus, the results were in line with findings of early activation of hand-related areas after concrete verbs processing, with a prolonged or delayed activation of these same areas by abstract verbs explained by referring to the different MoA of concrete and abstract words. Finally, also a very recent study by Sakreida et al. (2013) confirmed this evidence using a similar paradigm in a fMRI setting. Participants were required to read sentences composed by the same concrete noun coupled with concrete and abstract verbs, as well as an abstract noun with either kind of verbs previously used. As expected, the results showed that comprehension of concrete and abstract content activated the core areas of the sensory-motor neural network, precisely the left lateral and medial premotor cortex (i.e., supplementary motor areas). Additionally, pure concrete sentences elicited activations within the left inferior frontal gyrus and two foci within the left inferior parietal cortex, whereas pure abstract sentences were represented in the language processing system, namely the anterior left middle temporal gyrus. Although the sensory-motor neural network was engaged in both concrete and abstract language contents, the present findings on functional activity confirmed that concrete sentences relied more on the sensory-motor system, while abstract sentences reading relied more on the linguistic system.

The results reviewed are perfectly in line with WAT theory predictions, that explicitly suggest MoA as determining the representation of language at the neural level. Indeed, a word referring to a sensory-motor experience-based concept, like GLASS, might present a more extended grounding in perception-action systems than words learned through linguistic mediation, like FREEDOM. This idea is confirmed by data showing precocious activation of motor areas in a somatotopic fashion for concrete words (e.g., Pulvermüller, 1999, 2005; see the evidence reviewed in par. 3.3), whereas abstract words might elicit more linguistic information. However, WAT does not imply that abstract language rely simply on phono-articulatory experiences: if language representation was based only on speaking and listening experiences, it would not be enough to consider it as embodied. In contrast, opposing to disembodied approaches to language abstract concepts/words (e.g., DEMOCRACY) would activate as well sensory-motor information in WAT perspective. WAT is not a dual code theory (e.g., Paivio, 1991), according to which abstract words would rely only on the verbal linguistic system. WAT is instead a multiple representations approach, according to which both concrete and abstract words are grounded in perception, action and emotion systems, and even if the linguistic system plays a major role for abstract words representation they rely for a larger extent on the re-enactment of subjective and introspective experiences (see also Scorolli et al., 2011). It is worth noticing here that an embodied proposal which, in a way similar to WAT, may offer the possibility of accounting for both abstract and concrete words has been put forward by Vigliocco et al. (2009). The principal hypothesis of the authors is that two source of information underlie the representation of all concepts. The first source is experiential (e.g., sensory, motor, and also affective) and the second is linguistic (e.g., verbal associations determined by co-occurrence and syntactical patterns), with differences between concrete and abstract word meanings resulting from the relative proportions of experiential and linguistic information they rely on. Thus, analogously to what proposed by WAT, Vigliocco et al. (2009) affirmed that the supposed dichotomy of concrete and abstract words/concepts arises in function of the statistical preponderance of sensory-motor information that underlie concrete word meanings and the reciprocal statistical preponderance of affective and linguistic information to underlie abstract word meanings. This approach strongly underlies that emotion has to be considered as another kind of experiential information, that play a crucial role in acquisition and representation of abstract words especially.

To close, even if conclusive evidence on the topic is still far from being produced, proposals such as the WAT theory, that refers to both developmental phenomena and multiple kinds of experiences and mental representations to explain the grounding of language, seem to be really comprehending approaches very promising in the direction of solving the problem that abstractness posit to embodied theories.

3.6. Embodiment and the evolution of language

A discussion of the embodied views on language cannot be concluded without referring to its possible phylogenetic evolution. Indeed, one of the classic question in philosophy and psychology is about how language evolved. Especially following the discovery of a human mirror neurons system (e.g., Rizzolatti et al., 1998) a number of scholars have approached this topic inside the embodiment framework, producing interesting proposals (e.g., Arbib, 2005; Corballis, 2002; Gallese, 2008; Gentilucci & Corballis, 2006; Gentilucci & Dalla Volta, 2008; Hurford, 2011; Ramachandran & Hubbard, 2001; Rizzolatti & Arbib, 1998; Rizzolatti & Craighero, 2004).

The basic idea that links embodied perspectives on language evolution is the possibility of a gestural, iconic origin for contemporary conventionalized languages (e.g., Arbib, 2005; Corballis, 2002, 2009, 2010; Flumini, 2011; Gentilucci & Corballis, 2006; Ramachandran & Hubbard, 2001). The proposal that language may have originated from manual gestures is long-standing, dating at least the XVIII century. For example, the philosophers Vico (1744), Condillac (1746) and Rousseau (1782) referred to a gestural theory more or less openly in their essays, while in the following century Darwin (1871) explicitly described the role of gestures for human language: “I cannot doubt that language owes its origins to the imitation and modification of various natural sounds, and man’s own distinctive cries, aided by signs and gestures” (1871/2004, p. 76). However, the gestural theory languished in modern psychology until the late XX century, as it was freshened by the influential work of Pinker & Bloom (1990). A first source for the renewed interest was the growing evidence on signed languages that suggested they present similar grammatical and semantic sophistication, when compared to spoken languages (e.g., Emmorey, 2002). At least in principle, human language could have existed in an original form that, analogously to contemporary signed languages, was entirely gestural, with a gradual, phylogenetic blending of vocal components in this kind of linguistic system (Corballis, 2009). Supporting evidence also came from the study on the communication systems of non-human primates, which highlighted their ability in progressing toward the learning of a form of manual sign language (Gardner & Gardner, 1969), or to the pointing to visual symbols on a keyboard (Savage-Rumbaugh et al., 1998), while the attempt to teach them to speak have been always a failure. Finally, as said before, the gestural theory was pushed also by the discovery of mirror neurons in the primate brain. Indeed, as described in par. 2.4 and par. 3.3, the human homologue of area F5 in monkey's ventral premotor cortex is Broca's area (e.g., Rizzolatti & Arbib, 1998), traditionally associated with speech production. Nowadays there is well-supported evidence that this neural region is involved not only in speech, but also in motor functions including complex hand movements and sensory-motor learning (e.g., Binkofski & Buccino 2004). Moreover, the extended mirror system in monkeys has been shown to overlap with the homologue cortical circuits in humans that sustain language, leading some scholars to assume that language is originally a part of the mirror system itself (e.g., Fogassi & Ferrari, 2007). Indeed, the primate mirror system has to do mainly with manual gestures, although facial movements also play a role, and even if it does not seem to incorporate vocalization, it is receptive to acoustic as well as visual input (Rizzolatti & Sinigaglia, 2008). This is supported by results on single neurons recorded in monkey area F5 that showed responses to the sounds of manual actions, like breaking peanuts, in absence of visual presentation, while no response to monkey calls was detected (Kohler et al., 2002). Also other properties of the primate mirror system suppose a close connection between hand and mouth. For example, area F5 neurons responded when the monkey grasped an object with both the hand and the mouth (Rizzolatti et al., 1988). As suggested elsewhere, "the close neural associations between hand and mouth may be related to eating rather than communication, but later exapted for gestural and finally vocal language" (Corballis, 2009, p.25). A linkage between hand and mouth was also indicated in humans by behavioral results. For example, in the kinematic study by Gentilucci et al. (2001) participants were required to open their mouths while grasping objects, and the kinematics of the maximum opening of the mouth showed increased/decreased size in function of the manually grasped object size. Reciprocally, when participants were required to open their hands while grasping objects with the mouth, the hand grip also showed increased/decreased size in function of the grasped object size. A following study showed also that manual grasping of larger objects induced selective modulations (i.e., increasing) in lip kinematics parameters as in voice spectra of syllables produced during action execution (Gentilucci et al., 2004). This finding indicated that grasping with the hand is able to immediately affect the kinematics of speech itself. Thus, it has been proposed that in the course of phylogenetic evolution such a hand/mouth joint control mechanism could have favored the transfer, based on the mirror system, from a communication system based on hand movements to the contemporary one, based on mouth movements (e.g., Corballis, 2009; Gentilucci & Corballis 2006).

Following the assumption that speech evolved from gestures, it is very likely that, during the gradual phylogenetic shift to verbal language, speech initially reproduced the characteristics of the gestural communication medium it emerged from. If we look to actually observable gestural languages, that are contemporary signed languages, the principal property they exhibit is *iconicity*. F or example, it has been estimated that in Italian Sign Language (LIS) about the 50% of the hand signs and the 67% of the bodily locations stem signs have an iconic representations (Pietrandea, 2002). Also in American Sign Language (ASL) the larger part of the signs are iconic, while only some are purely arbitrary (Emmorey, 2002). Iconicity means here that there is a degree of spatiotemporal resemblance in the mapping of meanings to signs. However, if for example the ASL sign for ERASE reproduces the action of erasing the surface of a blackboard, many signs are not so transparently iconic, thus often they are not recognized by inexpert observers (Pizzuto & Volterra, 2000). An important consequence of sign languages analyses is the discovery that signs during recursive usage tend to become less iconic and more arbitrary. This process is known as conventionalization (Burling, 1999), and operates in order to maximize the efficiency of a communication system, reshaping its constraints and increasing its speed. Thus, at any step by which conventionalization has been phylogenetically established to a larger extent in a human gestural communication medium, there was less necessity to maintain iconic components and even to depend on visual indexes. However, even if humans seem to be very quick in learning arbitrary labels, manual and bodily gesturing might be still needed to establish referential links in the first developmental steps. For example, a young child could hardly acquire the meaning of the word CAT, if nobody draws the attention to the animal pointing to it while producing the name. Moreover, the labels themselves may facilitate the establishment of referentiality being directly based on certain patterns of sounds that immediately resemble their meanings, following a mechanism called sound-symbolism or phonosemantics (Hinton et al., 1994). Sound-symbolism is the iconicity of vocal languages. Indeed, it is not striking that words denoting for example concepts such as the calls of animals (e.g., MIAOW) rely on sound (rather than sight) in a way that the words' sound results iconically related to the calls' sound. This is the most evident case of sound-symbolism in modern lexicon, that is defined onomatopoeia. Anyway, as the largest part of words in contemporary languages seem not to bear any resemblance relation with their referents, the well-known principle of arbitrariness was indicated as the defining property of human (verbal) language (de Saussure, 1916), ending up it with the supposition that signed languages, with their foundation on iconic representations, are not true languages. The arbitrariness of words and morphemes, rather than being a necessary property of language, is better conceived as a matter of expedience depending on the kind of constraints imposed by language mediums. Indeed a more lucid analysis of speech realized that speech itself requires the information to be linearized and squeezed into sounds sequences that are necessarily limited in terms of how much they resemble the physical nature of what they refer to (Corballis, 2009). In this direction it has been clearly noticed that "when a representation of some four-dimensional hunk of life has to be compressed into the single dimension of speech, most iconicity is necessarily squeezed out. In one-dimensional projections, an elephant is indistinguishable from a woodshed. Speech perforce is largely arbitrary, if we speakers take pride in that, it is because in 50,000 years or so of talking we have learned to make a virtue of necessity" (Hockett, 1978, pp. 274–275). On the contrary, signed languages gesturing does not seem to be so constrained, as hands and arms movements can freely mimic real-world objects shapes or reproduce action, with linguistic information being delivered in parallel in place of being forced into rigid schemata. Even so, conventionalization allowed signs in deaf languages to be simplified, speeding up communication, to the point that many of them may lose (at least in part) their iconic aspect. Thus, following well-known principles of cognitive load economy (Zipf, 1949), it is frequent use what made signs become more compact and, in a similar way, what made words shorter and detached from their referents (Corballis, 2009). Thus, arbitrariness is the result of this recursive process, rather than being the original form of human languages.

Several scholars have already suggested that sound-symbolism played an important role in the evolution of contemporary languages, being a sort of phylogenetic grounding for verbal language (e.g., Kita, 2008; Kita, Kantartzis, & Imai, 2010; Kovic et al., 2009; Ramachandran & Hubbard, 2001). Indeed, it is very likely that a key step in evolving languages would have been the emergence of a system for agreeing upon the referents of that communication medium signs, being them gestural or vocal. If visual iconicity sustained this step for gestural languages, sound-symbolism - intended as the natural, non-arbitrary link between the sound and the meaning of a word – did it for speech. Supposing the existence of a universal sound-symbolism seems to be the easiest and most economic way in which such an agreement could have been reached in vocal languages. Indeed, if a sound-to-meaning linkage exists in everybody's mind, then the listener would be able to rapidly identify the referent of the speaker's novel word, making communication easier and efficient (Kantartzis et al., 2011). Thus, given that sound-symbolic words in modern languages can refer to sound-meaning correspondences about information from various modality, such as vision, touch, smell, taste, movement manners, emotion, attitude (e.g., Kita, 1997), our ancestors' sound-symbolic proto-words may have had a considerable expressive power (Kita, 2008) and universal sound-symbolism could have facilitated a rapid growth of shared lexicon (Ramachandran & Hubbard, 2001). Thus, universal sound-symbolism would have had a strong adaptive value for primordial hominins, and probably – as evolution is conservative - all humans are nowadays endowed with implicit abilities for sensing universal sound-symbolism and using it for word learning. If the emergence of such dispositions was really a crucial step in language evolution, and if the capacity of using sound-symbolic relations in learning is really rooted in the evolutionary history, it should be observable in children who are learning languages (Kantartzis et al., 2011). And indeed it was; for example, it has been shown that Japanese children have a strong preference, even greater than in adults, to use sound-symbolic words when describing the manner of motion in a narrative task (Kita et al., 2010). Previous research also demonstrated that generalization of novel verbs, which has been shown to be a complex task for children (e.g., Imai et al., 2005), becomes easier for 3-year-olds when the verbs sound-symbolically match the action they represent (Imai et al., 2008). Novel verbs that sound-symbolically matched or not the actions they refer to were taught to Japanese 3-year-olds participants (the novel sound-symbolic verbs were created on the basis of real Japanese sound-symbolic words). Participants failed to generalize a novel verb to an instance of the same action performed by a different actor if it did not have a sound-symbolic relation to the referent, but they succeeded in the task when the novel verb sound-symbolically matched the action it denoted. The results clearly showed that Japanese children acquired new verbs in a way that made then possible to correctly generalize their meaning if the novel word sound-symbolically resembled the action it referred to. These findings led the authors to explicitly propose a sound-symbolism bootstrapping hypothesis which affirms that sound-symbolism helps children to single out the referent of a novel word in the complex reality, thus favored symbols grounding, allowing them to store the semantic representation in such a way that after they can correctly generalize the verb to new situations.

Given the cross-linguistic recognizability of sound-symbolism, the question about whether the Japanese children had picked up regularities in the Japanese sound-symbolic lexicon or were sensitive to universal sound-symbolism was unsolved. To disentangle the two possible explanations, Kantartzis et al. (2011) performed a verb generalization task in which English-speaking 3-year-olds were taught novel sound-symbolic verbs created starting from Japanese sound-symbolic words, or novel non sound-symbolic verbs. English-speaking children performed better with the sound-symbolic verbs, just like Japanese-speaking children. These findings seem to confirm that children are sensitive to universal sound-symbolism and can rely on it in word learning and meaning generalization regardless of their native language. Nygaard et al. (2009) also extended this results to adults by examining the influence of potential non-arbitrary sound-meaning mappings on word learning. In a vocabulary-learning task, English-speaking monolinguals learned meanings for Japanese words, with each spoken Japanese word being paired with English meanings that could either match the actual meaning of the Japanese word (e.g., HAYAI paired with FAST), or be antonyms for its meaning (e.g., HAYAI paired with SLOW), or be randomly selected from the total set of antonyms (e.g., HAYAI paired with BLUNT). The results showed that participants learned the actual English equivalents and antonyms for Japanese words more accurately and responded faster than when learning randomly paired meanings. Thus, these findings confirmed that natural languages contain non-arbitrary sound-to-meaning mapping, to which learners are sensitive within spoken language (see also Brown, Black, & Horowitz, 1955; Iwasaki, Vinson, & Vigliocco, 2007). In addition, also sound-to-shape correspondences (i.e., strident sounds assigned to jagged shapes; sonorant sounds assigned to rounded shapes, e.g., Kohler, 1947) have been reported by Maurer et al. (2006) in both 2.5-years-old children and adults, using a classic double-forced choice paradigm (my empirical research on this topic, with a number of methodological improvements in respect to this setting, is reported in chapter 9) . Participants were shown simultaneously two shapes, a rounded and a jagged one, and had to name each shape choosing the name within two alternatives, one with rounded vowels and the other with unrounded ones (a different explanation, based also on consonant sounds, is suggested by Nielsen & Rendall, 2011). Results showed that children, as well as adults, matched sonorant words to rounded shapes and strident words to the jagged ones, confirming that if the sound-shape correspondences are at work at the earliest stages of language acquisition, this preference is then conserved during ontogenetic development.

The evidence reviewed here boost recent hypotheses (e.g., Perniss et al., 2010) that indicated iconicity as playing a decisive role at different levels for both spoken and signed language. Indeed, it is very likely that symbolic resemblance relations might be crucial for more than one fundamental aspect of language. For example, in phylogenesis iconicity may facilitate displacement, thus the ability to refer about things distant in time and space. In ontogenesis, it might provide the mechanism for establishing referentiality, linking linguistic form and meaning facilitating word learning and following lexical access. Finally, in language processing iconicity might facilitate grounding of words in sensory-motor and emotional systems, determining embodied grounding (Perniss & Vigliocco, in press).

On the whole, it seems that universal sound-symbolism in modern languages may be the heritage of an iconic (gestural and vocal) communication system once used by our ancestors. Such heritage might have been preserved in contemporary spoken languages because children, and along with them adults, have a preference in using sound-symbolic words over non sound-symbolic words during early development and in novel words learning. Thus, both language phylogenetic evolution and linguistic ontogenetic development seem to be more natural affairs than what traditionally thought.

PART II

Experimental work

4. Affordance-based compatibility effects in different contexts

A growing body of research shows tight (bidirectional) relationships between visual objects and affordance activation. In the literature, affordance-based compatibility effects have been observed in the context of a number of cognitive tasks (e.g., deciding whether an object is upright/inverted, artificial/natural etc.; see chapter 2), implying a fundamental role for sensory-motor simulation in cognitive processing predicted by embodied cognition theories. This evidence, though compelling, as one evident limitation: these experimental studies typically focused on objects in isolation, while in every-day interaction objects are embedded in specific contexts (see chapter 5), which enclose also other agents' actions. In the present study, affordance activation was investigated in visual contexts defined by different relations between 2 objects and a hand. Participants were presented with pictures displaying 2 manipulable objects linked by a functional relation (e.g., scissors-sheet), by a spatial relation (e.g., scissors-stapler), or by no relation (e.g., scissors-bottle), and had to decide by key-presses (performed with the hands in Experiment 1, and with the foot in Experiment 2) whether the objects exhibit a relation or not. To determine if observing others' actions and understanding their goals would facilitate the judgments, a hand was displayed in the pictures: a. near the objects; b. grasping the active object to use it; c. grasping an object to manipulate/move it; d. not displayed. The results showed faster responses if the objects were functionally rather than only spatially related, revealing a modulation of the visual context on affordances. Crucially, in Experiment 1 manipulation postures were the slowest in the functional context and functional postures were inhibited in the spatial context, probably due to the mismatch between the inferred goal and the context. The absence of this interaction for foot responses in Experiment 2 confirms that the effects observed in Experiment 1 were due to a motor simulation, rather than to simple associations between contexts and hand-postures. On the whole, these results confirm that affordance-based compatibility effects are modulated by the context in which objects are embedded, as spatial and social cues related to action influenced the responses of participants.

4.1. Introduction

The ability to act appropriately with objects, to respond to objects affordances, and to flexibly adapt our actions to the current situation, are important building blocks of human capability to interact with the environment. While affordances have been intensively studied, the mechanisms according to which their activation is modulated by the context, and particularly by the context in which actions of others are displayed, are poorly understood.

About 30 years ago, Gibson (1979) used the term affordance to indicate properties the environment provides to acting organisms which are relevant for an organism’s goals. According to Gibson, affordances are variable and relational, as they emerge from the interaction between organisms, objects and the environment. For example, a chair affords sitting for human adults but not for other organisms such as elephants or mosquitos, nor for human toddlers. Today, the contribution given by Gibson is widely recognized, though the term affordances is used in slightly different ways. An example is given by Ellis and Tucker (2000) proposal to use the term microaffordances to indicate the activation of action components (e.g., reaching and grasping components) suitable for interacting with specific objects. The continuity with the view of Gibson is obvious. However, in antithesis with Gibson’s view, recognizing an object is necessary to activate its microaffordances. In addition, microaffordances would be represented in the brain, that is they are conceived of as the product of conjoining in the brain of specific visual and motor patterns.

In the last 10 years a lot of evidence on affordances has been provided. On the neural side, many brain imaging studies have shown that observing an object activates possible actions to perform with it (for a review see Martin, 2007). Specifically, activation of parietal and premotor cortex has been linked to perception of tools' affordances (e.g., Jeannerod, Arbib, Rizzolatti, & Sakata, 1995; Johnson-Frey, 2004). On the behavioural side, many issues related to affordances have been investigated, particularly with compatibility paradigms (e.g., Borghi et al., 2007; Bub, Masson, & Cree, 2008; Caligiore, Borghi, Parisi, & Baldassarre, 2010; Gianelli, Dalla Volta, & Gentilucci, 2008; Girardi, Lindemann, & Bekkering, 2010; Riggio et al., 2008; Tipper, Paul, & Hayes, 2006; Tucker & Ellis, 1998, 2001; Yoon & Humphreys, 2007). A typical way to study activation of affordances is to verify whether an object characteristic related to action, such as object size, has an impact on a task (e.g., categorization task) for which size is not relevant. If this is the case, this would mean that affordances related to object's size and graspability are automatically activated. To notice that here the term ''automatically'' means ''independently from the task''. For example, Tucker and Ellis (2001) found that mimicking a precision or a power grip to decide whether an object is an artifact (e.g., hammer, nail) or a natural object (e.g. apple, cherry) was influenced by the object size, which was not relevant to the categorization task. A compatibility effect was found, that is small objects (e.g., nail) were responded to faster with precision grip than with power grip responses, whereas the opposite was true for larger objects (e.g., bottle). The authors interpreted their results claiming that observing an object automatically activates its affordances.

This evidence, though compelling, has one limitation we have considered in this work. In the majority of current studies objects (which are typically images and, less frequently, real objects) are often considered independently from the context in which they are usually embedded. This is striking, given the great relevance of contexts for object recognition and categorization. We perceive the world in scenes: perceiving objects embedded in a context facilitates object recognition (Bar, 2004) and it is not surprising that humans have the peculiar ability to be very fast in categorizing scenes (Thorpe, Fize, & Marlot, 1996). Furthermore, studies on categorization have shown that presenting objects in scenes facilitates categorization, particularly of superordinate level categories (e.g., musical instruments, Murphy & Wisniewski, 1989; see also Borghi, Caramelli, & Setti, 2005). In spite of the relevance of contexts for a variety of processes, only a few studies have accessed to what extent the activation of affordances is modulated by the context. To our knowledge there are only a few exceptions. For example, Pezzulo, Barca, Lamberti Bocconi, and Borghi (2010) investigated how expert and novice climbers remembered routes of different difficulty on a climbing wall. To perform the task climbers had to take into account the relationship between each hold (affordance) and the context given by the presence of other holds on the climbing wall in order to simulate how they could grasp the holds with the hands and use them as support for the feet. More directly relevant to the present work is the study by Yoon, Humphreys, and Riddoch (2010), who focused on affordances elicited by pairs of objects that appear in the same scene and are positioned for action, such as a frying pan and a spatula. The authors found an effect they called ‘‘paired object affordance effect’’: the time taken by right-handed participants to respond whether the two objects were used together was faster when the active object (e.g., the spatula) was to the right of the other object.

The aim of our work is to verify the effects of different kinds of contexts on activation of object affordances for tools (see also chapter 5). The context is suggested by the presence of a second object, which can be either spatially or functionally related to the tool. The active object and the second object are presented either alone, with a hand in potential interaction with them, or with a hand in effective interaction with them. Specifically, in our study we address three issues pertaining the relationship between affordances evoked by the active object and the context in which it is embedded. They concern the behavioural effects of the relation between the 2 displayed objects, of the eventual presence of a hand near the object, and of the kind of interaction between the hand (when present) and the object. We will introduce these issues below.

(1) Spatial vs. functional relations between objects. Our work aims to verify whether the activation of affordances of the active object is modulated by the presence of a second object, thematically related to the first, either through a spatial or a functional relation (Borghi & Caramelli, 2003; Estes, Golonka, & Jones, 2011; Kalénine, Bonthoux, & Borghi, 2009; Kalénine, Peyrin, et al., 2009; Lin & Murphy, 2001; Natraj et al., 2013; Yoon & Humphreys, 2007). The notion of thematic relation is common in categorization literature. A given object may be categorized taxonomically, as a member of a given category (e.g., both elephants and foxes are animals), or thematically, as part of the same context or action (e.g., trees and houses, dogs and bones are linked by a thematic relation). Thematic relations between objects are known to be the preferential mode of categorization in young children (Lucariello, Kyratzis, & Nelson, 1992), but are still easily available in older children and adults (Borghi & Caramelli, 2003; Lin & Murphy, 2001). Here we consider two different kinds of thematic relations: we considered two objects as spatially related when they typically appear in the same context but are not directly used together; we classify them as functionally related when the two objects not only appear in the same context, but are also typically used together. For example, we have often experience of knives in the same context as coffee mugs – they are typically found in the kitchen, possibly on the table, but the two objects are only spatially and not functionally related. Instead, a knife is not only spatially but also functionally related to a stick of butter.

(2) Presence of a hand near the objects. Previous studies using a categorization task (Borghi et al., 2007; Vainio, Symes, Ellis, Tucker, & Ottoboni, 2008; see also Fischer, Prinz, & Lotz, 2008) provide evidence of a compatibility effect between the posture of a hand-prime (displaying either a power or a precision grip) and the grip (power vs. precision) required to

grasp a target-object (see also studies on predictions we form based on observation of objects grasping, e.g., Becchio et al., 2012). However, the two aforementioned studies do not fully disentangle the effects given by the observation of the object and of the hand in potential interaction with it. Indeed, it is possible that the activation of different neural areas underlies observation of objects alone or of hands potentially interacting with objects. Studies on the premotor cortex in monkeys have demonstrated the existence of two different classes of visuomotor neurons, canonical and mirror neurons (see chapter 2). Canonical neurons (Murata et al., 1997; Raos, Umiltà, Murata, Fogassi, & Gallese, 1996) discharge during interaction with graspable objects and also during simple object presentation; mirror neurons, instead, discharge both during action execution and during observation of a conspecific interacting with an object (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992), but do not fire in response to a simple object presentation. Further neuroimaging studies have confirmed the existence of canonical and mirror neuron systems in humans (for a review see Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2010). Borghi et al. (2007) and Vainio et al. (2008) argued that it is possible for two different mechanisms, one related to canonical (affordances) and the other to the mirror neuron systems (motor resonance while observing others acting with objects), to be responsible for the effects found. Liuzza, Setti, and Borghi (2012), using a weight decision task (light vs. heavy), showed that when a grasping hand prime preceded the object, participants responded slower to heavy than to light objects, while no difference was found when no hand prime was present. This confirms that two different mechanisms might be at play. This is similar to what has been proposed in the perceptuo-motor domain in considering living vs. non-living action affordance effects (Bennet, Thomas, Jervis, & Castiello, 1998; Castiello, 2005). The present study allows us to disentangle the contribution played by these two different mechanisms without using a priming paradigm but presenting the hand together with the object. Indeed, in one condition subjects saw the two objects alone, without any hand: in this condition we hypothesize that only the Canonical neuron system would be activated. In a further condition, a hand was displayed close to both objects but not interacting with them (Still-Hand): this condition was aimed to verify whether the simple presence of a hand, even if not in a prehensile posture, produced a facilitation in processing action relations between objects. Since the Mirror Neuron System is typically activated when the action goal is present, it is possible that an action simulation occurs only when the hand is presented with a prehensile posture. If a mirroring mechanism is responsible of this simulation, then the simulation should occur only with manual responses (Experiment 1), not with foot responses (Experiment 2), due to an effector-specific motor resonance mechanism (Paulus, Lindemann, & Bekkering, 2009). However, given the behavioural nature of our study, these proposals cannot be conclusive claims on the underlying neural mechanisms.

(3) Relations between hand and object: manipulation vs. function. Psychological (Costantini, Ambrosini, Scorolli, & Borghi, 2011; Jax & Buxbaum, 2010; Pellicano, Iani, Borghi, Rubichi, & Nicoletti, 2010) and neuropsychological studies provide support for two different ways of interacting with objects (Buxbaum, Sirigu, Schwartz, & Klatzky, 2003; Jeannerod, Decety, & Michel, 1994), which have been termed by Bub et al. (2008) as volumetric and functional (see chapter 2). Volumetric gestures are associated with the overall shape and weight properties of objects and concern the hand posture used to grasp an object to lift or move it, rather than to use it for its defined purpose. Functional gestures, on the other hand, involve specific manipulation of objects in accordance with their proper conventional use. Recent studies by Buxbaum have focused on the so called ‘‘conflict objects’’ (see chapter 5), that is objects that evoke different actions (and different ways to manipulate them) depending on the action goal (Jax & Buxbaum, 2010). In addition, Buxbaum and Kalenine (2010) have recently proposed that two different circuits underlying different affordances

might be activated, one based on the object structure (dorso-dorsal stream), and another related to object function (dorso-ventral stream; Rizzolatti & Matelli, 2003) (see also Young (2006) and Borghi and Riggio (2009) for a similar proposal of different kinds of affordances, more related to manipulation vs. to use; see also Pellicano, Thill, Ziemke, & Binkofski, 2011). In our study the hand can interact with the object in a manipulative vs. functional way (Manipulation-Hand vs. Function-Hand). Thus observing an action suggests two different underlying goals. Consider for example grasping a fork in order to place it somewhere else; the fork does not necessarily have to be grasped by the handle. Instead, if a fork is grasped for use, then it has to be held by the handle with a very specific grip (Creem & Proffitt, 2001). If a specific motor program is activated when the hand interacts with the objects, and if by observing an action with an object we infer the underlying goals of the agent, we expect a different effect depending on whether the hand is presented in a functional interaction with the object or in a manipulative interaction with it.

Based on these three issues concerning the relationship between affordances and the context in which objects are embedded, our predictions are the following:

1. If the activation of the affordances of the active object (possibly mediated by the activation of the Canonical Neuron System) is sensitive to the context given by the presence of a second object, then participants should respond faster and more accurately to the functional than to the spatial context, given that the functional context allows using the two objects together, while the spatial context does not allow a combined action with the two objects.

2. If observing a hand together with an object activates a simulated interaction with it (possibly through the mediation of the Mirror Neuron System), then participants should respond faster and more accurately when a hand is presented than when no hand is displayed, particularly if the hand is in a grasping posture.

3. If the activation of affordances is sensitive to fine-grained contextual information, then we should find an interaction between kind of context and kind of posture, due to a mismatch between the inferred action goal and the context. Thus, manipulation postures should be processed slower in the functional than in the spatial context; functional postures, instead, should lead to slower responses in the spatial than in the functional context.

4.2. Experiment 1

Experiment 1 was aimed to verify the aforementioned predictions; participants were required to respond by pressing two keys on the keyboard.

4.2.1. Method

4.2.1.1. Participants

Sixty-two participants volunteered for participation in the experiment (24 males; mean age = 28.18). All were right handed by self-report and had normal or corrected-to-normal vision. All were naive as to the purposes of the experiment.

4.2.1.2. Materials

A special care was taken in selecting the materials. 24 everyday manipulable objects were chosen. Every target-object was presented in three different pictures, paired with three different artifacts; thus we obtained a total of 72 pictures displaying pairs of objects, in which the active member of the pair was located on the right. For example, a picture displayed scissors (the active member of the pair) located on the right with respect to a sheet of paper. All images were presented in an egocentric position (Bruzzo, Borghi, & Ghirlanda, 2008). In each pair the target-object was composed by a part graspable for object use and a manipulable part: for example, when we use a fork to eat we typically hold it by the handle (Function hand), but not by grasping it by its teeth (Manipulative hand). There were 3 kinds of pairs depending on the relation existing between the two objects. This relation could be: (1) functional, when objects are both typically located in the same context and are used together (e.g., knife–butter); (2) spatial, when objects typically occur in the same context but are not used together (e.g., fork–spatula); (3) no relation, when the two objects are unrelated (e.g., knife–nail, scissors–bottle).

Before the experimental study, a separate group of subjects evaluated the pictures (without the hand) for familiarity. After asking participants to evaluate items familiarity, one item (‘‘potato-peeler’’) was eliminated. This lead to the removal of three pictures. Thus we obtained 69 object pairs that were selected to be used for the experiment. The pictures, presented in four different random orders, were rated by an independent groups of 20 subjects for visual complexity, using a 7-point-scale (7 very complex, 1 not complex at all), and were invited to use not just the poles but also the intermediate values of the scale. The analyses on visual complexity of the single objects revealed that there was no difference between the Functional ($M = 2.59$), the Spatial ($M = 2.60$) and the No Relation Context ($M = 2.76$), $F(1, 42) = 0.43$, $MSe = 0.47$, $p = .65$ (consider that in some cases the same passive object was used in different pairs).

For the experimental stimuli, the frame size was 730 pixels wide and 548 pixels high. Four pictures of each pair were taken, as each pair was presented in 4 different conditions. In one condition only the objects were displayed (No-Hand condition), whereas in further three conditions a hand was presented as well. The hand was always presented in an egocentric perspective, since it has been shown that processing is faster when the hand and the participant's perspective match (Bruzzo et al., 2008; Vogt, Taylor, & Hopkins, 2003), on the right side of the picture. The hand could simply be displayed near the object (Still-Hand condition), or it could interact with the object in a posture relevant for using it, for example grasping the fork handle (Function-Hand condition), or in a posture apt to manipulate the object, for example holding the fork teeth (Manipulation-Hand condition). Note that in our study the hand configuration is exactly the same for both functional and manipulative postures, but realized in a way that the manipulative posture simply does not afford tool-object interaction (see Figure 1, in which from the right top corner are displayed Function-Hand, Manipulation-Hand and Still-Hand conditions) but some other manipulation movements (e.g. turning the object around, rotating or relocating it etc.). We think that creating identical physical hand postures for both manipulative and functional interactions makes it possible to measure more clearly what is the extent of the context influence on affordances activation. Furthermore, at a methodological level, reducing the visual difference between two basilar experimental conditions keeps them more controlled and comparable. The complete list of the selected materials can be found in the Appendix of the thesis.

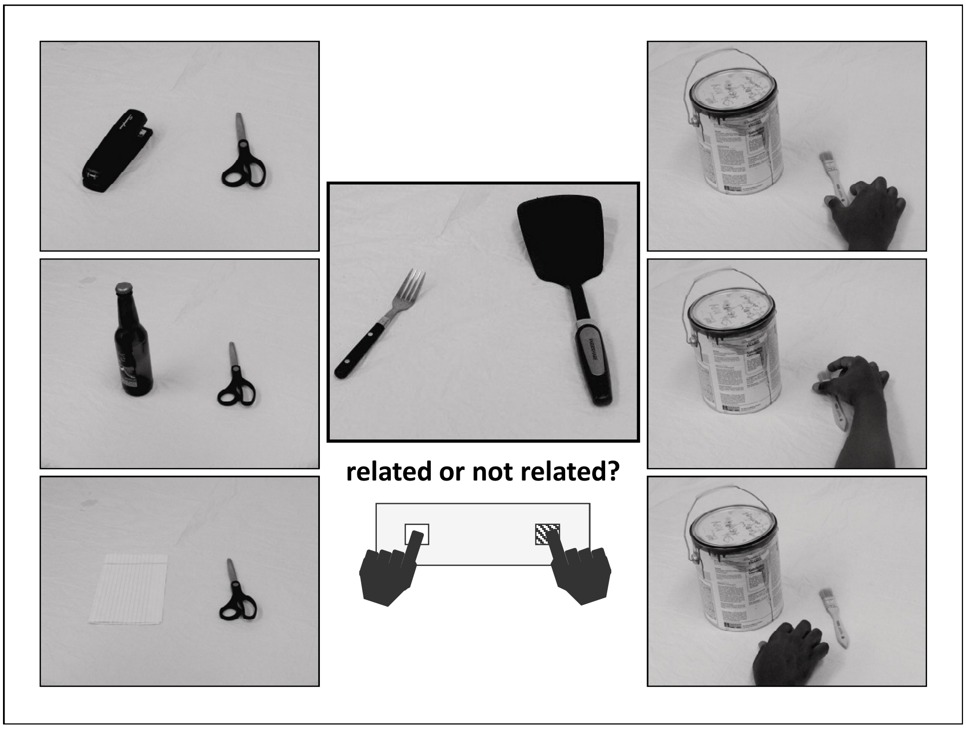


Figure 1. Stimuli and procedure. Left side, from the top: Objects with spatial relation (scissors–stapler), no relation (scissors–bottle), functional relation (scissors–sheet). Right side, from the top: Hand in functional posture, manipulative posture, close to the objects.

4.2.1.3. Design and procedure

Participants sat 50 cm from the computer screen, with their right and left hands placed over the ‘‘3’’ and the ‘‘9’’ key on the keyboard. Each trial began with a fixation point (+) that remained on the screen for 500 ms. Then one of the photographs was displayed at the center of the screen and remained on the screen until a response was made. Participants read the following instructions: ‘‘In the center of the screen a little cross will appear, followed by a picture showing two objects and sometimes a hand. You are required to decide if the two objects are usually seen/used together or not. If the two objects are usually seen/used together (e.g., a flowerpot and flowers) press the 9 key with your right hand, if the two objects are not usually seen/used together press the 3 key with your left hand. Please respond as quickly and as accurately as you can. The experiment lasts approximately 15 min. Press a key to start’’. We decided to ask participants to simply decide whether the objects were linked by some sort of relationship, and not by a functional relation, since we wanted to avoid rendering the aims of the study too transparent for participants. Indeed, our aim was to simply assess what differentially drove the strength of the association, whether the context, the hand, or both.

Since participants were required to decide whether the two objects were functionally or spatially related, or not, a ‘‘yes’’ response should occur in 2/3 of the trials, while a ‘‘no’’ response would occur in 1/3 of all trials. They had to respond ‘‘yes’’ with their dominant hand. Participants were instructed to respond as quickly and as accurately as possible and received feedback for both correct and incorrect responses. Each pair was presented once for each of the four hand conditions. Overall, the experiment consisted of one practice block of 12 trials and one experimental block of 276 trials.

4.2.2. Results

We performed separate analyses on the ‘‘yes’’ trials (i.e. trials requiring a ‘‘yes’’ response with the right hand, characterized by functional or spatial relations between the two objects) and the ‘‘no’’ trials (i.e. trials implying a ‘‘no’’ response with the left hand, characterized by the absence of relation between the two objects).

From the ‘‘yes’’ trials 12.79% of the trials were removed as errors. The low number of errors reveals that the task was easy to perform. Reaction times (RTs) more than two standard deviations from each participant’s mean were excluded from the analysis; this trimming method lead to the removal of 2.06% of the data. Errors and correct RTs were entered into a 2 x 4 within-subject ANOVA with the factors Context (Functional, Spatial) and Hand (No-Hand, Still-Hand, Manipulation-Hand, Function-Hand). Significant interactions were evaluated with Newman-Keuls post-hoc test ($p < .05$).

The ANOVA on errors demonstrated reliable main effects of both Context and Hand. The Spatial context ($M = 3.78$) elicited more errors than the Functional ($M = 2.30$) context, $F(1, 61) = 69.80$, $MSe = 3.89$, $p < .001$. The factor Hand, $F(3, 183) = 4.14$, $MSe = 2.09$, $p < .01$ was significant due to the fact that the Functional-Hand ($M = 2.73$) elicited less errors than both the No-Hand ($M = 3.18$) and Still-Hand condition ($M = 3.33$).

The ANOVA on RTs demonstrated reliable main effects of both the factors Context and Hand; the interaction was significant as well. The Spatial context reaction time ($M = 803$ ms) was longer in duration than the Functional ($M = 767$ ms) context, $F(1, 61) = 56.28$, $MSe = 2947.82$, $p < .001$. The factor Hand, $F(3, 183) = 14.41$, $MSe = 1544.72$, $p < .001$ was also significant. Post-hoc analysis revealed that it was due to the fact that the No-Hand condition ($M = 769$ ms) was significantly faster than all other conditions, and to the fact that Manipulation-Hand condition ($M = 802$ ms) was significantly slower than all other conditions. The Context x Hand interaction, $F(3, 183) = 2.78$, $MSe = 1424.95$, $p < .05$, depicted in Figure

2, reveals that with the Spatial Context RTs in the No-Hand condition (*M* = 787 ms) are faster than Manipulation-Hand (*M* = 814 ms) and Function-Hand conditions (*M* = 812 ms), probably due to the lower visual complexity of the first. To testify the sensitivity to the combination between the hand posture and the context, with the Functional context RTs in the Function-Hand condition (*M* = 760 ms) are slower than in the Still-Hand condition (*M* = 752 ms). However, with the Functional context, Manipulation (*M* = 789 ms) is slower than all other conditions (Newman-Keuls $p < .05$).

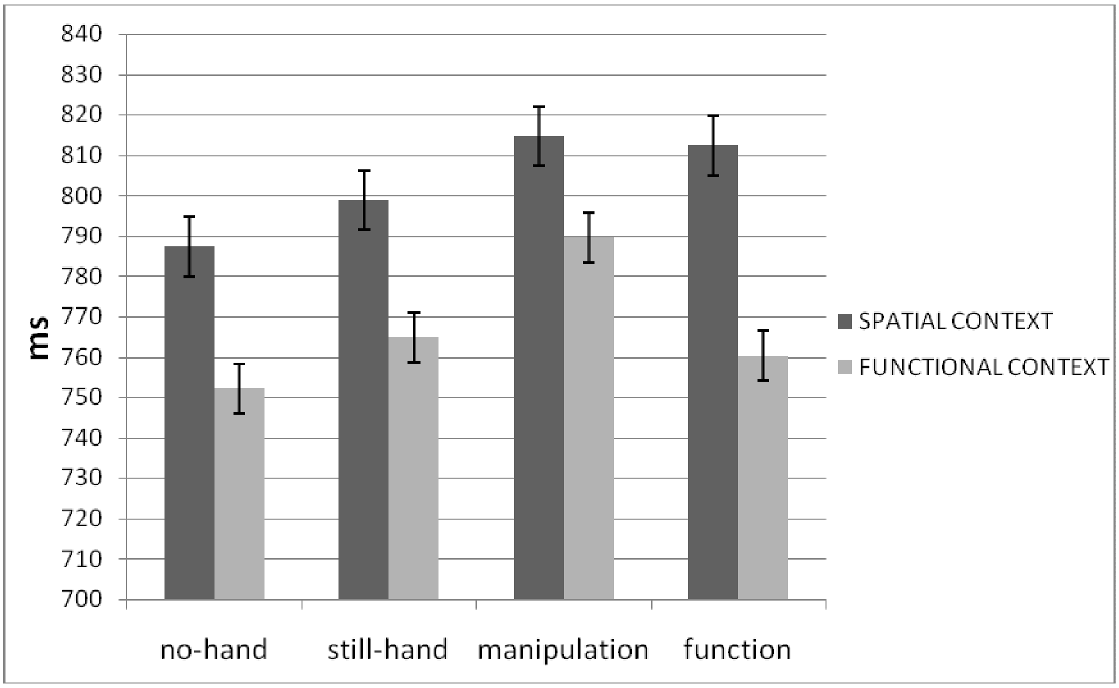


Figure 2. Experiment 1 - Interaction Context x Hand posture (Response times).

From the ‘‘no’’ trials 21.35% of the trials were removed as errors. We used the same trimming method as for ‘‘yes’’ trials: this method lead to the removal of 3.02% of the data. Errors and correct RTs were entered into a one-way within-subjects ANOVA with the four levels factor Hand (No-Hand, Still-Hand, Manipulation-Hand, Function-Hand). The ANOVA on errors did not show any reliable main effect, $F(3, 183) = 0.22$, $MSe = 2.63$, $p = .87$.

In the ANOVA on RTs, instead, we found a reliable main effect of the Hand factor, $F(3, 183) = 12.74$, $MSe = 2329.46$, $p < .001$, due to the fact that Manipulation-Hand (M = 967 ms) was significantly slower than all other conditions. Function-Hand (M = 927 ms), No-Hand (M = 916 ms) and Still-Hand (M = 929 ms) did not significantly differ.

4.2.3. Discussion

Based on our first hypothesis, as predicted, participants were more accurate and faster with the Functional context than with the Spatial context, suggesting that seeing functional pairs of objects heightens activity of the motor system. As to our second hypothesis, the presence of a hand did affect RTs. More crucially, related to the third hypothesis, the interaction we found indicate that manipulation postures were processed slower in the Functional than in the Spatial context. At the same time, the functional postures, which were processed rather fast in the Functional context, were slower in the Spatial context. This result can be explained in two different ways. The interaction can be simply the product of an association between the context and a specific hand posture. Alternatively, our results might be interpreted in terms of activation of a motor simulation. Participants would have more difficulties in simulating a manipulative action when the context implies using the object rather than simply manipulating it. Similarly, because in the Spatial context no functional use of the object is allowed, it is possible that the motor system continues to try to make sense of the scene, leading to longer RTs with the functional posture. It is therefore possible that in the Function-Hand condition not only effector-independent action information is activated, but that the perception of a functional grip evokes an effector-specific simulation. Experiment 2 was aimed at ruling out the first purely associative interpretation of the interaction: for this reason we had participants using foot instead of hand responses.

4.3. Experiment 2

The aim of Experiment 2 was to demonstrate that the interaction found in Experiment 1 was due to a motor simulation, related to a specific effector (i.e., the right hand). Indeed, it could be argued that Manipulation-Hand posture was slower in the Functional context and Function-Hand posture was slower in the Spatial context because of the lower association degree between a given context and a given posture. Experiment 2 was aimed at ruling out this explanation. It was identical to Experiment 1, but responses were provided with the foot instead than with the hand. We predicted that, with a different effector, no grasping motor simulation would occur, thus the Context x Hand interaction effect should not be present. As to the hand, our results do not allow us to determine definitively whether the lowest RTs obtained with the No-Hand condition are due to the lower visual complexity or to the concurrent activation of two different mechanisms. Experiment 2 can help us in solving this issue, as the response effector is the foot instead of the hand.

4.3.1. Method

4.3.1.1. Participants

Sixty-two participants volunteered for participation in the experiment (20 males; mean age = 23.53). All were right handed by self-report and had normal or corrected-to-normal vision. All were naive as to the purposes of the experiment.

4.3.1.2. Materials

The same materials as in Experiment 1 were used; the frame size of the stimuli was 730 pixels wide and 548 pixels high.

4.3.1.3. Design and procedure

Participants sat 50 cm from the computer screen, with their right and left feet placed over two pedals high 9 cm and wide 7.6 cm, placed 13.5 cm far from each other and at 20.5 cm from the frontal legs of the chair in which they sat. The procedure was exactly the same as in Experiment 1. The only difference was that participants were required to use foot responses: they had to respond ‘‘yes’’ with their right foot; ‘‘no’’ responses were 1/3 of the overall trials as in Experiment 1.

4.3.2. Results

As in Experiment 1, we split the data collected in two different groups depending on the required response (‘‘yes’’ right foot responses vs. ‘‘no’’ left foot responses), and performed separate analysis on them.

From ’’yes’’ trials 8.97% of the trials were removed as errors. Reaction times (RTs) more than 2 standard deviations from each participant’s mean were excluded from the analysis; this trimming method leads to the removal of 2.10% of the data. Errors and correct RTs were entered into a 2 x 4 within-subject ANOVA with the factors Context (Functional vs. Spatial) and Hand (No-Hand, Still-Hand, Manipulation-Hand and Function-Hand). Significant interactions were evaluated with Newman-Keuls post-hoc test ($p < .05$).

The ANOVA on errors demonstrated the reliable main effects of both the factors Context and Hand; the interaction was significant as well. As in Experiment 1, the Spatial context ($M = 3$) elicited more errors than the Functional ($M = 1.83$) context, $F(1, 61) = 65.22$, $MSe = 2.58$, $p < .0001$. The factor Hand was significant too, $F(3, 183) = 4.80$, $MSe = 1.67$, $p < .01$, due to the fact that Function-Hand ($M = 2.09$) elicited less errors than all other conditions, differing significantly from Manipulation-Hand ($M = 2.42$) and Still-Hand ($M = 2.71$), and almost significantly from No-Hand ($M = 2.45$) (Newman-Keuls $p = .07$). The Context x Hand interaction, *F*(3, 183) = 6.30, *MSe* = 1.62, $p < .001$, revealed that, while within the Spatial Context there were no significant differences between the 4 Hand conditions, in the Functional context the Still-Hand condition (*M* = 2.43) elicited significantly more errors than the Manipulation-Hand (*M* = 1.61) and Function-Hand (*M* = 1.24) conditions, but not differing from the No-Hand condition (*M* = 2.04) as well (Newman-Keuls $p = .09$).

The ANOVA on RTs demonstrated the reliable main effects of both the factors Context and Hand, but as predicted their interaction was not significant. RTs were slower in the Spatial context (*M* = 856 ms) than the Functional context (*M* = 800 ms), *F*(1, 61) = 91.92, *MSe* = 176.48, $p < .0001$. The factor Hand, *F*(3, 183) = 29.42, *MSe* = 1943.38, $p < .0001$, was also significant, due to the fact that Manipulation-Hand (*M* = 859 ms) was significantly slower than all other conditions. The Context x Hand interaction was not significant, *F*(3, 183) = 0.80, *MSe* = 1784.62, $p = .49$ (see Fig. 3).

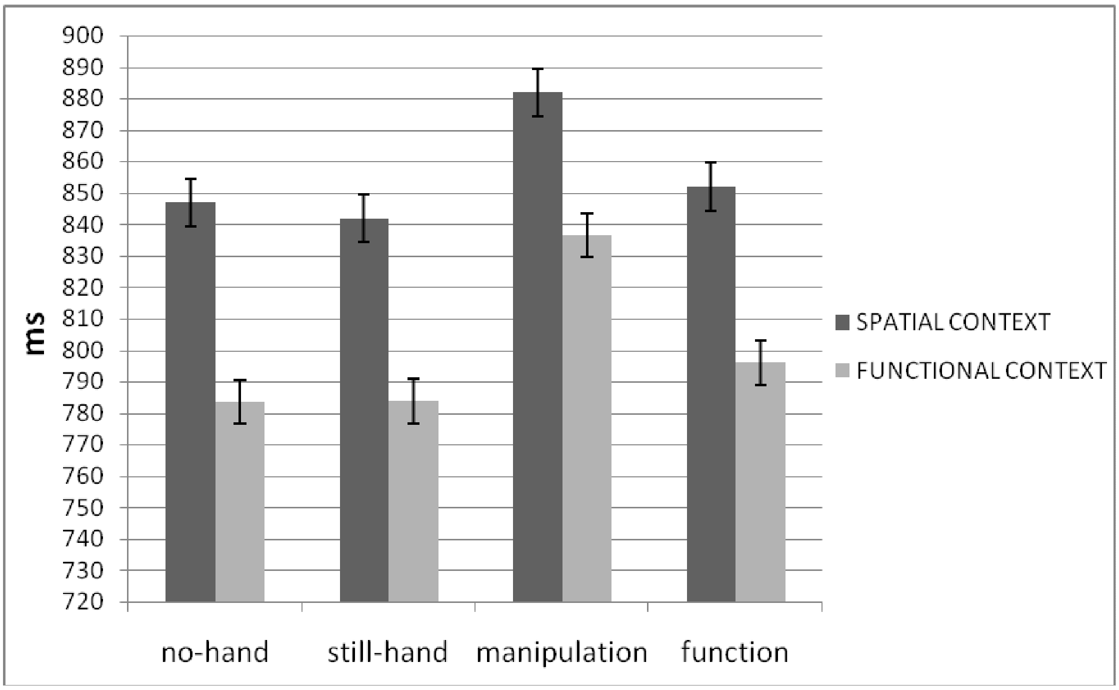


Figure 3. Experiment 2 - Interaction Context x Hand (Response times).

We removed as errors 18.04% of the trials from the ‘‘no’’ response trials. Using the same trimming method as before, 3.83% of the data were eliminated. Errors and correct RTs were entered into a one-way ANOVA with the 4 levels factor Hand (No-Hand, Still-Hand, Manipulation-Hand, Function-Hand) manipulated within-participants.

The ANOVA on errors demonstrated the reliable main effect of the factor Hand, $F(3, 183) = 3.26$, $MSe = 2.55$, $p < .05$: Manipulation-Hand ($M = 4.61$) elicited more errors than Function-Hand ($M = 3.79$) and No-Hand ($M = 3.94$), but not of Still-Hand ($M = 4.26$) (Newman-Keuls $p = .22$).

The main effect of Hand was significant also in the ANOVA on RTs, $F(3, 183) = 24.80$, $MSe = 2026.69$, $p < .0001$, due to the fact that Manipulation-Hand ($M = 1022$ ms) was significantly slower and that No-Hand ($M = 954$ ms) was significantly faster than all the other conditions; the difference between Function-Hand ($M = 985$ ms) and Still-Hand ($M = 972$ ms) did not reach significance (Newman-Keuls $p = .13$).

4.3.3. Discussion

The sensitivity to the difference between Spatial and Functional context found in Experiment 1 was confirmed. As to the role played by the displayed hand, the fact that we did not find an advantage of the No-Hand condition, as in Experiment 1, confirms that the result is not due to the lower visual complexity but to the activation of two mechanisms, one related to observation of others interacting with objects, the other to observation of objects evoking actions. More crucially for us, the absence of an interaction between Context and Hand for foot response rules out one of the possible interpretations of the interaction found in Experiment 1. Given that this interaction was present with the hand but not with the foot responses, we can argue that our main result is not due to the association between a given context and a given hand posture. Rather, it suggests that the interaction is due to a motor simulation, which is related to the effector involved (and it might activate grasping).

4.4. Comparison between the two experiments

To better understand the differences between the two experiments, we performed a 2 x 2 x 4 ANOVA for the ‘‘yes’’ responses. The factor Experiment (Manual responses vs. Foot responses) was manipulated between subjects, while the already used Context (Functional, Spatial) and Hand (No-Hand, Still-Hand, Manipulation-Hand, Function-Hand) factors were manipulated within subjects. Where permissible, interaction effects were evaluated with Newman-Keuls test ($p < .05$).

The ANOVA on errors demonstrated reliable main effects of Experiment, Context and Hand. The first effect was due to the fact that Manual responses ($M = 3.04$) elicited more errors than the Foot ones ($M = 2.42$), $F(1, 122) = 10.29$, $MSe = 9.48$, $p < .01$. The Spatial context ($M = 3.39$) yielded more errors than the Functional ($M = 2.07$) context, $F(1, 122) = 69.80$, $MSe = 134.05$, $p < .0001$. The factor Hand, $F(3, 366) = 7.34$, $MSe = 1.89$, $p < .0001$ was significant as with the Function-Hand ($M = 2.41$) errors were less than in all other conditions. The interaction Context x Hand, $F(3, 366) = 7.48$, $MSe = 1.80$, $p < .0001$, was significant due to the fact that, with the Functional context, Manipulation and Function-Hand had an advantage over the other two conditions, while within the Spatial context the four hand conditions did not differ.

In the ANOVA on RTs all main effects were reliable. RTs with Foot responses ($M = 827$ ms) were longer in duration than RTs with Manual responses ($M = 785$ ms), $F(1, 122) = 5.35$, $MSe = 84864.67$, $p < .05$. The factor Context, $F(1, 122) = 205.87$, $MSe = 2562.33$, $p < .0001$, was significant as RTs with the Functional Context ($M = 783$ ms) were faster than RTs with the Spatial (*M* = 829 ms) one. The factor Hand was also significant, *F*(3, 366) = 41.06, *MSe* = 1744.04, *p* < .0001, due to the fact that the Function-Hand was significantly faster and that the Manipulation-Hand was significantly slower than all other conditions. The interaction between Experiment and Context, *F*(1, 122) = 8.81, *MSe* = 2562.33, *p* < .01, was significant due to the combined effect of the advantage in both experiments of the Spatial Context over the Functional Context, and of the Manual over the Foot responses. The other significant interaction, the Experiment x Hand, *F*(3, 366) = 4.49, *MSe* = 1744.04, *p* < .01, showed that, while with the Manual responses the Manipulation-Hand factor was slower compared to all others and the No-Hand factor was the fastest one, with the Foot responses only the Manipulation-Hand condition differed from the others as it was the slowest one. The Context x Hand interaction just approached significance, *F*(3, 366) = 2.18, *MSe* = 1822.69, *p* = .09.

4.5. General discussion

Our results allow us to address the three principal issues advanced in the introduction. First, they indicate that the relations existing between objects have a strong effect on the responses. Our results suggest that positioning the objects for action facilitated the responses. As predicted, in both experiments participants were more accurate and faster with the Functional context compared to the Spatial context. In this respect, referring to work by Yoon et al. (2010) might be useful. Yoon et al. (2010) presented pairs of objects and submitted participants to two different tasks, an action decision task (they had to decide whether two objects were typically used together) and a contextual decision task (they had to decide whether both objects were for example kitchen items). They propose that responses to Task 1 depend on the time necessary to access action knowledge, whereas responses to Task 2 are dependent on the time necessary to access semantic knowledge. In our experiment, participants were required to decide whether the two objects were related or not, and the kind of relation linking the two objects could be either spatial (the two objects are typically located in the same place, e.g., in an office or in a kitchen) or functional (the two objects are typically used together).

The finding that responses were faster with functional contexts suggests that action knowledge is available earlier and accessed faster than the knowledge of objects' common location. In addition, the advantage of functional over the spatial context in both experiments might suggest that the first evoked a lateralized compatibility effect for the right effector, either hand or foot, since in both experiments the ''yes'' response was associated with a movement of a right effector.

One could argue that this might depend on the differences in semantic association between functionally and spatially related objects. This is certainly possible, and merits further investigation. In any case, we believe that the very fact that action relations between common objects lead to faster responses compared to spatial relations between common objects is in itself informative. A possible cause of this different accessibility can be found in the differences between functional and spatial relations. While functional relations between two objects are normally clear and, in some cases, even socially established (e.g., in Western societies you need to use a fork or spoon to bring food to the mouth), spatial relations are more subject to individual differences and less conventionalized (e.g., some people keep their scissors with their cutlery, others in their office desk). This higher variability might explain why participants needed longer and produced more errors to verify a potential spatial relation than a functional relation between two objects.

A further, less plausible interpretation is that the possibility to interact with both objects is activated in all cases, independent from the relations linking them. When the objects are linked only by a spatial relation, this possibility is activated and then discarded; this would slow down response times. However, in future studies we will give consideration to the idea that spatially related objects might be useful for a single bimodal action goal (Swinnen, 2002; Swinnen & Wenderoth, 2004) which is beyond what was presented in the present stimuli (one hand grasping one object).

The evidence we found that observing functional objects together activates possible actions does not imply that this activation is always effector specific. Indeed the finding that we respond faster to objects linked by a functional than by a spatial relation could depend on their being typically acted upon together, and suggests that a simulated action is activated. However, the same effect was present in Experiments 1 and 2, with both hand and foot responses. A possible explanation is that the difference between functional and spatial context concerns the overall action goal, not the specific effector to use, and it is therefore present with both the hand and the foot responses. This interpretation is coherent with the idea that observing objects activate the canonical neuron system (Murata et al., 1997; Raos et al., 1996), but not the mirror one. In other words, objects might evoke the simulation of an action that does not imply the involvement of a specific effector (hand, foot). When, instead, the functional hand posture is observed, then the Mirror Neuron System is activated as well, and the action is programmed at a more detailed and fine-grained level, through the recruitment of a specific effector. This interpretation is compatible with recent evidence on action hierarchies and action chains. It has been suggested that actions can be comprehended at different levels: the overall action goal (e.g., preparing a coffee), which can be segmented in short-terms goal (e.g. grasping the coffee-pot), as well as the kinematic level describing the hand posture (e.g. opposing the thumb to the index to take a mug) (Hamilton & Grafton, 2008). Studies on action chains, both in monkeys (Fogassi et al., 2005) and in humans (Cattaneo et al., 2007; Iacoboni et al., 2005), have confirmed that actions have a goal-based hierarchical organization. In terms of the Theory of Event Coding (Hommel, Müsseler, Aschersleben, &

Prinz, 2001), we could say that the absence of an effector-specific effect confirms that actions are primarily coded in terms of their distal features, not of their proximal ones.

Our results do not allow us to clearly address the second issue we discussed in the introduction, i.e. the role played by the presence of a hand near the objects. Indeed, contrary to our expectations, in Experiment 1 we found that when no hand was displayed, RTs were faster than when the hand was present. A possible explanation is based on the higher visual complexity of the scenes in which the hand is presented together with the objects. However, if visual complexity were the determining factor, RTs in the No-Hand condition should always be faster than RTs in all conditions with a hand, regardless of Context, in both Experiments 1 and 2. This was not seen. An alternative explanation suggests the more complex mechanism mentioned above. Our results might reflect the contemporary activation of two different systems, one triggered by the observation of the objects in the context (affordances), the other triggered by the hand together with the object (mirror mechanism). When only one mechanism is activated, as in the No-Hand condition, the responses are faster. Instead, when the hand interacts with the object, the concurrent activation of the two mechanisms slows down response times unless the context strongly activates action (as it happens in the Functional context). As to the Still-Hand condition, given that the fingers were flexed it is possible that it was perceived as slowly moving towards the object (Gangitano, Mottaghy, & Pascual-Leone, 2001). To verify this we submitted a group of 14 participants a rating task. They were required to determine using a 7 point scale whether the hand was moving or not. An ANOVA with the factor Relation manipulated within items was conducted. Results were significant, $F(1, 44) = 4.75$, $MSe = .2496$, $p < .05$, due to the fact that in the No-Relation context the hand was perceived less in motion as in the functional context. This suggests that, even if the hand displayed was in the same position, the context, and particularly the functional one, suggested implied motion of the hand toward the object.

As to the third issue, our results show that participants were not only sensitive to the kind of context (Spatial vs. Functional), but also to the adequacy of the hand posture to the kind of context (see also chapter 5). Importantly, the interaction between context and posture was significant in Experiment 1, when the response effectors were the hands, but not in Experiment 2, when participants provided foot responses; the effector dependency of the effect suggests possible involvement of motor simulation, not a simple association between contexts and hand postures. Specifically, in Experiment 1 in the Functional context observing a hand in the manipulative posture was inhibited, as the slowest RTs obtained with the Manipulation-Hand reveal. It is possible that object recognition was more difficult with the manipulation posture, as it occluded the object a bit. In order to rule out this hypothesis, we performed a rating task asking participants to determine using a 7-point-scale how easily recognizable was the object. Fourteen participants were shown with the images of the objects; each image was presented with the hand both in the manipulative and the functional posture. Participants were asked to what extent they found the object recognition easy. The ANOVA with the within-items factor Manipulation vs. Function did not reveal any difference, $F(1, 22) < 1$; $p = .45$. This result strongly reduces the probability that the delay with manipulation posture is due to the fact that it occludes the object. Even if Manipulation-Hand is the slowest posture in not related pairs and when responses are provided with the foot too, in Experiment 1 in the Spatial context there is no difference in RTs between manipulation and functional postures. This result clearly suggests that in the Spatial context functional postures were inhibited, while they were facilitated in the Functional context. We propose that, when a functional grip is perceived, the effector-specific functional knowledge about the object is retrieved. However, in the spatial context no clear functional use of the object is possible. Given that effector-specific information is activated, this provokes an interference with the hand response, but not with the foot response. The idea that a motor resonance mechanism is activated while observing a functional grip in interaction with an object is in line with recent findings on tool recognition and effector-dependency. Witt, Kemmerer, Linwenauger, and Culham (2010) have recently demonstrated, with a motor interference task (squeezing a rubber ball in one hand), that participants were faster in naming tools with the handle faced away from the ball than facing towards it. Paulus et al. (2009) manipulated the kind of motor interference (hand, foot, and attentional interference) during acquisition of functional knowledge of objects. They found an effector-specific interference during a subsequent object detection task: verbal learning of object function was impaired when a manual motor action was executed. Literature on selection for action is relevant to our results as well. For example, Tipper, Howard, and Jackson (1997) found with a kinematics study that, when two or more objects are presented in a scene, non-target objects evoke competing responses, slowing down the reach. Consistently, in our study the shortest RTs were found when the two objects were functionally compatible, thus the possibility for the non-target object to evoke a competing response was minimized. Our results suggest that participants infer from the context the goal underlying the observed action (Gallese, 2009). In this respect, this behavioural study complements and extends results found with fMRI study by Iacoboni et al. (2005) with a single object. They presented participants with grasping hand actions without a context (e.g., a hand grasping a cup), with the context only (scene containing object) (e.g., a table with objects arranged as before or after having tea) and with grasping hand actions performed in two different contexts, suggesting two different intentions, such as drinking from a cup or putting it away after tea. Results revealed that the context, beyond activating visual information, recruits the motor system as it prepares for situated action: observing both action and context videos activated the parietofrontal circuit for grasping. Our results are compatible with the Predictive Coding model (Kilner, Friston, & Frith, 2007) proposed to account how the Mirror Neuron System would interpret and predict actions. According to this model the observed kinematics of an action can be interpreted at different levels, which are hierarchically organized. At each level of this hierarchy the mismatch between the predicted and the observed activity might lead to a prediction error. As highlighted by the authors, the goal inferred while observing an action is matched with information (priors) received from the context. Similarly, in our study, it is possible that participants observe the functional posture and infer both the short-term goal (grasping the object) and the long-term goal (using it). However, even if the short-term goal doesn't conflict with the context, the long-term goal inferred by the kinematics of the functional posture does not match with the information provided by the Spatial context. This could explain why functional postures are interfered in the Spatial context when responses are provided with the hand.

This result confirms what to our knowledge has not yet been found in a behavioural study, i.e., that functional information is more accessible than manipulation information, and that the activation of both functional and manipulation information is modulated by an action goal, which in this case is made explicit through the context. This result is in line with evidence by van Elk, van Schie, and Bekkering (2008) who demonstrated that objects presented in a location associated to the action goal were recognized earlier than objects in another location (e.g., cup at eye). This result bolsters previous findings showing that both manipulation and function are activated, and that a competition between the two takes place (Jax & Buxbaum, 2010). This competition is rather easily solved when the context disambiguates the situation. Given that typically we interact with objects in a functional way, the competition is more easily solved when the object's function has to be taken into account.

Overall, our study shows that affordances activation is modulated by the context. Other objects in a scene as well as cues related to action/interaction with objects, such as a hand, influence RTs. It remains an open issue, to be investigated in further research, whether and how the two mechanisms interact, one triggered simply by observation of objects and another by observation of others in potential interaction with objects. Brain imaging studies are required, in order to investigate whether two different neural circuits underline object manipulation and object use. Further studies are needed, in order to understand the precise time course of activation of motor information associated to one object, to two objects and to the hand. In relation to context, there is evidence of ventral stream activation to images of man-made artifacts shown in incorrect contexts (Mizelle & Wheaton, 2010a, 2010b, 2011).

The finding that affordance activation is modulated by the context might have important theoretical implications and might contribute to the ongoing debate on automaticity of activation of affordances. Indeed, it is unclear from current evidence and it is still hotly debated in the literature whether the object affordances are activated in an automatic way or whether they are modulated by the task and by the context (e.g., Buxbaum & Kalenine, 2010; Creem-Regehr & Lee, 2005). As clarified before, the automaticity of activation is often inferred from the fact that, even if the task requires processing a given aspect of an object (for example assigning it to a given a category, or deciding its color, etc.), affordances related to other aspects (e.g., grip, orientation) are activated. Studies so far have shown that affordances might be activated independently of the task – for example, affordances related to object grasping might be activated in a categorization task. However, very recent evidence has indicated that the kind of task flexibly modulates the activation of affordances: for example, Pellicano et al. (2010) and Tipper et al. (2006) have shown that affordance effects are not present with tasks implying simple perceptual processing of the stimuli, such as color discrimination tasks, whereas they emerge when the task implies deeper processing, as in categorization and decision on objects shape. Initial neuroimaging evidence further suggests that ventral stream areas for awareness of correct vs. incorrect contexts of man-made artifacts are not active when subjects are not seeking functional relationships between objects (Mizelle & Wheaton, 2010a). This evidence suggests that activation of affordances might be less automatic and more dependent on the task and situation than previously thought. Our evidence bolsters these results, showing that not only the task, but also the context modulates activation of affordances, and that our cognitive systems responds flexibly to changing contexts (see also chapter 5). In other terms, we side with the ‘‘affordance competition hypothesis’’ (Cisek, 2007), according to which a competition between different available action opportunities is activated. In our study we demonstrated that context and relations between objects, as well as the presence of the hand of someone suggesting a specific action goal, can orient this competition.

5. Semantic categorization of conflict objects in different contexts

Compelling evidence in the literature suggests that during conceptual processing of manipulable objects an embodied simulation of appropriate action is evoked (see chapter 2). These findings have been interpreted as evidence that objects are recognized by accessing to action features. However, as discussed in par. 2.5, many objects may be endowed with multiple, contrasting affordances (i.e., conflict objects) associated to different actions in relation of the agent intentions (e.g., a kitchen timer may be clenched with a power grip to move it, whereas it may be pinched with a precision grip to use it). In the present study, the hypothesis that affordance simulation is responsive to the visual scene in which objects are embedded was tested using conflict objects. Participants were asked to categorize object pictures presented in different naturalistic visual contexts that could evoke either move- or use-related actions. Categorization judgments (i.e., natural/artifact) were performed by executing a move- or use-related action (i.e., clench/pinch) on a manipulandum ad-hoc, with response times being assessed as a function of the action-context congruence. Although the actions performed to respond were totally irrelevant in order to successfully accomplish the categorization judgment, responses were significantly faster when actions were compatible with the visual context. This compatibility effect was largely driven by faster pinch responses when objects were presented in use- compared to move-compatible contexts. Thus, the present results confirm a strong influence of the visual scene in which objects are embedded on the emergence of affordance-based compatibility effects during semantic object processing. On the whole, these findings support the hypothesis (advanced in par. 2.5, and confirmed by the study reported in chapter 4) that the activation of different affordances linked to the same tool is flexibly biased by properties relevant for appropriately acting in the context in which the object is embedded.

5.1. Introduction

Evidence from numerous behavioral studies suggests that conceptual processing of manipulable objects is associated with potentiation of action (e.g., Craighero, Bello, Fadiga, & Rizzolatti, 2002; Ellis & Tucker, 2000; Girardi, Lindemann, & Bekkering, 2010; Tucker & Ellis, 1998, 2001). Many of these studies show that conceptual processing of a visually-presented object is facilitated when the motor response required for the task is compatible with the action typically associated with that object, even when that action is task-irrelevant. For example, participants are faster to categorize a small, "pinchable" object (such as a strawberry) as a natural rather than manufactured object when they indicate their categorization choiceby performing a precision (pinch) grip compared to a power (clench) grip on anexperimental apparatus (Tucker & Ellis, 2001). Such stimulus-response compatibility effects have been taken as evidence that conceptual object representationsare composed in part of sensorimotor features associated with object manipulation (e.g., Barsalou, 2008).

Many manipulable objects, however, are associated with several actions, as discussed in par. 2.5. For example, a kitchen timer may be clenched with a power grip to move it, but pinched with a precision grip to use it. Recent studies have shown that object processing may recruit both of these action types (Bub, Masson, & Cree, 2008; Lee, Middleton, Mirman, Kalénine, & Buxbaum, 2013). In one such study, for example, participants were first trained to associate different actions with distinct colors,thenviewed objects whose color signaled the action to be performed on an experimental device. Despite the apparent irrelevance of the motor response to the object identification task, responsesthat were congruent with using or moving the objects (e.g. poking-calculator; clenching-spray bottle) were executed faster than incongruent actions (Bub et al., 2008).

More recently, Jax and Buxbaum (2010, 2013) demonstrated that use- and move-related actions may compete with each other within single objects. In particular, initiation of use actions is slower for objects associated with distinct move-related actions (hereafter, “conflict” objects, e.g. calculator) as compared to objects for which use- and move-related actions are similar (“non-conflict” objects, e.g. drinking glass). This, and associated data indicating that initiation of move-related actions is no slower for conflict- than non-conflict objects, suggests that move-related activations may be relatively rapid, thus interfering with planning of use-related actions. Jax and Buxbaum (2010) proposed that the intention to act on an object triggers a race-like competition between functional and structural responses during action selection. Only functional responses require activation of long-term conceptual representations; thus, structural responses can be activated more quickly than functional responses.

The evidence for two classes of actions associated with a given object raises questions about the factors that may influence the strength and time course of their activation. One possibility is that both types of action are invariably activated during object recognition. Alternatively, and more likely in our view, action activation may be responsive to task goals and context (see Buxbaum & Kalénine, 2010; see also chapter 4). In support of this latter possibility, a recent eye-tracking study demonstrated that activation of move- and use-related competition between objects in a visual arraymaybe accelerated by congruent verbal context (Lee et al., 2013). For instance, cueing of target identity with action sentences such as “he picked up the calculator” or “he used the calculator” accelerated competition between the target (calculator) and distractor objects that are picked up or used similarly, respectively. These data suggest that verbal context may influence the activation of both of these classes of action.

To our knowledge, the question of whether visual scene context may modify activation of move- and use-related actions has not previously been addressed. In the present study, we tested the hypothesis that evocation of move- or use-related actions is indeed responsive to the congruence of the visual context in which objects are presented. To this aim, we used a stimulus-response compatibility paradigm first developed by Tucker and Ellis (2001) and presented conflict objects in move-compatible or use-compatible visual scenes.

5.2. Method

5.2.1. Participants

Twenty-five healthy adults (10 females, mean age =62, SD =6.4, mean education = 15.5 years, SD = 2.7 years) took part in the study.All participants were recruited from the Moss Rehabilitation Research Institute Research Registry(Schwartz, Brecher, Whyte, & Klein, 2005), Philadelphia, USA. They had no history of traumatic brain injury, neurologic disorders, alcohol or drug abuse, or history of psychosis, and achieved at least a score of 27 on the Mini-Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975). They gave informed consent according to guidelines of the Institutional Review Board of Albert Einstein Healthcare Network and were paid $15 for their participation.

5.2.2. Materials and procedure

The study included a baseline experiment designed to control for individual grasping time differences and a main experiment designed to test the influence of visual context on action activation during object semantic processing. Critical stimuli, which were only involved in the main experiment, were colored pictures of 20 manufactured objects associated with different move and use hand postures (e.g., kitchen timer). Objects were presented in either a MOVE environment, in which the visual scene was a context in which the object would be clenched with a power grip (e.g., kitchen timer in drawer) or a USE environment in which the object would be pinched with a precision grip (e.g. kitchen timer on countertop, with food). The association between the MOVE and USE scenes and the gestures evoked by the conflict objects (clench or pinch) was confirmed in a norming study (see the Supplementary materials of chapter 5 in the Appendix of the thesis). There were 40 photographs corresponding to the two visual contexts for each of the 20 conflict objects (see example in Figure 1; the complete list is reported in the Appendix). The scenes represented an office, kitchen, or bathroom. In addition to the critical conflict objects, each scene also contained 4 distractor objects, both man-made and natural (e.g., fruit, vegetables, plants, flowers). A subset of these distractors was used as target objects on filler trials. Thirty natural and 10 man-made distractor objects appeared in both MOVE and USE context pictures. The other natural and man-made distractors objects only appeared in one picture. Distractor objects could afford either power or precision grips or both/none (e.g. plants). For each conflict object, we ensured that the different affordances were represented in equivalent proportions between use and move contexts. For instance for the kitchen timer (Figure 1), all distractor objects would be grasped with a clench, expect for broccoli (use context) and lime (move context) that may afford both clench and pinch grips.

Figure 1. Conflict object (kitchen timer) presented in a move (left) or a use (right) scene.

Sound files corresponding to category labels "natural?" and "man-made?" were recorded by a female native speaker of American English.

The response apparatus consisted of a 4-inch long by1-inch diameter cylinder[2] that afforded both a power grip by clenching the whole cylinder and a precision grip by pinching the tip of the cylinder. The response device was programmed in E-prime to record reaction times when participants squeezed the cylinder (Figure 2).

Figure 2. Experimental set up and use of the response device for clench and pinch grasps.

5.2.3. Baseline experiment

The goals of the baseline experiment was, first, to provide individual mean reaction times for clenching and pinching the device without visual stimuli or a semantic task, and second, to train them to associate the yes/no response to clench/pinch grip. A fixation cross appeared on the screen and participants began the trial by using the index finger of the left hand to press the middle key of a response box positioned in front of them. Participants were always asked to respond with their left hand while their right arm was immobilized in order to allow a future comparison with left hemisphere stroke patients. Thus, as left hemisphere stroke patients frequently have reduced right arm mobility, the mobility of the right limb of each participant was limited with an arm sling. Participants reached to and grasped the apparatus with either a pinch or a clench in response to "YES" and "NO" verbal cues. They were asked

---

[2] Thank to Rob Ellis and Mike Tucker for providing the response apparatus.

to keep the key depressed until presentation of a verbal cue. After a variable delay of between 1500 and 3000ms, the word “YES” or “NO” was delivered through speakers. As soon as they heard the word, participants released the key and grasped the response device, which was a cylinder mounted vertically in a wooden support, located 13 inches from the response box (Figure 2). Participants were randomly assigned to one of 2 groups. Participants in Group 1 had to clench the cylinder when they heard “YES” and pinch it when they heard “NO”, whereas those in Group 2 had to perform the opposite mapping. They were instructed to respond as quickly and accurately as possible. Initiation times (word offset to liftoff) and transport times (liftoff to cylinder contact) were recorded automatically in E-Prime. Accuracy was coded online by the experimenter (c = clench, p = pinch, n = none). Gesture videotaping was used for offline accuracy checking. There were 10 practice trials with reaction time and accuracy feedback (5 “YES” and 5 “NO” in random order), followed by 24 baseline trials where no feedback was provided (12 “YES” and 12 “NO” in random order).

Correct movement initiation times were computed as a function of the Gesture performed on the device in response to “yes” and “no” labels (pinch or clench). After the 10 practice trials, all participants were 100% correct on the 24 baseline trials. One participant was particularly slow in initiation times (3SD below the group mean) and was excluded from further analysis. For the remaining 24 participants, initiation times that were either shorter than 200ms or longer than 3 standard deviations below the mean of the group in the pinch and clench conditions were considered outlier trials and removed from the data (1.5%).

5.2.4. Main experiment

On each trial, a fixation cross appeared in the center of the screen. Participants began each trial by pressing and holding the middle key of the response box with the index finger of their left hand. As in the Baseline experiment, the mobility of the right limb of each participant was limited with an arm sling. Immediately after pressing the key, the scene picture appeared on the screen. After a 1250 ms delay, a red box appeared around the target or one of the four distractors. Location of target and distractors was randomized. Simultaneously, they heard an auditory cue, either “natural?” or “man-made?”. Participants then indicated whether the category label matched the object in the box by using the response device to indicatea “YES” or “NO” response. This was accomplished by releasing the response box key and reaching to grasp the cylinder with either a clench or a pinch. The picture disappeared when the start button was released. Participants in Group 1 clenched the device to respond “YES” and pinched it to respond “NO”, whereas participants in Group 2 performed with a reversed mapping. They were instructed to respondas quickly and accurately as possible. Movement initiation and transport times were recorded automatically in E-Prime. Accuracy was coded online by an experimenter (c=clench, p=pinch, n=none). Gesture videotaping was used for offline accuracy checking.

Participants performed 12 practice trials with feedback on accuracy, using pictures that were not displayed in the experiment. The experimental block contained 120 trials. Indeed, each of the 40 scenes was presented 3 times in randomized order resulting in 120 experimental trials. On 40 critical trials, the red box appeared around the conflict object in the scene. For the remaining filler trials, the box appeared around a natural distractor object on 60 trials and around a man-made distractor on 20 trials. Thus, the target object was natural and man-made on an equal number of trials. Each scene was repeated 3 times: once with the conflict object as target and twice with a distractor object as target. Since each conflict object was the target twice, once in the MOVE and once in the USE scene, the number of repetition of distractor objects as target was varied among filler trials so that overall, object category, object repetition across pictures, and target repetition were not informative in predicting which object in the scene would be the target on a given trial.

On half of the trials, the target object was coupled with the label “natural?” and on the other half coupled with the label “manmade?”. Repeated target objects could be associated with the same label or a different label on both occurrences. Hence, when a given object was the target for the second time, the likelihood of hearing a repeated or new label was equivalent.

5.2.5. Data Analysis and results

In the baseline experiment, individual initiation times[3] for pinch and clench were calculated and used to reduce between-subject variability in the data from the main experiment (see below; also see Supplemental Materials for additional detail).

In the main experiment, data were trimmed and adjusted as follows. First, participants who were at chance level in at least one condition (accuracy < 75% according to binomial probability) were excluded from further analysis (N=3). One participant was particularly slow in baseline initiation times (3SD below the group mean) and was also excluded. Thus, the final data set included 21 participants. Second, analyses on initiation times were conducted after removing incorrect trials (4% data) (Notrialswere excluded for being shorter than 200ms or longer than 3 standard deviations fromthe group meanin the corresponding condition). Finally, adjusted initiation times were computed at the individual level in each condition by subtracting initiation baseline times for pinch and clench fromthe respective initiation times in the main experiment.

A 2 x 2 Analysis of Variance was conducted on mean adjusted initiation times from critical trials with Gesture (pinch, clench) and Context (MOVE, USE) as within-subject (F_1) or within-item (F_2) factors. Distribution normality and variance homogeneity were verified.

---

[3] Our analyses focused on movement initiation times since object action-related features have been shown to affect grasp planning prior to movement execution (e.g., Bub, Masson, & Cree, 2008; Jax & Buxbaum, 2010; Girardi, Lindemann, & Bekkering, 2010). Nonetheless, note that we did not observe any effect of the variables of interest on transport times (*all ps* > .25).

Errors were extremely rare: of the total of 840 trials run by all subjects in the experiment, only 35 trials had errors. Error distribution was highly skewed and not suited to a similar analysis as the one conducted on initiation times. Nevertheless, proportions of correct responses between conditions were compared using chi-square.

Initiation times

There was no main effect of Gesture [$F1_{(1,20)}$= 0.31, R^2=0.02 , p= .58; $F2_{(1,19)}$= 1.08, R^2= 0.05 , p= .31] or Context [$F1_{(1,20)}$= 1.15, R^2= 0.05 , p = .29; $F2_{(1,19)}$= 3.18, R^2=0.14, p = .09]. Critically, the Gesture x Context interaction was significant in both the by-subject [$F1_{(1,20)}$= 4.8, R^2= 0.19, p = .04] and by-item [$F2_{(1,19)}$= 6.31, R^2=0.25 , p =.02] analyses. As shownin Figure 3, there was a greater advantage of the use context compared to themove context in the pinch gesture condition compared to the clench gesture condition.

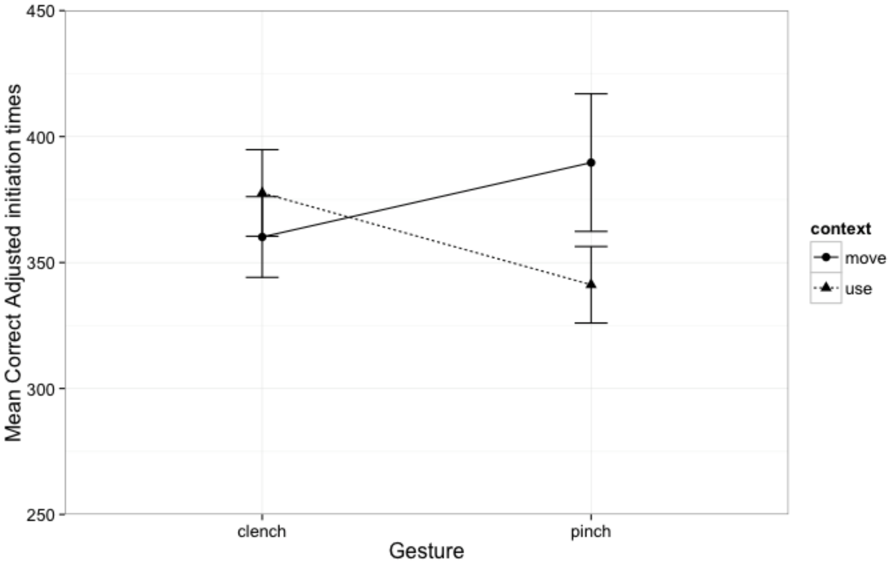


Figure 3. Mean correct adjusted initiation times (and standard errors) for clench and pinch categorization responses as a function of context (move/use).

Post-hoc comparisons of the by-item analysis indicated that the interaction between Gesture and Context was likely due to shorter initiation times in the use than in the move context for pinch (t = 2.74, p = .01), whereas there was no difference between use and move contexts for clench (t = -0.2388, p = 0.81). None of the post-hoc tests reached significance in the by-subject analysis, though the results were consistent with those demonstrated in the by-item analysis (t=1.6891, p = .10 between move and use contexts for pinch; t=-0.5447, p = .59 between move and use contexts for clench).

Correct responses

Chi-square test on accuracy data did not show any significant different in proportion of correct responses between the four Gesture x Context conditions (χ^2= 6.65, p = .08). As can be seen in Table 1, the number of correct responses was numerically inferior for pinch responses in the use context, but this was anecdotal considering the absence of significant difference between conditions and the very limited number of errors. Consequently, accuracy data will not be further discussed.

Context	*Gesture*	*Number of correct responses*	*Proportion of correct responses*
Move	Clench	203	96.2%
Use	Clench	204	96.6%
Move	Pinch	202	97.6%
Use	Pinch	196	92.9%

Table 1. Number and percentage of correct responses in each condition.

5.3. Discussion

We report context-dependent compatibility effects between the motor responses performed during object semantic categorization and the action evoked by the object in a given visual context. Prior demonstrations indicate that action evocation during object processing may be modulated by verbal context (Costantini, Ambrosini, Scorolli, & Borghi, 2011; Lee et al., 2013), affordances of distractor objects (Caligiore, Borghi, Parisi, Ellis, Cangelosi, et al., 2013; Ellis, Tucker, Symes, & Vainio, 2007; Pavese & Buxbaum, 2002; Tipper, Howard, &

Jackson, 1997), and relationships to other objects or agents (Borghi, Flumini, Natraj, & Wheaton, 2012; Ellis et al., 2013; Girardi et al., 2010; Natraj et al., 2013; Yoon, Humphreys, & Riddoch, 2010; see chapter 4). The present data extend such findings by demonstrating that activation of move- and use-related gestures during semantic object processing may additionally be modulated by the visual environment in which objects are presented. The visual environments used here were composed of 5 objects naturally displayed on a furnished room background. The fact that we observed compatibility effects with complex visual contexts provides additional ecological validity to action evocation phenomena during object processing and reinforces the idea that affordances are flexibly activated in natural environmental conditions. In addition, the data suggest that the contextual modulation observed in the present study is the outcome of a global visual processing of the scene that can be distinguished from the influence of single object affordances. Although distractor objects may have also activated the actions associated with them, their affordances were equivalent between contextual conditions. Thus, the context-dependent compatibility effects reported here are likely related to the meaning conveyed by the array and by the action intention that emerges from the visual scene.

The existence of such effects raises the challenge of identifying when and how visual context influences compatibility effects in the cascade of perceptual and motor processes. It is well-recognized that preparation of a motor response orients attention towards action-relevant features and may facilitate visual processing of stimuli that are congruent with that action (the "motor-visual attention" effect, e.g., Allport, 1987; Bekkering & Neggers, 2002; Botvinick, Buxbaum, Bylsma, & Jax, 2009; Craighero, Fadiga, Rizzolatti, & Umiltà, 1999; Hannus, Cornelissen, Lindemann, & Bekkering, 2005; Pavese & Buxbaum, 2002). Preparing a clench or a pinch may facilitate processing of distinct conflict object features (e.g., the entire kitchen timer or the timer dial, respectively). Consequently, faster object processing may be observed when the features highlighted by response preparation are compatible with one of the actions evoked by the object. At the same time, visual object processing appears to activate action representations, even in tasks not involving a motor response (e.g., Kalénine, Mirman, Middleton, & Buxbaum, 2012; Lee et al., 2013; Myung et al., 2010). Additionally, visual context influences object processing (e.g., Gronau, Neta, & Bar, 2008; Mudrik, Lamy, & Deouell, 2010). In objects associated with more than one action, such as the conflict objects presented here, we may speculate that the visual context serves to amplify the action associated with it (e.g., Wurm, von Cramon, & Schubotz, 2012). In an iterative manner, this "bottom up" facilitation of an object-related action by the contextmay resonate with the intention-driven facilitation of action by the planned action (see Chambon et al., 2011; Shen & Paré, 2011 for related accounts). Further investigations of context-dependent compatibility effects could potentially employ variations in the timing of experimental perceptual and motor events to specify how environment-based and intention-based processes interact during object processing.

Another main issue concerns the stage of object processing at which the observed context-relevant action effects emerge. While most studies on effect of context on action evocation from objects have induced "deep" object processing by using semantic decision tasks, a few studies have contrasted different processing levels and showed that affordances are not activated when the task requires shallow object processing (e.g., color judgments; Pellicano, Iani, Borghi, Rubichi, & Nicoletti, 2010; Tipper, Paul, & Hayes, 2006). One possibility is that context-relevant action modulation arises before conceptual object processing is completed, perhaps on the basis of associations between the target object, context and actions. Context-dependent activation of object affordances could then impact semantic processing while emerging from earlier (pre-conceptual) stages of perceptual processing. Alternatively, object-related actions might be automatically evoked during early processing stages (Goslin, Dixon, Fischer, Cangelosi, & Ellis, 2012) and context modulation might arise later on during conceptual processing. Context could work as a late filter, which would enhance relevant action features and turn off irrelevant ones. Regardless, results overall suggest that all action features are not systematically integrated to object concepts and that context and goals play a decisive role in this integration.

The affordance-based compatibility effects observed in the present study were largely driven by faster initiation of use-related actions when the object was presented in a use-compatible context compared to a move-compatible context. In contrast, initiation of move-related actions did not appear sensitive to visual context. This asymmetry could have been related to the fact that participants were required to respond with their left hand. Indeed, manual asymmetries have been reported in visually primed grasping (Vainio, Ellis, Tucker & Symes, 2006). However, manual differences were observed in the opposite direction, with an absence of object size-grip type compatibility effects when precision grip responses were performed with the left hand. A reduction of affordance effects has also been recently observed when right-handed participants used their left hand to execute memorized instructions on objects with handles that were spatially congruent or incongruent with the dominant hand (Apel, Cangelosi, Ellis, Goslin, & Fischer, 2013), suggesting that compatibility effects may be more difficult to observe when responses are performed with the left hand. Moreover, while manual asymmetries could possibly account for a main effect of grip type in the present paradigm (which we did not observe), they could not explain the observed context effects on precision grips. If compatibility effects are overall enhanced/reduced for precision grips depending on the response hand, this should affect move and use context conditions equally. Thus, reasons for the asymmetry reported here remain uncertain, but several potential explanations can be formulated. First, pinch grasps might be more context-specific than clench grasps. For instance, pinch might be more associated with opening a bottle with a corkscrew than clench is associated with moving this item. Second, use-related actions are often preceded by move-related actions, particularly in naturalistic environments. For example, one must first pick up a corkscrew with a clench prior to using it with a pinch. Accordingly, clenches may be equally triggered by use- and move-compatible contexts while pinches would be more strongly activated in use-compatible contexts. Finally, at the action planning level, one could consider the clench hand posture less specified than the pinch hand posture. In other words, the first phase of any grasping movement (pinch or clench) could start in some cases with a clench-like posture, and the position of the different fingers that are opposed to the thumb could require further determination. This possibility accords with neurophysiological data showing additional fronto-parietal recruitment for the control of precision grips compared to power grips (Ehrsson et al., 2001). Hence, clench action initiation would be as relevant for use-compatible and move-compatible environments and context would show little influence on clench responses.

In summary, the present study confirms the influence of visual scene on stimulus-response compatibility effects during semantic object processing (see also chapter 4). This finding brings additional support to action models that consider both action sub-types and context as key determinants for understanding interactions between object and action processing (e.g., Buxbaum & Kalénine, 2010). Moreover, this finding may have strong implications for object processing in naturalistic tasks where objects are perceived in their natural visual environments.

6. Activation of affordances by visual objects and their names

Much research in the literature on embodied language has been collected with the aim of demonstrating the similarity between the simulation triggered by language a simulation the triggered by objects or actions processing. This evidence is compelling (see chapter 3), however, as proposed in par. 3.4, there might be differences that remains unexplored. In order to better understand this issue, this study investigated affordance-based compatibility effects with a categorization task (i.e., artificial/natural) in two experiments, identical in all but the stimuli used: in Experiment 1 the categorization was performed on pictures of objects, in Experiment 2 the categorization was performed on the names of the same objects of Experiment 1. Furthermore, as the real-time dynamics of affordance-based compatibility effects are currently unknown, the time course of the responses was tracked using MouseTracker software. Participants were required to report their choice using either a big mouse (requiring a power grip, a hand-posture compatible with the grasping of big objects) or a small mouse (requiring a precision grip, a hand-posture compatible with the grasping of small objects). The results of Experiment 1 showed a compatibility effect between the grip required by the mouse and the grip elicited by picture of objects, even if it was irrelevant to the task. The results of Experiment 2 on linguistic stimuli referring to the same objects of the previous experiment failed to exactly reproduce this effect, nevertheless a compatibility effect mediated by the target-word category (i.e., artificial/natural) was observed. On the whole, the present study confirmed at the behavioural level the hypothesis, advanced in par. 3.4, that visual and linguistic stimuli trigger simulation processes that show not only similarities, but also differences as well. These finding are predicted by theories of reuse and neural exploitation, which suggest that language recruits structures and mechanisms characterizing the sensory-motor system to modify and build on them.

6.1. Introduction

The ability to grasp objects in the appropriate way, using the adequate kind of grip and timing of opening and closing the hand, represents one of the more complex and sophisticated motor abilities humans are endowed with, as its progressive refinement during development testifies. While grasping has been mostly and extensively studied in the framework of motor control (Oztop & Arbib, 2002; Shadmehr et al., 2010), in the last years the interest for grasping actions and grasping postures has risen in the literature on visuo-motor transformations and affordances. Building on the notion of *affordance* proposed by (Gibson, 1979), according to which objects invite organisms to act, recent studies have shown that observing objects activates possible motor responses.

Studies on action preparation have provided evidence of a shifting of attention toward the action-relevant property of the objects, leading to compatibility effect between the hand posture used to respond and some characteristics of the stimuli. For example, Craighero et al., (2002) asked participants to prepare to grasp a bar that could have different orientations; when the picture of a hand was displayed, they had to grasp the bar as fast as possible. Results revealed a compatibility effect between the orientation of the bar (clockwise, counter-clockwise) and the grasping hand final position. Most relevant to the present work are studies on the *affordance-based compatibility effect* between the object size and the kind of grip used to respond. In an influential work, Tucker & Ellis (2001) instructed participants to categorize as natural or artifact real objects differing in size by mimicking either a precision or a power grip with a customized device. The compatibility effect they found between the object size (large, as apple, hammer, or small, as cherry, nail) and the grip used to respond (power or precision) indicates that observing objects potentiates their affordances. Importantly, the action-relevant dimension (i.e., size) influenced response times even if it was not relevant to the categorization task. Further experiments with briefly presented objects (Tucker & Ellis, 2004) revealed that the effects are maintained even when the object disappears: the objects do not need to be visible during the response selection; additionally, also the presentation of the object name exerted the effects. Thus, they seem to be due to long-term associations between objects and actions. Further recent studies have investigated the compatibility effects induced by the context. In some studies the context was given by the presence of a hand in potential interaction with the object and by another object which might be functionally connected to the first or not (e.g., Borghi et al., 2012; Natraj et al., 2012, 2013; Yoon et al., 2010; see chapter 4). In a recent work, Kalenine et al. (2013) used conflict objects, i.e., objects that had different affordances related to use and to movement (Jax and Buxbaum, 2010; Lee et al., 2013; see also Bub et al., 2008; Creem & Proffitt, 2001) and found that the compatibility effect between hand posture (precision vs. power) and objects was modulated by the visual scene in which objects were embedded, eliciting either use-related or move-relate actions (see chapter 5).

Further studies on visuo-motor priming investigated the effect of showing different hand postures on subsequent tasks. Vogt et al. (2003) and Bruzzo et al. (2008) manipulated the perspective of the hand prime demonstrating its effect on grasping and categorization tasks. More relevant to the present work are studies on compatibility effects between the hand posture and the object size. Borghi et al. (2007) asked participants to categorize pictures of objects differing in size into artifact and natural objects by pressing two different keys on the keyboard; the stimuli were primed by pictures of hands displaying either a precision or a power grip. A compatibility effect between the hand prime (power, precision) and the object size (large, small) was found, provided that before the experiment participants mimicked the displayed hand postures. The compatibility effect between the hand prime and the size of the targets was replicated and extended by Vainio et al. (2008) with dynamic hand stimuli.

As this brief overview shows, a number of experiments have demonstrated the presence of an interaction between the hand posture and objects action-based characteristics, particularly size. In some studies different hand postures were used to provide the response, in other studies different hand postures were displayed as primes, or embedded in a scene. In all cases compatibility effects were found (see chapter 2). Overall, the evidence on the interaction between hand posture and object size raises questions about the factors that influence the involved processes and their time course. However, to our knowledge no study so far has focused on how the effects of the compatibility vs. incompatibility between the information derived from the object and from the hand posture unfolds in time and is reflected in an explicit movement. In the present study we intend to investigate the time course of the congruency effect. We aim to assess when does the conflict between the information derived from the posture of the hand used and the object size come into play, and how it is reflected in overt movements. In addition, we intend to verify the role played by a distractor compatible in size with the target in deviating the trajectory to reach for the object.

With respect to the current literature, our work presents several novelties. The first is that it investigates static hand postures rather than full prehension movements. In real life we often use static precision and power postures: we hold nails and nuts, coconuts and umbrellas, etc. In the studies on affordance-based compatibility participants were either required to simply press a button on the keyboard (e.g., Borghi et al., 2007, 2012; Riggio et al., 2008; Fagioli et al., 2007) or alternatively they were asked to squeeze the device mimicking a power grip and to press a switch mimicking a precision grip (e.g., Tucker & Ellis, 2001); this resembles more to the experience of squeezing some fruit or vegetable, while artifacts are often hard and not squeezable (Anelli et al., 2010). In the present study we used a mouse that participants held and dragged with their dominant hand; the mouse could be small, graspable with a precision grip, or large, graspable with a power one. Compared to previous study, in the current one the precision grip was not characterized by the opposition between the thumb and the index finger but was the one required to hold a small mouse; that is, closer to the power grip in terms of fingers configuration, but it was adapted to a smaller object.

Using the mouse has a further advantage, which represents the second novelty of our study. In the experiments conducted by Tucker & Ellis (2001, 2004) the hand posture was relevant for the response to provide, while in studies with hand primes (e.g., Borghi et al., 2007) it was not, since a simple key-press response was required. In our study, participants' motor response consists in moving the mouse in different directions regardless of how the mouse is grasped; still we manipulate participants’ prehension. Participants saw a cue-word on the screen (“artificial” vs. “natural”) and were instructed to drag the mouse to two different locations to decide which of the displayed images represented an object of the category indicated by the word, thus mimicking the reaching of the object. Objects were either natural or artificial, differing in size (graspable with a precision vs. a power grip). Similar to the original study by (Tucker & Ellis, 2001), participants' hand posture was irrelevant to the task but was manipulated - by providing participants with a small or big mouse - to unveil compatibility effects with stimuli dimensions (e.g., size of target and distractors).

The third important novelty of the present work is that we used a continuous measure of performance: we tracked participants' mouse trajectories during the choice. This procedure is increasingly used to study the real-time dynamics of decision and is particularly useful to reveal the fine-grained effect of conflicting cognitive processes (Barca & Pezzulo, 2012; Freeman & Ambady, 2010; Song & Nakayama, 2009; Spivey, 2007). In our study, the mouse-tracking procedure allows us to investigate the effect played by congruent or by conflicting information as they unfold in time and are reflected in hand movements. The trajectory followed while moving the mouse to reach for the object can indeed inform us on the effects of congruent or conflict information on the response selection. To our knowledge the only study investigating similar issues is an EEG experiment conducted by Goslin et al. (2012) on compatibility effects between the response hand and the handle location of objects. Results revealed that visual processing and motor information are integrated very early, before 200 ms of stimulus onset (see also Bub & Masson, 2010, on the dynamics of aligned effects elicited by handled objects).

If participants are sensitive to static hand postures and this sensitivity is reflected in the trajectory followed while reaching for the object, we predict a compatibility effect between the grip required by the mouse (small, large) and the grip elicited by the object. Furthermore, we predict an inverse compatibility effect between the target and the distractor. The degree of uncertainty of the reaching trajectory should be higher when the target and the distractor are compatible in size than when they are not, and when the mouse and the distractor are compatible in size than when they are not.

6.2. Experiment 1

6.2.1. Method

6.2.1.1. Participants

Twenty-four under graduated students from the University of Bologna (9 males; mean age = 21.25 (2.88); all Italian monolingual and right-handed by self-report) participated for course credits. All participants had normal or corrected-to-normal vision and were naive as to the purposes of the experiment.

6.2.1.2. Materials

Participants performed a semantic categorization task. They were presented with color images of everyday objects. Sixteen pictures were used, 8 depicting natural objects and 8 depicting artifacts. Within each category, 4 objects afforded a power grip (e.g., 'courgette') and the other 4 afforded a precision grip (e.g., 'nut'). Two pictures were presented in the upper corners of the screen, one depicting an artifact and one depicting a natural object (i.e., one target and one distractor). Pictures were preceded by the central presentation of the word 'ARTIFICIAL' or 'NATURAL', which instructed the participants on which item they have to click with the mouse to respond correctly. Stimuli were combined in 64 pairs presented twice, once for the categorization of the 'ARTIFICIAL' target, once for the categorization of the 'NATURAL' target. Objects images were scaled to preserve the real size differences, and always presented in a 200x250 pixels box, color print on white background (see Fig. 1).

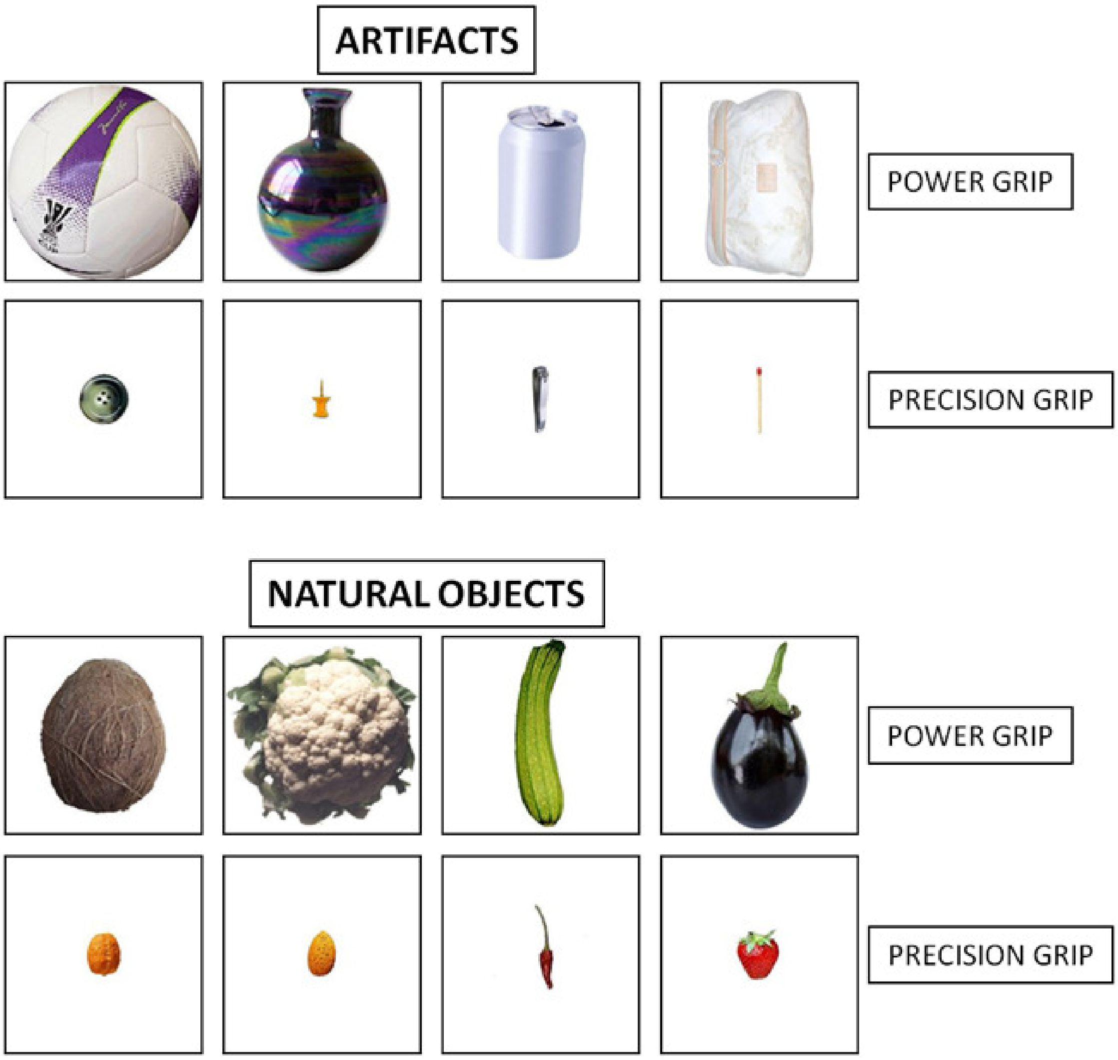


Figure 1. Sample stimuli used in the experiment

6.2.1.3. Design and procedure

Participants sat 60 cm from the computer screen, with their right hand placed over the mouse they found in front of them, already positioned to begin the experiment.

Each trial began with the appearance of the 'START' button (displayed at the bottom-centre of the screen) that remained on the screen until a single mouse-click was performed on it. The cue-word ('ARTIFICIAL' or 'NATURAL') was displayed after the mouse-click at the centre of the screen for 1500 ms (50% of the trials were preceded by the word 'ARTIFICIAL', the other half by the word 'NATURAL'). Then, the two experimental stimuli appeared on the top-left and top-right corners of the screen and remained on the screen until a response was made by mouse-clicking one of them. Participants were instructed to decide which among the two stimuli matched the category indicated by the cue-word (see Fig. 2). They were asked to respond as quickly and accurately as possible. A feedback message was provided in case of incorrect response (a red 'X' at the centre of the screen). To avoid any repetition effect, pairs of stimuli were presented in random order.

Stimuli were presented in two blocks where mouse dimension was also manipulated, so that in one block participants were asked to respond using a big mouse (length 11 cm x width 6 cm x height 3.5 cm) and in the other block they had to use a small mouse (length 7 cm x width 3.5 cm x height 2.2 cm). Each block consisted of 128 experimental trials preceded by 4 training trials, so each participant responded overall to 8 training trials and 256 experimental trials.

In each experiment the following factors were manipulated: Response Device (big mouse/small mouse), Target Type (artifact/natural), Target Dimension (big/small), Distractor Dimension (big/small).

MouseTracker software was used for stimulus presentation and data collection (Freeman & Ambady, 2010): an open-source software package, freely available at the web page http://www.dartmouth.edu/~freemanlab/mousetracker/dl.htm, which allowed us to record and analyze the continuous stream x-y coordinates of the hand movements performed by participants who decided among alternative responses. Thus, precise characterizations of both temporal and spatial dynamics of the mouse trajectories were available to be analyzed.

Individual trajectories were first rescaled to a standard coordinate space and then normalized into 101 time steps, see also (Freeman & Ambady, 2010). Data were then exported in Microsoft Office Excel using the utilities included in the MouseTracker package, then trimmed in Excel, while all the ANOVAs were performed in StatSoft STATISTICA 6.0.

6.2.2. Results

Accuracy, Initiation Time and Trajectory time

We removed 0.98% of trials as errors. This very low rate reveals that the task was easy to perform. Total trajectories times exceeding 2 standard deviations from each participant's mean were excluded from the analysis, leading to the removal of additional 8.27% of the data. The total trimming was of the 9.25% of trials.

The remaining data were entered into a 2 x 2 x 2 x 2 within subjects ANOVA, with the factors Response Device (Big mouse vs. Small mouse), Target Type (Artifact vs. Natural), Target Dimension (Big vs. Small) and Distractor Dimension (Big vs. Small). Where possible, interaction effects were evaluated with Newman-Keuls post-hoc test ($p < .05$).

The ANOVA on *Initiation times* showed significant main effects of Response Device and Target Type. The time to initiate the movement was longer when using the Big mouse than when using the Small mouse (422 ms and 290 ms, respectively; $F(1, 23) = 58.19$, $MSe = 29074.1$, $p < .001$); and for categorizing Natural than Artifact items (364 ms and 348 ms, respectively; $F(1, 23) = 9.13$, $MSe = 2664.66$, $p < .01$). The three-way interaction between the factors Response Device, Target Type and Target Dimension ($F(3, 92) = 4.135$, $MSe = 1623.09$, $p = .06$) almost reached significance, showing that different patterns for the two devices were modulated by the target category too (Mouse big / Artifact: Target big M = 431 ms - Target small M = 432 ms, Natural: Target big M = 405 ms - Target small M = 421 ms; Mouse small / Artifact: Target big M = 290 ms, Target small M = 302 ms, Natural: Target big M = 287 ms - Target small M = 280 ms).

No other main effects or interactions were significant.

The different size of response devices implies also a difference in their weight and friction, which might be partly responsible for the observed effect on initiation time. Given such side effect, no theoretical conclusion will be drawn on the effect of mouse dimension on temporal measures of the response.

The analyses on total Trajectory times, which here are the overall Response Times, revealed as significant the main effect of Target Type ($F(1, 23) = 15.39$, $MSe = 8805.26$, $p < .001$) with faster response for Natural ($M = 1312$ ms) than Artifact ($M = 1350$ ms) items.

No other main effects or interactions were significant.

Trajectory spatial analysis

The *Area Under the Curve* (AUC) is a measure of spatial attraction towards the opposite response alternative, i.e., the distractor item the influence of which has to be suppressed to give the correct response. Positive AUC mean values indicate that the mouse trajectory is above the idealized straight line between the START button and the target-object. Thus, the AUC values measure how much the hand movement is attracted toward the distractor item, indexing the indecision during the choice.

The ANOVA on AUC demonstrated the main effects of the factors Response Device and Target Type. Two interactions were reliable as well, whereas another almost reached significance.

The Response Device main effect showed that the Big mouse mean AUC ($M = 0.33$) was smaller than the Small mouse mean AUC ($M = 0.42$ ms), $F(1, 23) = 7.73$, $MSe = 0.10106$, $p < .05$, probably due to the lightness of the small mouse (more subject to involuntary deviations). The factor Target Type, $F(1, 23) = 12.03$, $MSe = 0.07857$, $p < .01$, was significant because the Natural items AUC ($M = 0.32$) was smaller than the Artifacts AUC ($M = 0.42$).

The expected interaction of Response Device and Target Dimension was significant, $F(2, 46) = 11.06$, $MSe = 0.04411$, $p < .01$, confirming our prediction about a compatibility effect between the hand posture and the dimension of the target,. Indeed, when participants were using the Big mouse it was easier to go straight over a Big target-object (*M* = 0.29) than a Small one (*M* = 0.37) (Newman-Keuls *p* < .05), while the opposite was true when using the Small mouse, with greater AUC for Big target-object (*M* = 0.45) than for Small ones (*M* = 0.38) (Newman-Keuls *p* <.05, see Fig. 2 and Fig. 3, Graph a).

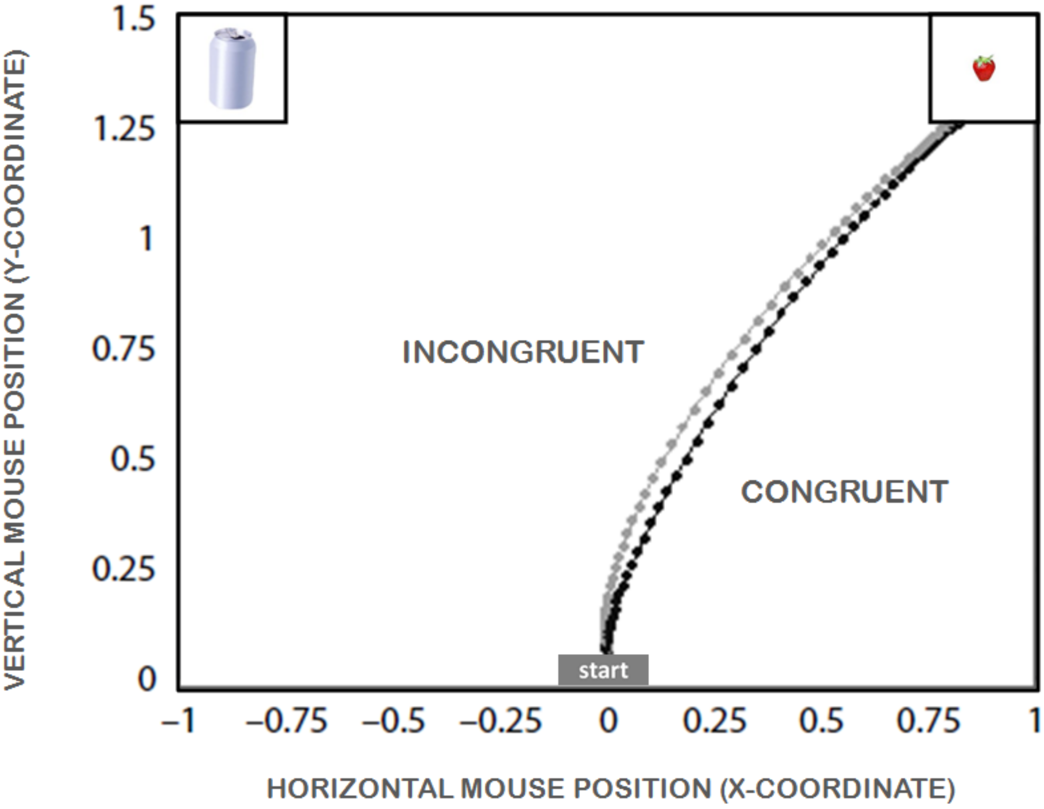


Figure 2. Experiment 1 - congruent (black line) vs. incongruent (grey line) trials mean trajectories

The two-way interaction between Response Device and Distractor Dimension was significant as well, *F*(2, 46) = 8.73, *MSe* = 0.03408, *p* <.01. It showed that an effect reciprocal to the one described above for the target stimuli was present for the distractor items, as when participants were using the Big mouse trajectories drew a greater AUC for Big (*M* = 0.36) than for Small distractor (*M* = 0.29) (Newman-Keuls *p* <.05), while the opposite was true when using the Small mouse (Big distractor *M* = 0.44, Small distractor *M* = 0.40) (Newman-Keuls *p* =.07) (see Fig. 3, Graph b).

Finally, the Target Type x Target Dimension interaction, *F*(2, 46) = 3.59, *MSe* = 0.03071, *p* =.07, almost reached significance. It showed that when the target stimulus was an Artifact it was easier to go straight to a Small target (*M* = 0.41) than to a Big one (*M* = 0.44), while the opposite was true for Natural target stimuli (Big target *M* = 0.30, Small target *M* = 0.34).

No other main effects or interactions were significant.

The *Maximum Deviation* (MD) is a further measure of spatial attraction to the opposite response alternative. It determinates which of the points in the trajectory is the most far from the idealized straight line between the START button and the target-object by measuring the perpendicular line from that point to the idealized straight line. As for the AUC, positive MD mean values indicate that the mouse trajectory is above the idealized straight trajectory, so the MD values index again how much the hand movement is attracted toward the distractor item. The ANOVA on MD demonstrated the main effects of the factors Response Device and Target Type. Three interactions were reliable as well.

The factor Response Device was significant due to the Big mouse mean MD (*M* = 0.18) being smaller than the Small mouse mean MD (*M* = 0.23 ms), *F*(1, 23) = 6.03, *MSe* = 0.0318, *p* <.05 (probably for the lightness of the small mouse). The factor Target Type, *F*(1, 23) = 14.48, *MSe* = 0.01748, *p* <.001, showed that the Natural items MD (*M* = 0.18) was smaller than the Artifacts MD (*M* = 0.23).

The interaction of the factors Response Device and Target Dimension was significant in this measure too, *F*(2, 46) = 17.01, *MSe* = 0.01307, *p* <.001. This confirmed again our prediction of a facilitation effect in case of compatibility between hand posture and target-object dimension. Indeed, when participants used the Big mouse it was easier to go straight over a Big target-object (*M* = 0.16) than over a Small one (*M* = 0.21) (Newman-Keuls *p* < .01); the opposite was true when using the Small mouse, with greater AUC for Big target-object (*M* = 0.25) than for Small ones (*M* = 0.21) (Newman-Keuls *p* <.05) (see Fig. 3, Graph c).

The Target Type x Target Dimension interaction, *F*(2, 46) = 5.13, *MSe* = 0.00861, *p* <.05, also reached significance. When the target stimulus was an Artifact it was easier to go straight to a Small target (*M* = 0.22) than to a Big one (*M* = 0.24), and the opposite was true for Natural target stimuli (Big target *M* = 0.16, Small target *M* = 0.19) (Newman-Keuls *p* < .05).

Finally, the interaction between Response Device and Distractor Dimension was significant as well, *F*(2, 46) = 10.56, *MSe* = 0.01004, *p*<.01, as when participants were using the Big mouse trajectories drew a greater MD for Big (*M* = 0.20) than for Small distractor (*M* = 0.17) (Newman-Keuls *p* < .05), while the opposite was true when using the Small mouse (Big distractor *M* = 0.21, Small distractor *M* = 0.24) (Newman-Keuls *p* < .05). Thus, it showed a compatibility effect for the distractor items too (see Fig. 3, Graph d).

No other main effects or interactions were significant.

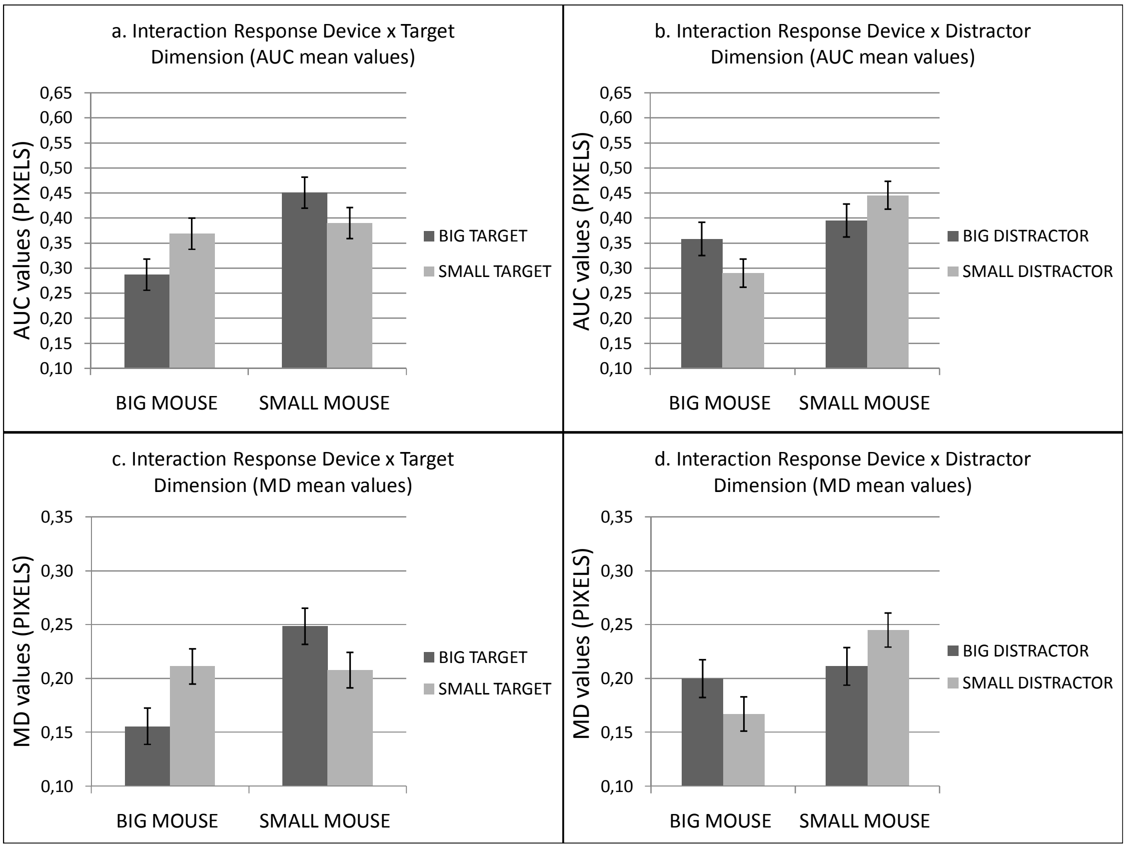


Figure 3. Experiment 1, all the results

6.2.3. Discussion

The continuous recording of mouse movements allowed us to study the influence of participants' hand postures and target object size. We found that response trajectories were affected by the power grip required by a big mouse and the precision grip required by the small mouse. The results confirmed our prediction of a compatibility effect between mouse dimension and stimuli. We found the predicted compatibility effect in both MD and AUC: the trajectories followed by participants were more direct, revealing less uncertainty in the decisional process, when the dimension of the mouse and the object size matched. In addition, we found a clear influence of the distractor size on the response. When the dimension of the mouse matched with that of the distractor, responses were more uncertain, as the interaction on MD indicated. Furthermore, the degree of uncertainty as revealed by AUC was higher when the object and the distractor size matched than when they did not. Overall, these results reveal that participants were sensitive to the static hand posture they used, and that the compatibility effect was present even if the object size was neither relevant to the task, a semantic categorization, nor to the response provided, consisting in moving the mouse in a given direction. Crucially, to our knowledge this is the first evidence of compatibility effect between object size and static hand posture; importantly, the effect is obtained analyzing the trajectory of a reaching movement.

Further results, less crucial for our main hypotheses, confirm and extend previous findings in the literature. They indeed complement studies showing that responses to artifacts are slower than to natural objects (Borghi et al., 2007; Vainio et al., 2008). This might appear counterintuitive, since artifacts are designed to be used, but as suggested in the literature it is probably due to the fact that they do not only activate manipulation but functional information as well, and also to the fact that these two kinds of information might collide (e.g., Jax and Buxbaum, 2010; see also chapter 5).

The effects obtained raise the issue of whether and to what extent our effects depend on online computation (plausibly supported by the dorsal stream), or whether they depend on information stored in memory. To better investigate this issue we performed a second experiment in which we presented the names of the objects instead of the images; this allowed us to determine whether the effects were due to information stored in memory rather than on online computations.

6.3. Experiment 2

6.3.1. Method

6.3.1.2. Participants

Twenty-four under graduated students from the University of Bologna (12 males; mean age = 22.37 (3.19); all Italian monolingual and right-handed by self-report) participated for course credits. All participants had normal or corrected-to-normal vision and were naive as to the purposes of the experiment.

6.3.1.3. Materials, design and procedure

Participants performed the same semantic categorization task of Experiment 1. In this case, they were presented with the names of the 16 objects of Experiment 1. Linguistic stimuli were presented in a 200x250 pixels box in ARIAL font upper case, black print on white background. Experimental procedure and task instructions were exactly the same as in Experiment 1.

6.3.2. Results

Accuracy, Initiation Time and Trajectory time

2.44% of trials were removed as errors, a very low rate which again confirmed that the task was easy to perform. Total trajectories times exceeding 2 standard deviations from each participant's mean were excluded from the analysis, leading to the removal of additional 8.99% of the data. The total trimming was of 11.43% of trials.

As in Experiment 1, the remaining data were entered into a 2 x 2 x 2 x 2 within subjects ANOVA, with the factors Response Device (Big mouse vs. Small mouse), Target Type (Artifact vs. Natural), Target Dimension (Big vs. Small) and Distractor Dimension (Big vs. Small). Where possible, interaction effects were evaluated with Newman-Keuls post-hoc test ($p < .05$).

The ANOVA on *Initiation times* showed the main effect of Response Device; two interactions were significant as well, while a third one almost reached significance.

As in Experiment 1, the initiation times were longer with the Big than with the Small mouse (416 ms and 284 ms, respectively; $F(1, 23) = 21.29$, $MSe = 78356.5$, $p < .001$).

The three-way interaction between the factors Response Device, Target Type and Distractor Dimension ($F(3, 92) = 4.34$, $MSe = 3787.41$, $p < .05$) was significant too (Mouse big / Artifact: Distractor big M = 416 ms - Distractor small M = 432 ms, Natural: Distractor big M = 413 ms - Distractor small M = 404 ms; Mouse small / Artifact: Distractor big M = 291 ms, Distractor small M = 281 ms, Natural: Distractor big M = 274 ms - Distractor small M = 292 ms), even if no one of the differences between critical conditions were reliable in the post-hoc analysis. The three-way interaction between the factors Response Device, Target Dimension and Distractor Dimension ($F(3, 92) = 4.34$, $MSe = 3787.41$, $p < .05$) almost reached significance (Mouse big / Target big: Distractor big M = 421 ms - Distractor small M = 411 ms, Target small: Distractor big M = 408 ms - Distractor small M = 424 ms; Mouse small / Target big: Distractor big M = 283 ms, Distractor small M = 295 ms, Target small: Distractor big M = 281 ms - Distractor small M = 277 ms).

A four-way interaction, the Response Device x Target Type x Target Dimension x Distractor Dimension ($F(4, 184) = 6.93$, $MSe = 3619.86$, $p < .05$) was significant as well (Mouse big / Artifact / Target big: Distractor big $M = 433$ ms - Distractor small $M = 418$ ms, Target small: Distractor big $M = 399$ ms - Distractor small $M = 445$ ms, Natural / Target big: Distractor big $M = 409$ ms - Distractor small $M = 405$ ms, Target small: Distractor big $M = 416$ ms - Distractor small $M = 403$ ms; Mouse small / Artifact / Target big: Distractor big $M = 281$ ms - Distractor small $M = 294$ ms, Target small: Distractor big $M = 301$ ms - Distractor small $M = 267$ ms, Natural / Target big: Distractor big $M = 286$ ms - Distractor small $M = 297$ ms, Target small: Distractor big $M = 262$ ms - Distractor small $M = 287$ ms).

No other main effects or interactions were significant.

The analyses on total Trajectory times revealed as significant only the main effect of Target Type (*F*(1, 23) = 11.58, *MSe* = 7252.6, *p* <.01), with shorter latencies for Natural (*M* = 1403 ms) than Artifact (*M* = 1433 ms) items. No other main effects or interactions were significant.

Trajectory spatial analysis

The ANOVA on AUC demonstrated the main effects of the factors Response Device, Target Type and Distractor Dimension. Two interactions were reliable as well, whereas another almost reached significance.

The Response Device main effect showed that, as in Experiment 1, the Big mouse mean AUC (*M* = 0.33) was smaller than the Small mouse mean AUC (*M* = 0.52 ms), *F*(1, 23) = 12.03, *MSe* = 0.29052, *p* <.01.The factor Target Type, *F*(1, 23) = 6.72, *MSe* = 0.05682, *p* <.05, confirmed also in the AUC results a advantage for Natural items (*M* = 0.40) over Artifacts (*M* = 0.45). The Distractor Dimension factor, *F*(1, 23) = 5.22, *MSe* = 0.05406, *p* <.05, showed a smaller AUC for the Small (*M* = 0.40) than for the Big distractor (*M* = 0.45).

The interaction Target Dimension x Distractor Dimension was significant, *F*(2, 46) = 4.57, *MSe* = 0.02833, *p* <.05, due to the fact that when the target-word referred to a Big object a greater deviation was observed for Distractor big (*M* = 0.48) than Small (*M* = 0.39) (Newman-Keuls *p* < .01), while no difference was observed for Small targets (Distractor big *M* = 0.42, Distractor Small *M* = 0.41) (Newman-Keuls *ns*) (see Fig. 4, Graph a).

The three-way interaction between Response Device, Target Type x Target Dimension was significant as well, *F*(3, 92) = 4.87, *MSe* = 0.06818, *p* <.05 (Mouse big / Artifact: Target big *M* = 0.37 - Target small *M* = 0.34, Natural: Target big *M* = 0.29 - Target small *M* = 0.33; Mouse small / Artifact: Target big *M* = 0.53-Target small *M* = 0.59, Natural: Target big *M* =

0.54 - Target small *M* = 0.43). It showed that for word presentation a subtle compatibility effect was modulated by the target stimulus category; the difference concerned especially the small target when using the small mouse, with the natural items showing significantly lower AUC mean values than the artifacts (Newman-Keuls *p* < .05) (see Fig. 4, Graph b).

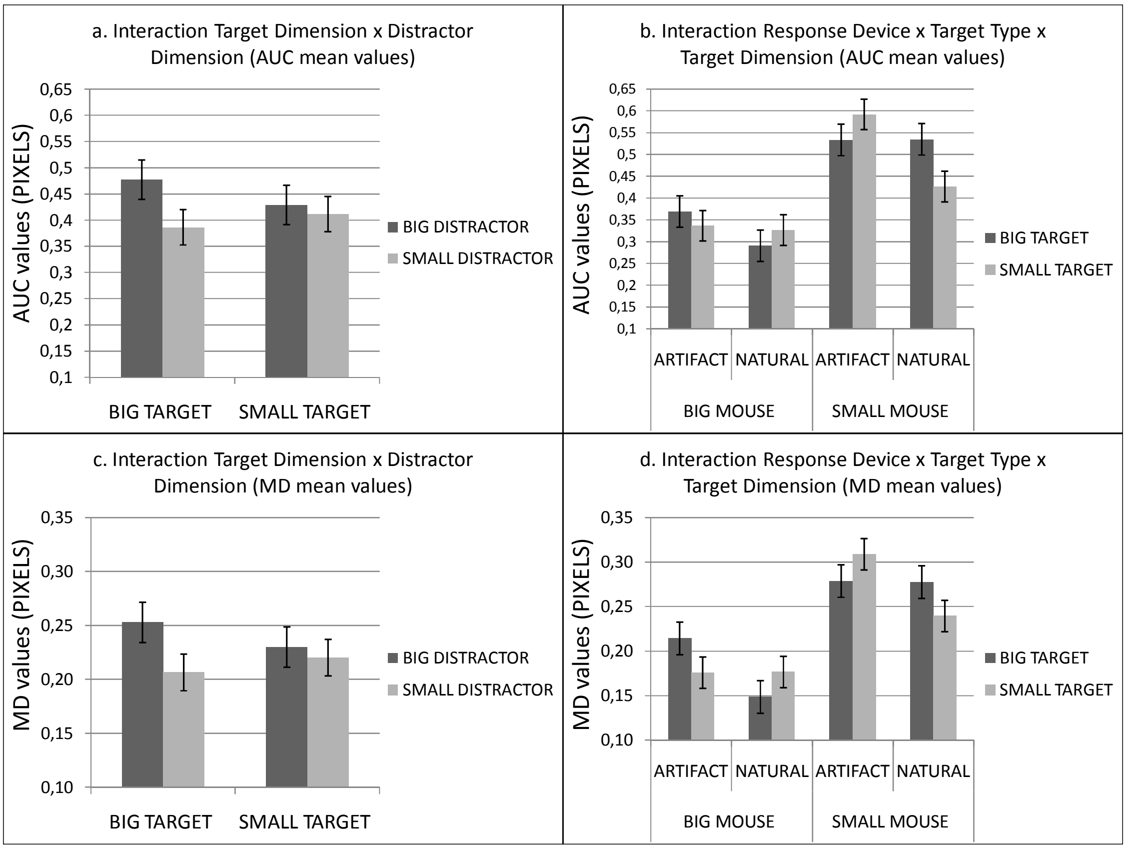
The ANOVA on MD demonstrated the main effects of the factors Response Device, Target Type and Distractor Dimension. Two interactions were reliable as well.

The Response Device main effect was due to the Big mouse mean MD (*M* = 0.19) being smaller than the Small mouse mean MD (*M* = 0.28 ms), *F*(1, 23) = 11.65, *MSe* = 0.07837, *p* <.01.The factor Target Type, *F*(1, 23) = 8.99, *MSe* = 0.01232, *p* <.01, confirmed again the significant advantage for Natural items (*M* = 0.24) over Artifacts (*M* = 0.21). The Distractor Dimension main effect, *F*(1, 23) = 6.93, *MSe* = 0.01095, *p* <.05, showed a smaller AUC for the Small (*M* = 0.21) than for the Big distractor (*M* = 0.24).

The interaction of the factors Target Dimension and Distractor Dimension was significant, *F*(2, 46) = 5.10, *MSe* = 0.00633, *p* <.05, because when the target-word referred to a Big object a greater MD was observed for Distractor big (*M* = 0.25) than Small (*M* = 0.21) (Newman-Keuls *p* < .05), while no difference was observed for Small targets (Distractor big *M* = 0.23, Distractor Small *M* = 0.22) (Newman-Keuls *ns*) (see Fig. 4, Graph c).

Finally, the three-way interaction between Response Device, Target Type and Target Dimension was significant too, *F*(3, 92) = 5.45, *MSe* = 0.02005, *p* <.05 (Mouse big / Artifact: Target big *M* = 0.21 - Target small *M* = 0.18, Natural: Target big *M* = 0.15 - Target small *M* = 0.18; Mouse small / Artifact: Target big *M* = 0.28-Target small *M* = 0.31, Natural: Target big *M* = 0.28 - Target small *M* = 0.24) (see Fig. 4, Graph d).

No other main effects or interactions were significant.



a. Interaction Target Dimension x Distractor Dimension (AUC mean values)

b. Interaction Response Device x Target Type x Target Dimension (AUC mean values)

c. Interaction Target Dimension x Distractor Dimension (MD mean values)

d. Interaction Response Device x Target Type x Target Dimension (MD mean values)

Figure 4. Experiment 2, all the results

So, in Experiment 2 the interaction of the factors Response Device and Target Dimension was *not significant* in the two considered measures, implying a difference for visual and linguistic stimuli, which will be discussed further.

6.3.3. Discussion

In Experiment 2, results with words are quite different from those obtained with objects. The continuous recording of participants' mouse movements did not show the predicted compatibility between the hand posture and the implied dimension of the target stimulus. Thus we failed to replicate with the present paradigm and the present measures the results obtained by Tucker & Ellis (2004), who found a compatibility effect in RTs not only with objects but also with words. However, the absence of the interaction between the factors

Response Device and Target Dimension in the AUC analysis may not tell the whole story, as indicated by the interaction of Response Device, Target Type and Target Dimension. The interaction between object size and distractor size in MD and AUC, due to higher response uncertainty when the big target was matched with the big than with the small distractor, reveals that participants are sensitive to the size of object the words refer to. Note that more interesting for our analysis would be a putative interaction between category, mouse and target, in particular the presence of a compatibility effect between the mouse and the target, even if confined to natural objects; but this interaction did not reach significance. Finally, the main effects of the object kind both in MD and AUC, due to the higher uncertainty with artifacts compared to natural objects, strictly matches the results of Experiment 1.

At first sight, the absence of the predicted compatibility effect might seem problematic for an embodied account of language processing, according to which words are grounded in perception, action and emotional systems, see (Barsalou, 2008; Fischer & Zwaan, 2008; Jirak et al., 2010; Meteyard et al., 2012). However, even if we did not find the same results with words and images, we found evidence of the activation of motor information with words as well. The presence of the effect of size suggests that words indeed elicit modal information as part of an embodied re-enactment or simulation of the associated sensorimotor experience (Barsalou, 2008; Pezzulo et al., 2011; Pezzulo, Barsalou et al., 2013; Pezzulo, Candidi et al., 2013). Still, the putative simulation is not so fine-grained as the one formed during object processing and so effect of object size seems to dominate over the compatibility one.

Our results can be read in terms of recent proposals emerging in literature on embodied cognition. Theories of reuse and motor exploitation suggest indeed that language recruits and reuses structures and mechanisms characterizing the motor system (Anderson, 2010; Pezzulo & Castelfranchi, 2009; Gallese, 2008). However, this is not the end of the story, since language also modifies these structures and mechanisms and builds on them

(Borghi, 2012; Gallese, 2008). For example, it has been shown that language recruits only some kinds of affordances, as those linked to stable characteristics of objects, as for example object size and not object orientation (Borghi, 2012; Borghi & Riggio, 2009; Ferri et al., 2012; Myalkykov et al., 2013). Our results interestingly indicate that, while the compatibility between the executed grip and the observed visual object occurs online, motor information on object size is processed offline and influences language comprehension.

6.4. General discussion

We report that static hand postures facilitate compatible responses with objects requiring either a precision or a power grip. Specifically, we demonstrated this investigating the effects of compatible or conflicting information as they were reflected in the trajectories of overt hand movements: participants were instructed to use a mouse to reach for objects or for words referring to objects on the computer screen. To our knowledge the present is the first work that provides evidence of this kind, obtained with kinematics measures.

This evidence clearly favors an embodied account of cognition, according to which observing objects activates the motor system. While object observation leads to the activation of fine-grained motor information aimed at preparing a specific kind of grip, the story is different for words. With words we found indeed evidence of activation of motor information, as the effect of size suggests, but we failed to replicate the compatibility effect previously found by Tucker & Ellis (2004) with a different paradigm. As argued in the discussion of Experiment 2, this can be interpreted in the framework of embodied theories of reuse, according to which language recruits some characteristics of the motor system, modifying and building on them.

The comparison of the results obtained with objects and words suggests that the compatibility effects found with objects occur online, thus are likely due to the activation of the dorsal route rather than of the ventral stream (Milner & Goodale, 1995). Further research should explore whether the effects would be similar with words and with not scaled images, i.e., with images that do not allow computing online the object size. Notice indeed that the images we used in Experiment 1 were scaled, i.e., they maintained some resemblance to the original size, even if larger objects were more reduced in dimension compared to small ones, to fit them within the square.

Less crucial to our main hypothesis but still important for an embodied cognition view is the advantage of artifacts over natural objects. This advantage, which is likely due to the activation (with artifacts) of both manipulation and functional information, is present with both objects and words, thus it is probably not merely due to the dorsal route activation.

Overall our study shows that static hand posture influences the on-line dynamics of a decision even if it is irrelevant to the performance of the task. Unexpectedly for opponents to an embodied cognition view, the way the mouse is held seems to have a number of effects on human thought processes.

7. Simulating concrete and abstract words acquisition

The Words As social Tools theory (WAT, Borghi & Cimatti, 2009, 2010), described in par. 3.5 while discussing the embodied accounts of abstract language grounding, formulates precise predictions about the grounding of both concrete and abstract concepts/words. According to WAT, abstract words meaning is more strongly related to the experience of being exposed to socio-linguistic situations than concrete words meaning, which rely more on sensory-motor experiences. Thus, the difference between abstract and concrete words is directly related by this theory to their different modalities of acquisition (MoA). This study tested in four experiments if different acquisition modalities effectively lead to the emergence of the differences typically found between concrete and abstract words in the literature. In order to mimic the acquisition of concrete and abstract concepts, participants either manipulated novel objects or observed groups of objects interacting in novel ways on a screen (Training 1). In a following recalling task (Test 1) participants decided whether two elements belonged to the same category or not. Then, each item were presented with a category label (Training 2), with labels being or not accompanied by an explanation of the meaning of the category. In a subsequent category-label verification task, participants were presented with the previously seen exemplars along with novel elements and one of the learned label, being required to decide which of the items corresponded to the label (Test 2). The results showed that, across the experiments, it was more difficult to form abstract than concrete categories (Test 1), even if the labels were added (Test 2). A third task (Test 3) differed across the experiments: in Experiment 1 participants performed a feature production task, the results of which showed that the properties produced with the novel categories matched the patterns evoked by real existing concrete and abstract words. In Experiment 2, 3 and 4, the third task (Test 3) consisted of a color verification task with manual and vocal responses (i.e., keyboard/microphone), with the main difference between the experiments being the fact that the color could be or not predictive of the categories. The results showed an advantage for vocal over manual responses for abstract words, especially if they were previously learned with the explanation. Interestingly, this advantage disappeared in the case in which linguistic information contrasted with perceptual one. On the whole, these results support WAT predictions: due to their different MoA, concrete words evoked more manual information, while abstract words more verbal information.

7.1. Introduction

This study on novel categories has its focus on what differs in the acquisition of concrete and abstract words. One standard way of differentiating between concrete and abstract words is to refer to their perceivability. Concrete words refer to entities that can be perceived through the senses. Abstract words refer to entities more detached from physical experience (Barsalou et al., 2003; Paivio et al., 1986; Crystal, 1995). However, the distinction between concrete and abstract words cannot be conceived of as a dichotomy (Wiemer-Hastings et al., 2001). For example, words referring to social roles (e.g., "physician") might be more abstract than words referring to single objects (e.g., "bottle"), but less abstract than purely definitional words (e.g., "odd number") (Keil, 1989). In addition, words referring to emotions probably require special classification (Altarriba et al., 1999). Further, basic and subordinate words, such as "cat" and "siamese cat", referring to single entities, can be seen as more concrete than superordinate words, such as "animal", that refer to sets of entities that differ in shape and other perceptual characteristics (e.g., Borghi et al., 2005). To summarize, the distinction between concrete and abstract words is not clear-cut, and should be rather intended as a continuum. However, we believe that this distinction captures some aspects of word meaning, and that it is important to understand how the process of abstraction occurs, from single instances to categories at different levels of abstraction. In particular, explaining the ways in which abstract words are represented constitutes a major challenge for embodied and grounded views of cognition, as well as for embodied computational models and robotics. The problem abstract words pose for embodied and grounded theories is clearly synthesized by Barsalou (2008, p. 634) as follows: “Abstract concepts pose a classic challenge for grounded cognition. How can theories that focus on modal simulations explain concepts that do not appear modal?”. We will first clarify why explaining abstract concepts is a crucial challenge for embodied cognition, and later clarify its importance for research in robotics.

According to the standard propositional view (e.g., Fodor, 1998), the representation of both concrete and abstract concepts is abstract, symbolic and amodal. In contrast, according to standard embodied accounts (e.g., Barsalou, 1999) both concrete and abstract concepts are grounded in the sensory-motor system, and therefore are modal (see chapter 1). Notice that both standard propositional and embodied accounts evoke a single kind of representation, either amodal or modal, for both concrete and abstract concepts (see par. 3.5).

In contrast, recent views propose that multiple representational systems are activated during conceptual processing (for a non embodied version of this view see Dove, 2009). According to these views both sensory-motor and linguistic information play a role in conceptual representation. This idea is not entirely novel. The seminal dual coding theory by Paivio et al. (1986) applies two different kinds of representations, a linguistic and a sensory-motor code, to explain how concrete and abstract words are represented and recalled. Concrete words are recalled more easily because they activate both sensory-motor and linguistic information; differently abstract words are not “grounded”, they only evoke linguistic information. Recent support to Paivio’s theory comes from studies on brain imaging showing that abstract word processing is strongly lateralized towards the left hemisphere, while activation during processing concrete words is bilateral (for a review, see Sabsevitz et al., 2005). However, this might be due to the fact that the majority of the studies employ single words and tasks requiring a superficial level of processing. Recent studies requiring deeper processing, such as sentence sensibility evaluation tasks, do not provide evidence in favor of a pronounced laterality (e.g., Desai et al., 2010). The major difference between Paivio's view and embodied accounts is based on the concept of multiple representation; to elaborate, Paivio argues that abstract words are not "grounded" in perception and action systems, whereas according to the embodied perspective both concrete and abstract words activate both linguistic and perception-action information, even if these two kinds of information are differently distributed.

The Language And Situated Simulation (LASS) theory is probably the most well-known of the multiple representation theories (Barsalou et al., 2008). In this view both the linguistic and the sensory-motor system are activated during word processing. The understanding of word meanings always implies activation of the sensory-motor system (simulation), but for tasks which do not require deep processing the linguistic system might suffice. While presenting the LASS theory, Barsalou et al. (2008) suggest that for abstract concepts, linguistic information might be more relevant than for concrete concepts, but they do not advance clear predictions pertaining the differences in processing between concrete and abstract concepts, independently from the task. Thus, they argue that "different mixtures of the language and simulation systems support the processing of abstract concepts under different task conditions" (Barsalou et al., 2008, pp. 267).

More precise predictions concerning the difference between concrete and abstract words are advanced by the Words As social Tools (WAT) proposal (Borghi & Cimatti, 2009; 2010), which assumes the existence of multiple representations and also idea, initially proposed by Wittgenstein (1953) and Austin (1962), that words are *tools* (see also Clark, 1998). Similarly to real tools, words can be considered as instruments to act in the social world, thus as social tools. The difference between concrete and abstract words is explained by WAT referring to the fact that, due to a different acquisition process, the role played by actions performed through words – by linguistic information – is more relevant for abstract than for concrete words. The present work aims to directly test the WAT proposal using novel categories and novel linguistic labels. According to WAT, perception and action are crucial in the acquisition of concrete words. Instead to acquire the meaning of an abstract word people also rely on verbal explanations (for example, explaining the meaning of "democracy" to a children requires many more other words than for explaining the meaning of "bread"). In this respect, the role played by words as social tools is more important for abstract than for concrete words. Evidence relevant to this issue was obtained by Wauters et al. (2003), who studied different Modalities of Acquisition (MoA) of words. They did not however, speak directly about concrete and abstract words. According to the authors, the meaning of a word like "ball" is acquired through perception, because every time the child hears the word, he/she sees a real ball, or a picture of it. The meaning of a word like "grammar", instead, has to be explained linguistically. Finally, the meaning of a word like "tundra" can be acquired in both ways, depending on the environment where it is learned. WAT predicts that this difference in the acquisition process can explain why, for concrete and abstract words both perception-action and linguistic information are activated. Linguistic and social information however, plays a more important role for abstract than for concrete words (e.g., Crutch & Warrington, 2005; Sabsevitz et al., 2005).

From a different perspective, an embodied and grounded account of the difference between concrete and abstract words is crucial in the process of developing intelligent machines capable of autonomously creating categories and using language. In computational cognitive science, robotics offers new opportunities for the design of artificial agents in which language is grounded on their ability to manipulate and experience the external world by means of physical interactions. The *symbol grounding problem* (Harnad, 1990) highlights the fact that, in traditional computational models, symbols are self-referential entities that require the interpretation of an external experimenter to identify the referential meaning of the lexical items. This issue has been widely discussed in the realm of cognitive science, and robotics offers a completely different way to solve the grounding problem. Indeed, in the last 20 years many different models were created with the explicit aim of grounding symbols and language in perception (e.g., Steels, 2003) and, more recently, in action (Marocco et al., 2010; Sugita & Tani, 2005). Although the embodied approach to language in robotics is gaining increased interest, both in terms of cognitive modeling and applications, the current trend is strongly focused on systems capable of autonomously acquiring concrete concepts and words, that can be grounded on perception and action processes of the robot. Existing models do not focus on the acquisition of abstract words, except for highlighting that such abstract concepts and words permeate the entire domain of human language experience and cannot be neglected. Nevertheless, an extension of the actual grounding approach in robotics to abstract words is not automatic. In this regard, we believe that the WAT proposal offers an interesting way to incorporate abstract words in future cognitive robotic models without compromising the grounding and the embodied approach, which should be the milestone of the future robotics. On the other hand, a robotic model could be useful to complement traditional psychological experiments, and provide further evidence on the feasibility of a novel theory, such as the WAT proposal presented.

In this research we used novel categories to mimic the different ways in which concrete and abstract word meanings are acquired and then represented. Reported experiments are designed in a way that allows for replication with a computational model. Similar stimuli and training processes can be used to create a cognitive based controller for a humanoid robot (Tikhanoff et al., 2008) that will be able to perform an identical categorization task. We defined concrete concepts as having a concrete, manipulable object as a referent. Abstract concepts, on the other hand, do not have a single, manipulable object as referent; instead they refer to rather complex relations between entities. We acknowledge that the distinction we made for operational simplicity is not exhaustive and that it covers only a subset of items. For example, it leaves out word meanings referring to perceivable but not manipulable objects or entities, such as "cloud", "mountain", and "moon". Even if the referents of these words cannot be manipulated, we would consider them as concrete, as their referents are clearly perceivable, can be scanned (acted upon) with the eyes, and are easy to imagine. We decided to address the distinction between concrete and abstract words starting from the extremes of the continuum: for this reason we decided to focus on concrete, manipulable objects. As for abstract word meanings, here we did not refer to purely definitional abstract word meanings, simply based on verbal explanations (as it might be the case for a word like "philosophy") but to word meanings that evoke complex relationships between entities. Due to their complexity, we predicted that applying a linguistic label and explaining their meaning would be crucial in order to form categories. Consider that the referents of our abstract categories were interacting moving objects – thus they were perceivable, similarly to the referents of concrete categories. As a matter of fact, in our view the formation of abstract categories always starts with some form of perception, be it visual, acoustic, tactile or otherwise.

Due to the difficulties involved in reproducing the acquisition of different kinds – concrete vs. abstract – of novel concepts/words in an artificial setting (i.e., laboratory), we operationalized the acquisition process considering two phases – the experience and the word acquisition – as follows:

a - Novel concepts acquisition: Training 1 (Experience) was designed to mimicking the acquisition of concrete and abstract concepts. The idea underlying these two different acquisition processes is that, where typically concrete concepts refer to category members which are perceptually similar or elicit similar actions, abstract concepts refer to entities that show complex interactions, or do not share an evident perceptual similarity (i.e., common features are not perceptually salient). We showed participants 3D figures of novel objects vs. 3D figures of objects interacting in novel ways. Then participants were tested (TEST 1: Categorical Recognition).

b - Novel labels acquisition: During Training 2 (Words Acquisition) participants were taught the category name; in some conditions a verbal explanation of the category meaning was added. Then participants were tested (TEST 2: Words-Objects Match). We predicted that in both tests participants would produce less errors with concrete than with abstract categories, as the first can be formed more easily on perceptual and motor basis. This difference should be reduced when a category label and a linguistic explanation of what the category members had in common was given. The manipulation of TEST 3 in the different experiments allowed us to check for the effectiveness of our operationalization of acquisition processes (Experiment 1), as well as to test if the verbal labeling, possibly strengthened by a verbal explanation, reinforces learning of both concrete and abstract categories in different ways (Experiment 2, 3 and 4).

c - Real words evidence match: TEST 3 of Experiment 1 consisted of a feature production task. We predicted that the pattern of produced properties would match that typically obtained in feature generation tasks with concrete and abstract words.

d - Linguistic vs. manual Information: In Experiments 2, 3, and 4, TEST 3 consisted of a property verification task. We chose to ask participants to respond to the items color because color was not relevant to the motor response. In one condition participants were required to provide a manual response (i.e. to press a key on the keyboard), and in another a verbal response (i.e. to respond “yes” with the microphone; see Scorolli & Borghi, 2007). We predicted facilitation for manual responses with concrete words and for mouth responses with abstract words. This would demonstrate that language is part of the representation of abstract words meanings.

7.2. Experiment 1

This experiment was designed to mimic the acquisition of concrete and abstract categories and to verify whether the novel categories we used reproduced the acquisition process that occurs with real world categories. As anticipated, in Experiment 1 TEST 3 consisted of a property production task. Before starting the experiment, participants were randomly assigned to two groups. One group was first shown the category and then tested on concrete items; later participants were shown and then tested on abstract items; the other group first learned and then was tested on the two kinds of items in reverse order. Across the experiment the order of presentation of the two blocks (concrete block; abstract block) was counterbalanced. The same methodological choice was applied to all the other three experiments.

7.2.1. Method

7.2.1.1. Participants

16 students of the University of Bologna took part in the study (3 men; mean age = 20.31 years; s.d. = 1.62). All were native Italian speakers, both right- and left-handed (2 left-handed) and all had normal or corrected-to-normal vision. The study was approved by the local ethics committee.

7.2.1.2. Materials

7.2.1.3. *3D figures of novel objects and related new labels*

We invented four novel words (*calona*, *fusapo*, *norolo*, *tocesa*) all having the same number of syllables and letters. We avoided using new words with ambiguous accents. Two of the four words ended with the vocal "a", which in Italian characterizes the female gender; the remaining two words ended with the vocal "o", which in Italian characterizes the male gender. The new words corresponded to four new categories of objects, composed of twelve exemplars each (4 x 12). The criteria we followed to construct the "original" three new objects were the following:

1. CALONA was a 3D concave figure ("C" shaped). The colors we used were sky-blue and light-grey;
2. FUSAPO was a 3D figures with five protuberances ("*"shaped). The colors we used were blue and yellow (Fig. 1);
3. NOROLO was a 3D figure with small convex nooks ("N" shaped). The colors we used were red and grey;
4. TOCESA was a 3D figure shaped as wavy slash, without internal convexities or concavities ("I" shaped). The colors we used were violet and beige.

The other nine exemplars for each category were both built by inverting the surface and depth colors (3 x 2), and by rotating the original figures by 180 degrees (6 x 2). Finally, we built 40 3D figures that were used as fillers: they did not belong to a category and were not assigned a name.

7.2.1.4. *3D figures of novel relations and related new labels*

We invented four new words (*cofiro*, *latofo*, *panifa*, *rodela*) by following the same criteria as described for the linguistic labels used for the 3D figures of novel objects. These new words referred to new categories of *relations* between two 3D figures; each of these categories was composed by twelve exemplars (4 x 12). We followed the following criteria to construct the "original" three new relations (that is, novel groups of 3D interacting objects):

a. COFIRO: two 3D moving figures. After the contact just one 3D figure remained, and it moved in a straight line or in a curved line;

b. LATOFO: one 3D static figure and two 3D moving figures. After the contact two 3D figures appeared at the opposite diagonal sides of the computer screen (e.g., one at the top right of the screen and the other at the bottom left of the screen), and they moved converging towards the central point of the screen;

c. PANIFA: two 3D moving figures. After the contact one of them moved in a straight line; the other one executed a turning movement with a different velocity (Fig. 2);

d. RODELA: one 3D static figure and two 3D moving figures. After the contact the two 3D figures moved in a same (straight) line and with the same velocity, but in an opposite direction, as if the figures were pushed away from each other.

All the 3D figures were sky-blue cylinders; they were arranged horizontally, one came from one part of the screen and the other from the other side. For LATOFO and RODELA we added a 3D static figure to the two interacting ones. This aimed to reproduce real life abstract word acquisition: some abstract words can evoke both relations between entities and static visual images (e.g., "freedom" can evoke a bird flying in the sky as well as an image of the Statue of Liberty). In other words, it can happen that objects which would be first categorized as exemplars of a concrete category (e.g. a statue) can be re-categorized and evoked by abstract words.

The other nine exemplars for each category were built by using parallelepipeds (3 x 2) instead of cylinders; the movement of the 3D figures followed a vertical instead of a horizontal direction (6 x 2). Finally, we built 40 3D figures to use as fillers, and we constructed 40 relations between 3D figures to use as fillers. They did not belong to a category and were not assigned a name. The duration of each relation was the same for both the categories' exemplars and the fillers (4 seconds).

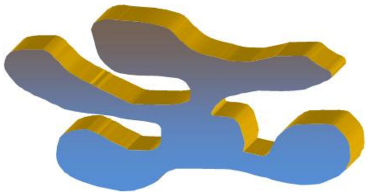
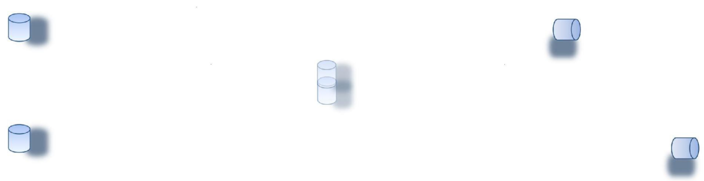


Figure 1. Example of entity pertaining to a concrete category (FUSAPO).



A. Initial phase B. Intermediate phase C. Final phase

Figure 2. Example of relation pertaining to an abstract category (PANIFA).

7.2.1.5. Procedure

Across all experiments, participants were trained and tested individually in a quiet laboratory room. They sat on a comfortable chair in front of a computer screen. All participants were submitted to 2 training phases (Experience; Word Acquisition) and to 3 different tests (Categorical Recognition; Word-Object Match; Production).

7.2.1.6. Training 1: Experience

Training 1 aimed to reproduce the different processes underlying the acquisition of concrete and abstract concepts. Whereas typically concrete words refer to category members which are perceptually similar or elicit similar actions, abstract words refer to entities that show complex interactions or do not share an evident perceptual similarity (i.e. common features are not perceptually salient). For example, the word “truth” binds experiences and situations that might be rather complex and different. During this training session participants were sitting in front of the computer screen. They were exposed to 20 trials. In each trial either three 3D figures (in the *concrete concept* acquisition condition) or three relations between 3D figures (in the *abstract concept* acquisition condition), were shown. Both the 3D figures and the relations were novel, i.e. participants had never experienced them before. In order to mimic the acquisition of concrete concepts (e.g., BOTTLE), participants were presented with 3D figures of novel objects as previously described. They were instructed to verify whether the objects could be inserted inside a doughnut shaped 3D figure. The experimenter invited them to manipulate the objects with the mouse for 12 seconds each. In order to simulate the acquisition of abstract concepts (e.g., TRUTH), participants were instructed to observe the groups of dynamic objects until the end of their interaction (12 seconds). The 3D figures interacted in ways that revealed the existence of a common structure. For example, two objects moved toward each other, then only one of them remained on the screen, moving in a straight line (COFIRO).

7.2.1.7. TEST 1: Categorical Recognition

Training 1 was followed by a categorical recognition task (TEST 1). Participants were instructed to look at a fixation cross that remained on the screen for 500 ms. Then they were shown two exemplars of the same or different categories, and were asked to judge whether the stimuli belonged to the same category or not by pressing two different keys (left, right). The key-response mapping was counterbalanced. They were shown 24 randomly ordered trials, with different combinations of the exemplars or of the exemplars and fillers, that is:

1) two exemplars of the same category;

2) two exemplars belonging to two different categories;

3) one exemplar of a category and one filler, that did not belong to any learned category.

Concrete concepts' exemplars remained on the screen for 2 seconds, while abstract concepts' exemplars were displayed for 10 sec. The 24 experimental trials were preceded by 2 training trials.

The Categorical Recognition task aimed to verify whether the training phase allowed participants to form a category on a purely sensory-motor basis, and to contrast it with a different category. We collected and analyzed errors, as this is the more reliable and informative measure for this particular task. We predicted that participants would produce less errors with concrete than with abstract categories, as the first can be formed more easily without the aid of language.

7.2.1.8. Training 2: Words Acquisition

After TEST 1, participants were trained to associate a linguistic label to each learned exemplar. Five exemplars from each category were randomly selected and they were presented once to participants together with the appropriate linguistic label. In order to mimic the acquisition of concrete words participants were shown 20 3D figures together with the related linguistic labels ("calona", "fusapo", "norolo", "tocesa"), presented in random order. Each trial lasted 2 seconds. Symmetrically, in order to simulate the acquisition of abstract words, participants observed the 20 relations together with the related linguistic labels ("cofiro", "latofo", "panifa", "rodela"), presented in random order. Each trial lasted 4 seconds. Participants were instructed to learn the linguistic labels associated with the 3D figures and with the relations.

7.2.1.9. TEST 2: Words-Objects Match

After the Training 2 participants had to perform a Words-Objects Match task. They were presented with 24 trials. One of the learned names and 2 figures/relations were displayed on the computer screen: the target object, corresponding to the label, and another nearby, which in half of the trials was novel and in the remaining 12 trials was an exemplar already associated with a different label. One of the two figures/relations was located on the left of the screen, the other on the right; the figure location was counterbalanced. Participants were required to decide by pressing a different key (left, right) on the keyboard which of the two was named with the shown label. This second test aimed to verify whether participants had associated a label with a category, and whether they were able to generalize it to a different category. We predicted that participants would produce fewer errors with concrete than with abstract categories, as the first rely more than the second on perception and action. However, the difference between concrete and abstract categories should be reduced compared to TEST 1, given that participants could now rely on linguistic labels as well.

7.2.1.10. TEST 3: Production task

After TEST 2, TEST 3 consisted of a feature production task with novel category names. The experimenter told participants each category name (in 4 random orders) asking them to produce the first properties that came to their mind. They were prompted to produce properties until they stopped for about 15 seconds. Properties produced were transcribed; both their frequency and production order was recorded. We predicted that the pattern of produced properties would match that typically obtained in production tasks with concrete and abstract words. Behavioral studies with production tasks, such as word association and property generation tasks, have shown that, whereas concrete words activate mainly perceptual and thematic relations, abstract words typically elicit more taxonomic relations (Borghi &

Caramelli, 2001); in addition, they elicit more situations and introspective relations compared with concrete words (Barsalou & Wiemer-Hastings, 2005).

7.2.2. Results

7.2.2.1. TEST 1: Categorical Recognition

We performed a one-way ANOVA on errors produced in the categorical recognition task, in which the factor Concept (Concrete vs. Abstract) was manipulated within participants. As predicted, Abstract Concepts (M = 5.21%) elicited more errors than Concrete Ones (M = 2.34%), *F* (1, 15) = 12.70, *MSe* = 5.17, *p* <.01 (see Tab. 1).

7.2.2.2. TEST 2: Words-Objects Match

An ANOVA was performed on the errors produced. Consider that on the screen two objects were presented, the target one and another object. Therefore in the ANOVA two factors were entered, both manipulated within participants: the factor Word (Concrete vs. Abstract) and the factor Other Exemplar (Novel vs. Learned). Both factors reached significance; Abstract Words (M = 5.01%) elicited more errors than Concrete Ones (M = 1.37%), *F* (1, 15) = 11.96, *MSe* = 17.79, *p* <.01, and more errors were produced when the target exemplar was presented with a Learned (M = 4.17%) than with a Novel Other Exemplar (M = 2.21%), *F* (1, 15) = 15.70, *MSe* = 3.89, *p* <.01 (see Tab. 1).

7.2.2.3. TEST 3: Production task

Different analyses were performed on the production task. The number of produced properties did not differ significantly between Concrete (M = 4.18) and Abstract Words (M = 3.73); *p* = .29. The properties produced with each word were put together, organized in 2 different random orders, and an independent group of 12 participants were asked to rate the produced properties on a 7 point scale. They were asked to select 1 if they believed that the property was typical of words having "concrete" referents, such as bottles, screwdriver, building, cellular, and cat, and 7 if they thought the property was typical of words having "abstract" referents, such as happiness, philosophy, risk, fantasy, democracy. The raters did not know which situation the properties had been produced in. We performed an ANOVA on the ratings of the properties produced with concrete and abstract words. As predicted, we found that abstract words elicited significantly higher scores than concrete words (M =3.93; M = 3.13), F (1, 11) = 27.51, MSe = 0.14, p <. 001. In addition, the scaled ratings were applied to the individual protocols in order to verify whether the properties produced and the production order of the properties for each word reflected the properties typically produced for concrete or for abstract words (the same method was used by Borghi, 2004; Wu & Barsalou, 2009). The average rating of each property was multiplied by the frequency of the produced property for each of the participants. A one-way ANOVA was performed on the obtained means, with participants as the random factor. The only factor manipulated was significant, F (1, 11) = 27.51, MSe = 0.14, p <. 001, as the means obtained with Abstract Words (M = 4.14) were higher than those produced with Concrete Words (M = 3.04), indicating that the novel Abstract Words we created elicited properties typical of real-life abstract words (e.g., "singularity"; "variation"; "linear motion"); this was symmetrically true for the novel Concrete Words which elicited a higher number of properties such as "hole in the middle", "stick-shaped", "crab-shaped". In addition, the average rating on each property was multiplied by the position of the property produced for each participant according to the formula $(n+1-p) / (n-1)*r$, where n is the total number of properties produced by each participant for each word, p the position in which each property was produced and r the average rating on that particular property (for a similar procedure, see Wu & Barsalou, 2009). This normalized p is the position in which each property was produced, in relation to n, the total number of properties produced by each participant. One ANOVA was performed on the obtained means, with participants as random factor; the factor manipulated was the kind of Word (Abstract vs. Concrete Words). The ANOVA again revealed lower means for Concrete (M = 3.11) than for Abstract Words (M = 4.48), *F* (1, 15) = 55.38, *MSe* = 0.27, *p* < .001. This indicates that with our novel Concrete Words properties typically elicited by real concrete words were elicited earlier, and the same was symmetrically true for our novel Abstract Words (see Tab. 1).

EXPERIMENT 1

Categorical Recognition

Concept	Concrete	Abstract
	2.34 (2.62)	**5.21 (2.95)**

Words-Objects Match

Word	Concrete	Abstract
	1.37 (1.64)	**5.01 (4.90)**
Other Exemplar	Novel Exemplar	Learnt Exemplar
	2.21 (3.66)	**4.17 (4.27)**

Production

Scaled ratings (1 concrete referent → 7 abstract r.) applied to the individual protocols		
Word	Concrete	Abstract
	3.04 (0.29)	**4.14 (0.29)**
Normalized position in which each property was produced (n+1-p) / (n-1)*r		
Word	Concrete	Abstract

	3.11 (0.25)	**4.48 (0.46)**

Table 1. Experiment 1 - Errors and standard deviations (in parenthesis) in percentage for TEST 1 and 2; for TEST 3 we reported results on rating scores.

7.2.3. Discussion

Results of Experiment 1 indicate that with our training with novel categories and words we were able to recreate the real-life situation in which concrete and abstract words are learned. Results for TEST 1 (categorical recognition) indicated that it is more difficult to form abstract categories than concrete ones. In addition, results of TEST 3 (property generation task) showed that the properties produced for the concrete and abstract words we created corresponded to those typically obtained with existing concrete and abstract words. Results of TEST 1 and TEST 3 revealed that abstract categories are more difficult to form, and that abstract words are represented differently from concrete ones, as they elicit less perceptual properties, such as properties related to shape, and more abstract and relational properties.

The higher difficulty of abstract words compared to concrete ones was also maintained in TEST 2 (Words-Objects Match), when participants learned to associate a novel word to a category. Results on TEST 2 showed that the use of linguistic labels did not further facilitate the acquisition of abstract in comparison to concrete words. This reveals that the higher complexity of abstract concepts is not reduced thanks to the use of linguistic labels. A possibility is that, in order to reduce the complexity of abstract words, a verbal explanation of the category meaning is needed.

7.3. Experiment 2

Given our results on Words-Objects Match (TEST 2) in Experiment 1, in Experiment 2 we decided to add a verbal explanation to the linguistic label used for abstract categories. This should mirror the way the acquisition process works. Abstract words differ from concrete words insofar that the first refer to a variety of situations, states, events. Due to this complexity, linguistic labels should be more relevant for abstract than for concrete words acquisition, and the first might also require a verbal explanation of their meaning. This is often not the case for concrete words, for which the linguistic label is usually associated with the presence of the object. Experiment 2 aimed to test whether there is a facilitation effect when the meaning of abstract words is explained linguistically, compared to when only the linguistic label is provided.

In addition, the aim of Experiment 2 is to verify whether the different acquisition modality has an impact on the response modality. We designed a property verification task (TEST 3), to be performed in substitution of the production task of Experiment 1 in order to address this aim. We chose to use color as the target property as color was not relevant to the motor response and to the response device that we used.

Specifically, we predicted that, given that for concrete words manual information is more relevant than for abstract ones, participants should be faster to perform a property verification task with concrete words when they had to respond using a keyboard instead of a microphone. Symmetrically, if it is true that linguistic information is more important for the acquisition of abstract word meanings than for concrete ones, faster responses should be noted with regard to abstract words while responding with the microphone than with the keyboard. We expect a stronger effect when abstract words are presented not only with novel verbal labels but with the explanations as well.

7.3.1. Method

7.3.1.1. Participants

Thirty-two students of the University of Bologna took part in the study (8 men; mean age = 20.44 years; s.d. = 1.41). All were native Italian speakers and right handed.

7.3.1.2. Procedure

All participants were submitted to 2 training phases (Experience; Word Acquisition) and to 3 different tests (Categorical Recognition; Word-Object Match; Property verification task). Training 1 and TEST 1 were identical to Experiment 1. However, Training 2 varied, as participants were randomly assigned to two different conditions, the Explanation or No Explanation condition. In the Explanation condition with abstract words half of the participants were told the name of the abstract category and were given an explanation clarifying the similarities of the members of a given category; in the No Explanation condition only the name was associated to the category. Training 2 for concrete categories was the same of Experiment 1.

In TEST 3 participants took part in a color verification task. Questions appeared on the screen, for example, “Is a LATOFO yellow?”. To respond “yes” or “no” they had to press two keys on the keyboard in one block (24 trials), or to pronounce the word “yes” or “no” in the microphone in another block (24 trials). The block order was counterbalanced. Both response times and errors were recorded. Forty-eight responses were recorded; “yes” responses corresponded to questions on five different colors (blue, red, violet, yellow for concrete words and sky-blue for abstract), and “no” responses corresponded to questions about 5 wrong colors (black, brown, green, orange, white).

7.3.2. Results

7.3.2.1. TEST 1: Categorical Recognition

In the one-way ANOVA conducted on error rates the factor Concept (Concrete vs. Abstract), which was manipulated within participants, was highly significant. As predicted and as in Experiment 1, Abstract Concepts (M = 6.18%) elicited more errors than Concrete Ones (M = 1.82%), *F* (1, 31) = 51.32, *MSe* = 5.92, *p* <. 001 (see Tab. 2).

7.3.2.2. TEST 2: Words-Objects Match

We performed two different ANOVAs on the errors produced, one for the No Explanation group (A) and another for the Explanation group (B). In the first ANOVA two factors were manipulated within participants, Word (Abstract vs. Concrete, both without explanation) and Other Exemplar (Novel vs. Learned). In the second ANOVA the same factors were manipulated but, as far as the Word factor is concerned, we contrasted Abstract Words *with* Explanation vs. Concrete Words *without* Explanation. In the first ANOVA, Abstract Words (M = 4.04%) elicited more errors than Concrete Ones (M = 1.17%), *F* (1, 15) = 12.01, *MSe* = 10.93, *p* < .01, and more errors were produced when the target exemplar was associated with a Learned (M = 3.52%) than with a Novel Other Exemplar (M = 1.69%), *F* (1, 15) = 13.35, *MSe* = 3.98, *p* < .01 (see Tab. 2). In addition, the interaction between Word and Other Exemplar was significant, *F* (1, 15) = 5.46, *MSe* = 3.19, *p* < .05. LSD post-hoc showed that all differences were significant (*p* <.05), with the exception of the difference between Concrete Words accompanied with a Learned vs. Novel Exemplar. With Abstract Words, instead, a Target Exemplar presented together with a Learned Exemplar elicited more errors than a Target Exemplar associated with a Novel Exemplar (*p* < .0004). In the second ANOVA both main effects were significant: Abstract Words with Explanation (M = 3.19%) elicited more errors than Concrete Words without explanation (M = 0.98%), *F* (1, 15) = 6.09, *MSe* = 12.87, *p* < .05, and more errors were produced when the target exemplar was associated with a Learned (M = 2.67%) than with a Novel Other Exemplar (M = 1.50%), *F* (1, 15) = 6.09, *MSe* = 12.87, *p* < .05 (see Tab. 2).

7.3.2.3. TEST 3: Property verification task with keyboard vs. microphone

In TEST 3 we collected both RTs and errors, for a number of reasons. First, previous work on the influence of action sentences on keyboard and microphone response devices was performed recording response times (e.g., Scorolli & Borghi, 2007). Second, differently from TEST 1 and TEST 2, no figures were presented, and participants had to read and respond to verbal questions. Thus there were no differences in the presentation timing of concrete categories (static figures) and abstract ones (videos). We will report results based on LSD test (*p* <.05) and discuss the results crucial for our hypotheses. Even though we collected RTs as well, we believe that, given that we study word acquisition, accuracy probably represents the most important measure of participants' performance.

24.77% of the trials were removed as errors. RTs above or below two standard deviations from each participant's means for correct trials were excluded from this analysis. This trimming method leads to the removal of further 3.39% of the data. The mean RTs for correct responses for true trials for each participant were submitted to two ANOVAs, one for the No Explanation group (A) and another for the Explanation group (B). In the first ANOVA two factors were manipulated within participants: Word (Abstract vs. Concrete, both without explanation) and Response Device (Keyboard vs. Microphone). In the second ANOVA we manipulated the same factors but, with the factor Word, we contrasted Abstract Words *with* Explanation vs. Concrete Words *without* Explanation. In both ANOVAs the factor Word was significant. Abstract Words (M = 958 ms; M = 950, respectively) were responded to significantly faster than Concrete ones (M = 1192 ms; M = 1200 ms), *F* (1, 15) = 12.52, *MSe* = 69871.63, *p* <. 005; *F* (1, 15) = 57.04, *MSe* = 17525.17, *p* < .001 (see Tab. 2). Crucially in the second ANOVA we found an interaction between the kind of Words and the kind of Device, *F* (1, 15) = 11.18, *MSe* = 91173.10, *p* <. 01: Concrete Words were responded to significantly faster with the keyboard (M = 1057 ms) than with the microphone (M = 1343 ms) (LSD post hoc, *p* < .05); symmetrically Abstract Words were responded to faster with the microphone (M = 841 ms) than with the keyboard (M = 1059 ms) (LSD post hoc, *p* = .06; see Fig. 3).

The main effect of Word on both the analyses is of marginal interest, as it is probably due to the fact that the task was easier to perform when using Abstract Words, as the figures / entities referred to through abstract words were always light blue colored, whereas objects referred to by concrete words differed in colors. Much more crucial for our hypotheses is the interaction between Word and Response Device found in the second ANOVA (group B): as predicted, with Abstract Words provided by a verbal Explanation RTs were faster with the microphone than with the keyboard; symmetrically with Concrete Words RTs were slower with the microphone than with the keyboard (see Fig. 3). Finally it is interesting to notice the difference between Abstract and Concrete Words, still present *without* the Explanation (group A, 234 ms), was increased by the introduction of the verbal Explanation (group B, 250 ms), particularly in case of mouth responses.

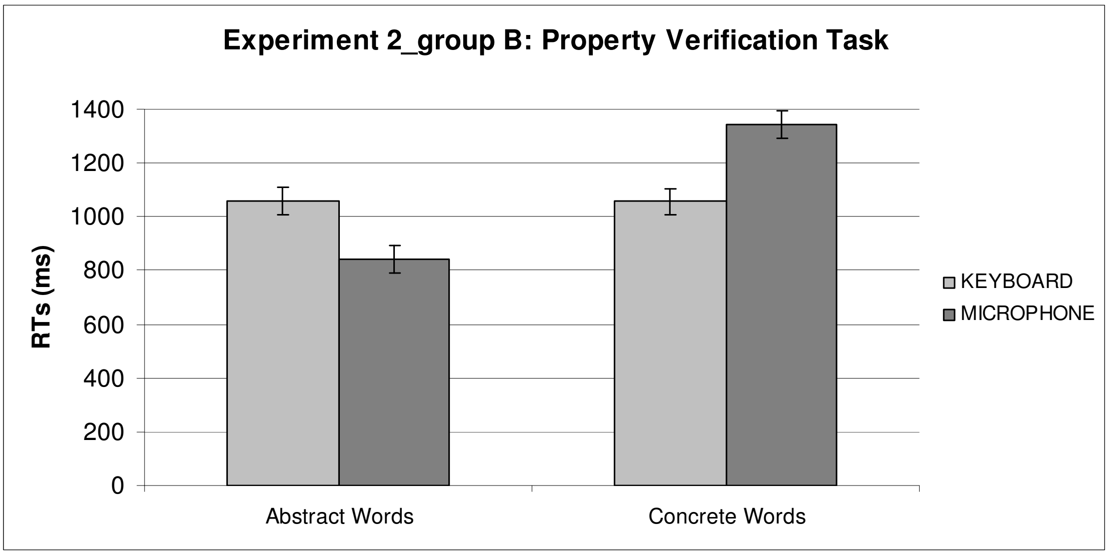


Figure 3. Experiment 2, group B: Interaction between Words (Abstract with Explanation, Concrete) and Response Device (Keyboard, Microphone).

Two further ANOVAs on errors were performed, in which the same factors were manipulated. In both analyses the factor Word reached significance: Concrete Words (group A: M = 15.69%; group B: M = 15.04%) elicited more errors than Abstract Words (group A: M = 10.55%, *F* (1,15) = 4.49, *MSe* = 94.38, *p* < .05; group B: M = 7.75%, *F* (1,15) = 26.04, *MSe* = 32.69, *p* < .001), probably due to the different difficulty level involved in processing the color property. Crucially, the introduction of the explanation strongly reduced errors with Abstract Words (10.55% vs. 7.75%) (see Tab. 2).

EXPERIMENT 2

Categorical Recognition

Concept	Concrete	Abstract
	1.82 (2.17)	**6.18 (3.45)**

Words-Objects Match

Group A Word	Concrete without explan.	Abstract without explan.
	1.17 (1.75)	**4.04 (3.99)**
Other Exemplar	Novel Exemplar	Learned Exemplar
	1.69 (2.62)	**3.52 (3.83)**
Group B Word	Concrete with explan.	Abstract with explan.
	0.98 (1.75)	**3.19 (3.55)**
Other Exemplar	Novel Exemplar	Learned Exemplar
	1.05 (2.44)	**2.67 (3.40)**

Property verification task: keyboard vs. Microphone

Group A Word	Concrete without explan.	Abstract without explan.
	15.69 (8.47)	**10.55 (8.52)**
Group B Word	Concrete without explan.	Abstract with explan.
	15.04 (5.77)	**7.75 (5.63)**

Table 2. Experiment 2 - Errors and standard deviations (in parenthesis) in percentage for each TEST.

7.3.3. Discussion

Results of Experiment 2 confirmed and extended those obtained in Experiment 1. Results on the recognition test confirm the results of Experiment 1, indicating that it is more difficult to form abstract categories than concrete ones. As in Experiment 1, TEST 2 showed that when participants learned to associate a novel word with a category, abstract words caused more difficulty in comparison to concrete words. Interestingly, abstract words without Explanation (group A) produced a significantly higher frequency of errors when the exemplar nearby has already been learned: this suggests that the categorical boundaries are less marked with exemplars referred to by abstract rather than by concrete nouns. By adding an Explanation to the label (group B), the categorical boundaries with exemplars referred to by abstract nouns become marked as the ones referred to by concrete nouns.

More crucial to our hypotheses are the results of TEST 3. As predicted, we found that Abstract Words produced faster responses with the microphone than with the keyboard; by introducing the Explanation (group B) this difference becomes significant. Symmetrically, Concrete Words (group B) were responded to more quickly with the keyboard than with the microphone. This clearly supports the WAT proposal, as it suggests that concrete words evoke more manual information, whereas abstract words elicit more verbal information.

7.4. Experiment 3

A potential problem of Experiment 2 was that TEST 3 (the property verification task) was submitted separately for concrete and abstract words. It is possible that, because abstract words always referred to blue objects, participants did not have to retrieve the perceptual properties of the single categories to respond, whereas this was necessary for concrete words. This could explain why RTs were faster with abstract than with concrete words. Experiment 3 is very similar to Experiment 2, with some modifications. First, given the interesting results obtained with explanations, we decided to use only the explanation condition with abstract words. Second, we balanced color information of objects referred to by both concrete and abstract categories, coloring the abstract figures. We used both concrete and abstract figures of different colors. We introduced this variation in order to solve the potential limitations of Experiment 2, thus to avoid any facilitation with abstract words in responding to the property verification task due to the fact that all abstract words' referents were blue in color. Third, in order to precisely control for the influence of learning the new labels of categorization we decided to perform the category recognition task both before and after learning the category labels. Fourth, and most importantly, we decided to perform the property verification task at the end of the experiment, so that both concrete and abstract words were presented. This modification was introduced in order to be sure that participants referred to the learned category names to respond.

7.4.1. Method

7.4.1.1. Participants

Eighteen students of the University of Bologna took part in the study (9 men; mean age = 23.00 years; s.d. = 2.30). All were native Italian speakers, both right and left-handed (1 left-handed).

7.4.1.2. Procedure

All participants were submitted to 2 training phases (Experience; Word Acquisition) and to 4 different tests (Categorical Recognition without labels; Categorical Recognition with labels; Word-Object Match; Property verification task). The procedure was identical to that of Experiment 2. We only introduced three variations: 1. all abstract words were presented using both the noun and the explanation, thus the No Explanation condition for abstract words was eliminated; 2. we added a further categorical recognition task after Training 2, in order to verify whether using category labels (for both concrete and abstract words) and explanations (for abstract words) would facilitate recognition; 3. the entities to which the abstract words referred to were presented in different colors. Similarly to what we did with concrete ones, we assigned to each abstract category a specific color (light blue, light green, orange, and pink).

7.4.2. Results

7.4.2.1. TEST 1: Categorical Recognition

In an ANOVA conducted on errors two factors were manipulated within participants, the factor Concept (Concrete vs. Abstract), and the factor Language (Before vs. After learning the label designating the category). Only the factor Language was significant, showing that more errors were produced before (M = 1.01%) than after learning the label (M = 0.29%), *F* (1, 17) = 36.26, *MSe* = 0.26, *p* <. 001. Thus, both concrete and abstract category formation appears to benefit from language (see Tab. 3).

7.4.2.2. TEST 2: Words-Objects Match

An ANOVA was performed on errors produced in the word-object match. Both the factors Word and Other Exemplar were significant. Abstract words (M = 4.46%) elicited more errors than concrete ones (M = 2.20%), *F* (1, 17) = 8.42, *MSe* = 10.89, *p* <. 01, and more errors were produced when the exemplar nearby had already been learned (M = 4.57%) than when it had not (M = 2.08%), *F* (1, 17) = 61.85, *MSe* = 1.80, *p* < .001 (see Tab. 3).

7.4.2.3. TEST 3: Property verification task with keyboard vs. microphone

In TEST 3 we collected both RTs and errors for the reasons previously explained (see 8.3.2.3). 12.93% of the trials was removed as errors. The same trimming method of Experiment 2 was used; this lead to the removal of 3.22 % of the data. An ANOVA was performed with two factors Words (Abstract vs. Concrete) and Response Device (Keyboard vs. Microphone) manipulated within participants. As expected, the difference between Abstract and Concrete Words found in Experiment 2 disappeared (means were respectively M = 1150 and 1151 ms): this demonstrates that this difference was due to the fact that processing color was easier in Experiment 2 for abstract words, as the entities they referred to were all of the same color. Crucial to our aims, the interaction between Word and Response Device was significant, *F* (1, 17) = 5.69, *MSe* = 6173.39, *p* <. 05 (see Fig. 4). LSD post-hoc showed that responses with the keyboard were slower than responses with the microphone for both Abstract and Concrete Words; however, with the first the difference was more marked (p < .001) than with the second (p < .01). In addition, responses with the Microphone in trend were faster with Abstract than with Concrete Words (*p* =.09).

The interaction was also significant in a further ANOVA we performed on errors with the same factors, *F* (1, 17) = 35.62, *MSe* = 0.80, *p* <. 001. Post-hoc LSD showed that, as predicted, Abstract Words elicited more errors than Concrete Words with the Keyboard (*p* < .001), while they elicited less errors than Concrete Words with the Microphone (*p* < .05). Abstract Words using the Keyboard produced more errors than all other conditions except Concrete Words using the Microphone. Whereas Concrete Words using the Keyboard produced fewer errors than all other conditions except for Abstract Words using the Microphone (see Tab. 3).

EXPERIMENT 3

Categorical Recognition: without/with label

Language	Before learning the label	After learning the label
	1.01 (0.68)	**0.29 (0.53)**

Words-Objects Match

Word	Concrete	Abstract
	2.20 (1.99)	**4.46 (3.63)**
Other Exemplar	Novel Exemplar	Learned Exemplar
	2.08 (2.49)	**4.57(3.22)**

Property verification task: keyboard vs. microphone

	Device	
	Keyboard	Microphone
Word Abstract	**4.17 (1.87**	**2.87 (1.34)**
Concrete	**2.35 (1.72)**	**3.56 (1.92)**

Table 3. Experiment 3 - Errors and standard deviations (in parenthesis) in percentage for each TEST.

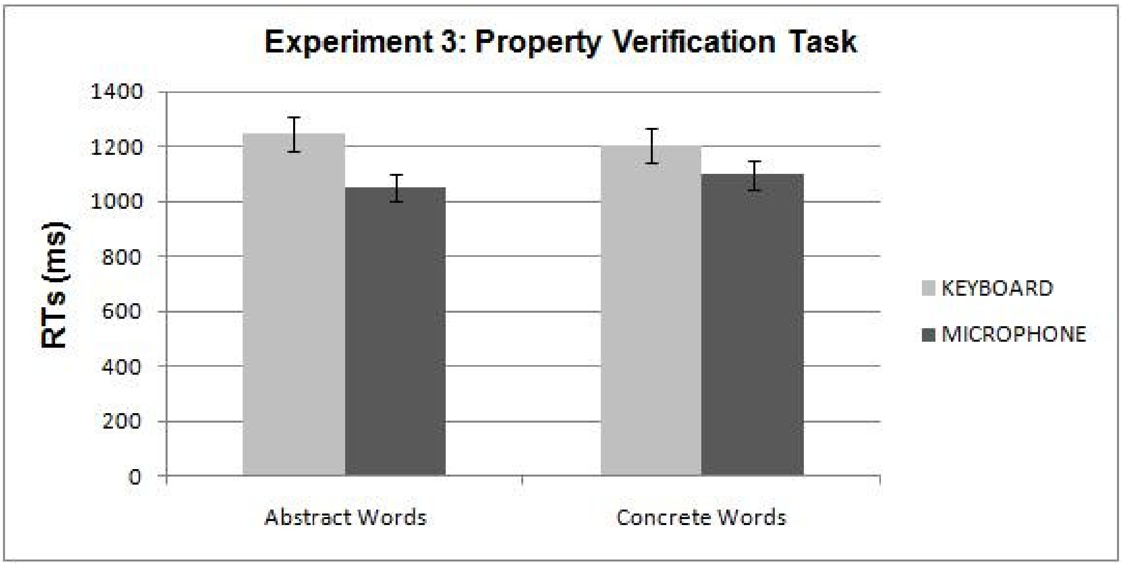


Figure 4. Experiment 3 - Interaction between Word (Abstract, Concrete) and Response Device (Keyboard, Microphone).

7.4.3. Discussion

Results of Experiments 3 confirmed and extend those of Experiment 2, eliminating some potential problems. Differently from Experiments 1 and 2, in TEST 1 (Categorical Recognition Task) we found no difference between abstract and concrete words, probably due to the fact that adding a property (color) to referents of abstract words increased their difference from contrast categories. Interestingly for us, in this experiment results of TEST 1 allowed us to conclude that the introduction of category labels facilitated categorization. The comparison between the same tasks performed before and after the linguistic training reveals this.

In TEST 2, the same pattern of results as Experiment 1 and 2 emerged: abstract words elicited more errors than concrete ones, thus confirming their higher complexity as well as the fact that their borders are not so clearly marked as those observed between referents of concrete words.

In TEST 3, as expected, the advantage of abstract words over concrete ones disappeared. This confirms that it was due to the modifications we made: we introduced color differences between the entities to which abstract categories referred to, in order to be certain that the task did not differ in difficulty for concrete and abstract words. The interaction between Response device and Words revealed that responses with the keyboard were always slower than responses with the microphone but that the discrepancy between microphone and keyboard was more marked with abstract than with concrete words. The pattern was complemented by the results on errors, which were fully in line with our predictions: more errors were elicited by abstract words using the keyboard, and by concrete words when using the microphone.

7.5. Experiment 4

The two last experiments left two issues unsolved. In Experiment 2 we manipulated the presence of explanations, but only for abstract words. In Experiment 3 participants were given explanations to clarify abstract word meanings because this would mirror the typical acquisition process of abstract categories. However, manipulating explanations only for abstract words did not allow us to precisely determine if there is an effect of explanation also for concrete words. Therefore, in Experiment 4 we presented only the category label or the label and the explanation for both concrete and abstract words. In addition, in this experiment for abstract words the information provided by perceptual input and that provided by the verbal label plus explanation were disentangled. To dissociate these two sources of information we used different colors for the members of abstract categories, in order to induce participants to categorize them on the basis of color, but the labels and explanations for these items still rested on items' reciprocal interaction, rather than on their perceptual features. Therefore, with concrete items the label and the explanation converged with the category formed on the basis of perceptual Experience (Training 1), whereas with abstract items the verbal and perceptual experience did not match. This manipulation was introduced in order to verify whether the advantage of the microphone responses was simply due to phono-articulatory aspects of the words or to their conceptual content as well. Our major predictions concerned TEST 3: 1) If the mouth activation found in Experiment 3 (TEST 3, vocal responses) is due to a motor phono-articulatory activation pertaining to the superficial linguistic information, in Experiment 4 (TEST 3) we should find an advantage of vocal responses both with concrete and abstract words, as well as a main effect of the verbal explanation. 2) If, consistent with the WAT proposal, the previously found advantage for vocal responses pertains also the category content, then it should play a major role if it complements information given by perception and action, not if it contrasts with it. Therefore we should find a difference with results of Experiment 3: there should be an advantage of the microphone over the keyboard only when the label and the explanation do not contrast with perceptually-based categories. In this experiment, this contrast characterizes abstract categories.

7.5.1. Method

7.5.1.1. Participants

Eighteen students of the University of Bologna took part in the study (7 men; mean age = 24.55 years; standard deviation = 3.66). All were native Italian speakers and right handed.

7.5.1.2. Procedure

The procedure was similar to that of Experiment 3, except for two variations. First, during Training 2 (Words Acquisition) half of the participants were taught the linguistic labels (Label group) vs. the linguistic labels plus the verbal explanation (Label+Explanation group), both for abstract and concrete items. The verbal explanations for abstract items were the same used in Experiment 2 and 3, so they basically described the kind of interaction. For concrete items the verbal explanations focused on the figure shape, avoiding any reference to its color (e.g., CALONA: “a figure having a concavity”). The number of words for each explanation across both the abstract and the concrete blocks was even. Second, in Experiment 4 we used different colors for each category member: for both concrete and abstract items, two members of each category had the same color as two members of another category. For example, FUSAPO surface could be yellow, blue, red or sky blue; its thickness was always the same, i.e. dark blue. NOROLO surface shared with FUSAPO surface yellow and blue colors, but it could be also green or violet; the color of the thickness was always dark blue.

7.5.2. Results

7.5.2.1. TEST 1: Categorical Recognition

We performed two different ANOVAs on errors: one for the Label group and another for the Label+Explanation group. In the first ANOVA two factors were manipulated within participants, the factor Concept (Concrete vs. Abstract), and the factor Language (Before vs. After learning the category label). In the second ANOVA we manipulated the same factors, but the levels of Language factor differed (Before vs. After learning the category label with explanation). In the first ANOVA, both main effects were significant: more errors were produced with Abstract (M = 7.41%) than with Concrete Concepts (M = 3.36%22), *F* (1, 8) = 7.73 , *MSe* = 19.12, *p* < .05, and more errors were produced before (M = 6.54%) than after learning the label (M = 4.22%), *F* (1, 8) = 17.31, *MSe* = 2.79, *p* < .01. The factor Concept was also significant in the second ANOVA: more errors were produced for Abstract (M = 11.17%) than for Concrete Concepts (M = 3.60%), *F* (1, 8) = 32.61, *MSe* = 15.38, *p* < .001. The factor Language did not reach significance, but we found a significant interaction between Concept and Language, *F* (1, 8) = 7.26, *MSe* = 1.83, *p* < .05 (see Table 4), due to the fact that after learning label+explanation errors decreased with concrete words (LSD post-doc, *p* < .01), but not with abstract ones.

7.5.2.2. TEST 2: Words-Objects Match

We performed two different ANOVAs on errors: one for the Label group and another for the Label+Explanation group. In the first ANOVA two factors were manipulated within participants: Word (Concrete vs. Abstract) and Other Exemplar (Exemplar already learned, with only linguistic label vs. Exemplar not learned). In the second ANOVA the same factors were manipulated, but the levels of the Other Exemplar factor differed (Exemplar already learned, with label+explanation vs. Exemplar not learned).

In both ANOVAs we found a significant main effect of the factor Word: fewer errors were produced with Concrete than With Abstract Words (group A: M = 2.55%22, M = 6.48% respectively, *F* (1, 8) = 8.31, *MSe* = 16.77, *p* < 0.05; group B: M = 2.66%, M = 7.29% respectively, *F* (1, 8) = 22.13, *MSe* = 8.71, *p* < .01; see Tab. 4).

7.5.2.3. TEST 3: Property verification task with keyboard and microphone

In TEST 3 for RTs 35.63% of the trials was removed as errors. We used the same trimming method as previous experiments; this lead to the removal of 2.38% of the data. An ANOVA was performed with three factors: Word (Abstract vs. Concrete), Response Device (Keyboard vs. Microphone) and Verbal Explanation (Without vs. With), the last one manipulated between participants. We found that vocal responses (M = 1128.73 ms) were 147.57 ms faster than manual responses (M = 1276.30 ms), even if the factor Response Device did not reach significance, *F* (1, 16) = 3.48, *MSe* = 112633, *p* = .08. The interaction between the factors Word and Response Device was significant, *F* (1, 16 = 4.58), *MSe* = 47804.8, *p* < .05. The advantage of the microphone over the keyboard did not reach significance with abstract words (M = 1221.67 vs. M = 1184.37, respectively), while with concrete words responses with the microphone (M = 1073.09) were faster than responses with the keyboard (M = 1330.93) (LSD, *p* <.01) (see Fig. 5).

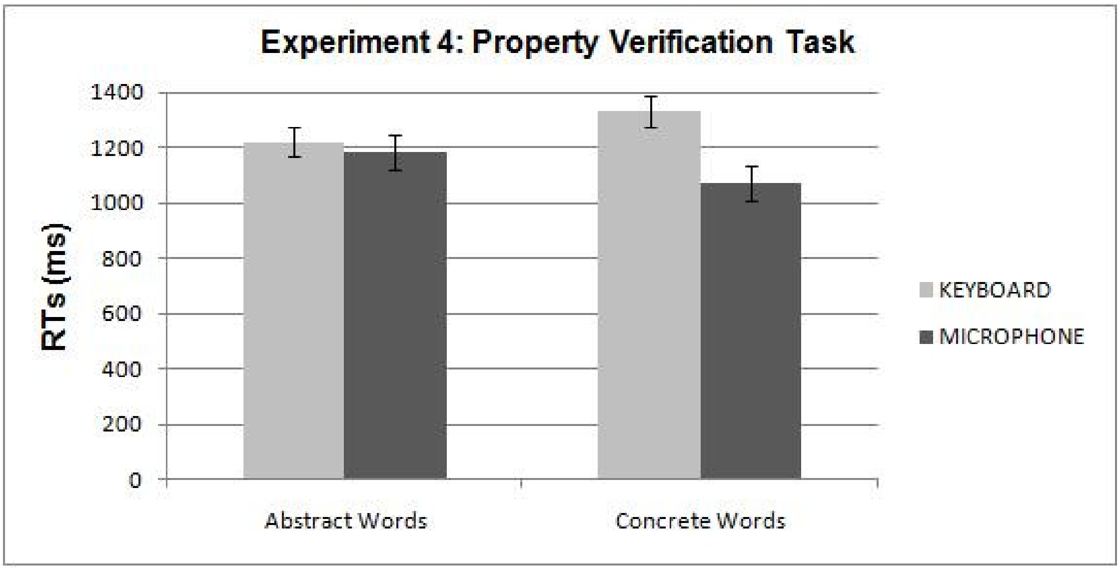


Figure 5. Experiment 4 - Interaction between Word (Abstract, Concrete) and Response Device (Keyboard, Microphone).

Finally in the ANOVAs on errors with the same factors, we found that abstract words (M = 20.01%) elicited more errors than concrete ones (M = 15.63%), *F* (1, 16) = 7.84, *MSe* =

44.13.08, *p* < .05. The significant interaction between Word and Response Device, *F* (1, 16) = 5.87, *MSe* = 37.90, *p* < .05, was due to the fact that abstract words with the microphone (M = 21.79%) elicited more errors than concrete words with both the keyboard (M = 17.36%) and the microphone (M = 13.89%) (LSD post-hoc, *p* <. 05) (see Tab. 4).

EXPERIMENT 4

Categorical Recognition: without/with label

Group A: label	Concept	
	Concrete	Abstract
	3.36 (2.59)	**7.41 (5.04)**
Group B: label+explanation	Concept	
	Concrete	Abstract
	3.70 (2.35)	**11.17(4.47)**

Words-Objects Match

Group A: label	Word	
	Concrete	Abstract
	2.55 (4.08)	**6.48 (4.57)**
Group B: label+explanation	Word	
	Concrete	Abstract
	2.66 (4.45)	**7.20 (4.36)**

Property verification task

	Device

	Keyboard	Microphone
Word Abstract	**18.23 (8.37)**	**21.79 (6.17)**
Concrete	**17.36(6.38)**	**13.89 (5.19)**

Table 4. Experiment 4 - Errors and standard deviations (in parenthesis) in percentage for each TEST.

7.5.3. Discussion

Results of TEST 1 indicate that the difference between the condition Label and No-Label increases when an explanation is added to the category name. Thus explanations facilitate categorization, as they render clearer category boundaries. However, the contribution of explanations is relevant only for concrete categories. For abstract categories, explanations do not help, as the information they provide is in contrast with perceptually based categorization.

Results of TEST 3 are the most intriguing. As predicted, participants were faster to respond with the microphone than with the keyboard with all words: this suggests that the phono-articolatory aspect of the words pronounced during acquisition affects performance. It is unclear, however, why no effect of explanation was present. The most important result is the interaction showing that the advantage of the microphone over the keyboard is more marked with concrete than with abstract words, both in RTs and accuracy. This suggests that not only phono-articolatory but also conceptual information is at play in explaining the advantage of responses with the microphone. In fact this advantage shows up only when there is a convergence between the linguistic information (label and explanation) and the category formed on sensory-motor basis, that is only with concrete words. One could object that the effect is due to the fact that explanations used with concrete words might have reduced the importance of manual interaction with objects. However, this doesn’t account for the advantage of the microphone with concrete over abstract words.

7.6. General discussion

Four experiments were designed to study the acquisition of concrete and abstract categories and words. We chose to use novel categories, in order to avoid confounds often associated with research on concrete and abstract words. We identified some characteristics which are typical of abstract but not of concrete categories, and we created novel categories according to these criteria. First, abstract categories do not refer to a single object but rather to a complex relationship between different objects. In addition, the entities to which abstract categories refer are not manipulable, even though they are perceivable, as they are interacting moving objects. Notice that our distinction does not cover the whole continuum ranging from abstract to concrete categories. Further work is needed for a thorough investigation of different typologies of concrete and abstract words (for attempts in this direction, see Setti & Caramelli, 2005). Here we used two different examples of concrete and abstract words and have shown that different processes are involved in their acquisition.

In Experiment 1 we controlled, using a standard properties production task, that the pattern of produced properties with our novel concrete and abstract categories was similar to that typically elicited by concrete and abstract words. In Experiment 2, 3, and 4 we introduced a modification: abstract words were not only learned by associating a label with the entities/relations they referred to, but also when an explanation of their meaning was provided. This learning situation should resemble the learning process of children, as studies on MoA show. We found that this learning process influenced a later property verification task: participants responded earlier to concrete words while using the keyboard, while responses with abstract words were faster while using the microphone. Similar results with action words and effectors showed that, while comprehending sentences referring to mouth-related actions, response times were faster with the microphone than with the keyboard

(Scorolli & Borghi, 2007). In addition, in line with WAT, participants performance with abstract words was improved when provided with a verbal explanation (Experiment 2, group B; Experiment 3). This effect was not observed in concrete words. The fact that the advantage of the explanation was confined to abstract words revealed that the difference is not simply due to phono-articulatory aspects, but that for accessing the meaning of abstract words linguistic information plays a major role. This was confirmed in Experiment 4, in which we found that, due to the fact that with abstract words the verbal label and explanation were in contrast with the already formed perceptually-based category, the advantage of the microphone over the keyboard was reduced compared to the other experiments.

Our results are in line with embodied and grounded theories of categorization and language comprehension. Namely, both the concrete and the abstract categories we used are grounded, as they have objects or relations as referents. We were able to demonstrate that they are not only grounded in perception and action systems, but that for forming them language is important. This leads to the prediction that abstract words would not only activate linguistic areas in the brain, but also classic motor and sensory-motor areas.

Results are in line with the predictions advanced by the WAT proposal. They reveal that a different learning process might lead to differences in performance on different tasks, such as a production task versus a property verification task. In addition, it provides evidence that for representing abstract concepts motor linguistic information is more important than manual information, whereas for representing concrete concepts the pattern is opposite.

In addition, our results clearly show that the formation of both concrete and abstract categories benefits from learning a linguistic label. As it emerges from the categorical recognition task in Experiment 3, language is relevant because it helps to better differentiate between categories (Mirolli & Parisi, 2011). The recognition test in Experiment 4 (TEST 1a-1b) shows that labeling is mostly helpful when an explanation of the category meaning is added. As shown in test 2 and test 3 of Experiment 2, the benefit provided by language is higher in the case of abstract categories when a verbal explanation (group B) supports the linguistic label. Nevertheless, when no explanation is provided, labeling is useful for both concrete and abstract categories. In sum: labeling helps categorization, independently of category complexity. However, even when no explanation is provided, given that abstract words do not refer to manipulable objects and are linked by complex relational properties, language plays a major role in their representation.

This opens an interesting scenario. Language is relevant for both concrete and abstract words because it helps better differentiate between categories. However, in tasks for which categorization is not relevant, such as the color verification task, it is more accessible in the representation of abstract than of concrete word meanings. This might occur because: a. the members of abstract categories are not manipulable; and b. more linguistic information is typically associated with the acquisition of abstract word meanings.

Further work should address unsolved issues in this research. One important expansion could be to introduce the social development component implied in word acquisition. We used language in a very simple way, through adding labels or explanations to read and to associate with the relevant categories. Thus, language was not associated with a real social experience, as in real life. Further work should take aspects of social development which characterize language acquisition into account. A further limitation is that the variety of real-life concrete and abstract words is very broad, and we were able to model only a subset of these.

Finally, we believe this work has important implications for modeling. The design of the task is particularly suitable for further modeling applications and replication. We succeeded in isolating some properties we believe to be relevant in real life categories and built novel categories based on our assumptions. We could verify that the behavioral responses produced within these novel categories were similar to the ones produced within real life categories and settings. This procedure demands an additional modeling development. We believe that computational models can integrate and generate a more general description the experimental results. For example, a robotic model, as discussed in the introduction, can benefit from a psychological theory, that provides a possible way to tackle a new and complex problem for robotics itself, such the theory described focusing on the grounding and acquisition of abstract words. On the other hand, the same robotic model can be tested in many different experimental situations, some of them not even applicable to human subjects. Experiments of this nature can complement and integrate experiments with human participants and can offer new insights and hypotheses to test. Furthermore, the process of developing artificial cognitive models always requires a profound articulation of the theory implemented. This fact forces the researcher to well define and to operationally describe every aspect of the theory and, at the same time, it emphasizes the importance of the central aspects of the theory, that can be fully exploited and validated by the model, at least as a preliminary proof of concept.

8. The activation of the left-right mental timeline is context-dependent

The central assumption of Conceptual Metaphor Theory (CMT, Lakoff & Johnson, 1980, 1999) is that the conceptual metaphors observed in every-day language are at work at a deep cognitive level. This basic principle of CMT is supported especially by those empirical findings on conceptual metaphors that discovered the existence of metaphoric mappings which are not explicitly attested in language. In par. 3.5, CMT view was related to the possible grounding of abstract concepts/words. Indeed, in the last years one of the most studied example of implicit conceptual mappings has been the TIME IS SPACE metaphor which maps temporal reference onto a horizontal spatial axis (e.g., left-to-right in Western cultures). Thus, the TIME IS SPACE metaphor consists in the use of a concrete, spatial schema (i.e., the mental timeline) to organize and represent the abstract contents of temporal expressions. In the literature, a growing body of research has investigated the characteristics of this mapping, but one important aspect that remains undetermined is whether these space-time mappings can be activated automatically (i.e., independently from the goals of the task). Prior attempts to settle this issue have provided mixed, unclear results. In this study, the design of prior investigations has been improved by performing both a time-relevant and a time-irrelevant task (i.e., temporal reference judgment and lexical decision, respectively) on the very same pool of tensed Spanish verbs and pseudo-verbs (i.e., both categories of stimuli were endowed with past and future tense markers). This allowed a trustworthy statistical comparison of their results (i.e., overall between tasks ANOVA). Furthermore, as in previous research only reaction times and accuracy were measured, in the present study also the MouseTracker software were used, in order to record the participants' responses mouse trajectories, as they can be considered as a more sensible index of the decision process. On the whole, the results confirmed that the left-right space-time mapping is active only when the task requires temporal discrimination. This finding support a flexible account for the activation of the TIME IS SPACE metaphor, and confirm the general hypothesis on conceptual effects that guides this thesis.

8.1. Introduction

How are internally represented concepts grounded in the external world? This question is the Symbol Grounding Problem (Harnad, 1990), one of the central problems of cognitive science (Barsalou, 1999; Glenberg, 1997; Johnson-Laird, 1983; Lakoff & Johnson, 1999; Santiago et al., 2011). The "hardest part" of the problem regards abstract concepts, which by definition refer to things that cannot be seen or touched (e.g., Borghi et al., 2011). In search of possible solutions, embodied and grounded cognition theories (e.g., Barsalou, 1999; Gibbs, 2006; Glenberg, 1997; Zwaan, 2004) suggest that in order to gain meaning, abstract concepts are grounded on more concrete dimensions, which are in turn grounded on sensory-motor experiences (see chapter 3).

Under the embodied and grounded view, language processing elicits an embodied simulation which is carried out by the very same neural systems used by perception, action and emotion (e.g., Barsalou, 2008; Gallese, 2008; Glenberg & Gallese, 2011; Glenberg et al., 2008; see chapter 3 and 7). When abstract concepts are referred to, such simulation follows the guide of stored mappings between abstract and concrete concepts. One of the strongest line of support for this idea comes from empirical studies on the abstract domain of TIME, as grounded on the concrete domain of SPACE. Indeed, a large number of studies have reported interactions between the processing of the temporal reference of words and sentences and irrelevant spatial dimensions of the task, such as lateralized keypresses or word presentation, forward-backward manual movements, and front-back word presentation (e.g., Boroditsky, 2001; Ouellet et al., 2010; Santiago et al, 2007; Sell & Kaschak, 2011; Torralbo et al., 2006; Ulrich & Maienborn, 2010; Ulrich et al., 2012). Analogous conceptual congruency effects have been reported in many other studies that focus on different concrete and abstract dimensions, such as left-right space and number (Dehaene et al., 1993), vertical space and evaluation (Meier & Robinson, 2004; Santiago et al., 2012), brightness and evaluation (Meier et al., 2004), or vertical space and power (Schubert, 2005). Conceptual congruency effects are usually interpreted as indexing the use of underlying concrete representations to think about the abstract dimension (see Santiago et al., 2011, 2012).

Although now well established, much research remains to be done on the boundary conditions of conceptual congruency effects. More specifically, it is still an open question whether such effects arise automatically whenever the concrete and abstract dimensions are processed simultaneously. One of the strongest cases of automatic activation in the literature concerns the mapping of affective evaluation to front-back responses. Chen & Bargh (1999) reported that participants were faster responding to positive items with a pulling response, and to negative items with a pushing response, and this occurred not only when the decision was explicitly based on the valence of the stimuli, but also when the subjects performed a lexical decision task (Wentura et al., 2000) and even a simple stimulus detection task (Chen & Bargh, 1999), which really minimized the task-relevance of the evaluative dimension. The mapping of numbers on left-right space has also been shown to be quite automatic (e.g., Fisher et al., 2003; Ruiz Fernández et al., 2011). In contrast, space-time mappings do not seem to be activated so automatically. Recently, two studies extending prior findings with temporal words (e.g., Santiago et al., 2007) to full sentences aimed to test the automaticity of the mental time-line in congruency tasks involving both the left-right (Ulrich & Maienborn, 2010) and front-back (Ulrich et al., 2012) spatial axes. These studies presented sensible and non-sensible sentences in a go-no go procedure: participants only responded when the sentence made sense. In one task (time-relevant), participants were required to decide whether the correct sentence referred to the past or the future by either pressing a left or a right button (Ulrich & Maienborn, 2010) or moving a lever forward or backward (Ulrich et al, 2012). Results in the time-relevant tasks were compared to time-irrelevant variants, where the responses were used to discriminate between sensible and nonsensical sentences. The results in both studies showed space-time congruency effects only on the former (time-relevant) task.

However, it is difficult to defend the absence of an effect. The findings of Ulrich & Maienborn (2010) on the left-right mental time line left open the possibility that participants simply did not need to process the temporal reference of the sentences to successfully complete the lexicality judgment task. The non-sensible sentences were constructed by matching agents and objects that did not fulfill the meaning restrictions of the verb (i.e., as in the past sentence "The fir trees have put on their coat while bathing", or in the future sentence "On next Sunday, the town-hall will marry the pea"). In order to judge whether these sentences were correct or not, participants might have only assessed whether the action mentioned by the verb could be done by the actor (with the object) on the patient. In order to address this concern, Ulrich et al. (2012) adapted the dual-task procedure of Ouellet et al. (2010) in the time-irrelevant experiment. After each response, participants were asked to answer a question about the temporal reference of the prior sentence, thus assuring that their temporal meaning was grasped. This experiment also failed to find any trace of interaction between front-back responses and past-future reference. Although it relieves concerns about the depth of processing of sentence meaning in the front-back task, no similar control is available for the left-right task of Ulrich & Maienborn (2010). It might be the case that the activation of the front-back timeline is not automatic, but we cannot be certain that this is true also for the left-right mental timeline.

In addition, their time-relevant and time-irrelevant tasks differed in several other respects. Firstly, the materials included equal amounts of sensible past, sensible future, and nonsensical sentences, what means that participants responded on 60% of the trials in the time-relevant task, whereas they did not have to hold any responses in the time-irrelevant task. This may have made the latter easier than the former. In the same direction, their time-relevant tasks interleaved two judgements (on lexicality and time), whereas the left-rigth time-irrelevant task of Ulrich & Maienborn (2010) only required judging lexicality, and the front-back time-irrelevant task of Ulrich et al. (2012) allowed participants to focus on each judgement sequentially: initially on lexicality, in order to produce the front-back response, and then on time, in order to answer the final control question). Again, the simultaneous double judgement in the time-relevant tasks may have increased their difficulty.

There is also the possibility that processing complex sentences does not afford automatic activation of the mental timeline. Meaning access at sentence level is less automatic than at word level, because the meaning of the sequence of words needs to be composed into the overall sentence meaning. This latter process could be sustained by different neural structures than those involved in tasks with single words (e.g., Friederici, 2004).

Finally, there always remain the possibility that a more sensitive measure might detect a smaller, but significant congruency effect in time-irrelevant tasks. Thus, the issue of the automaticity in the activation of the mental timeline is still far from being solved.

In the present study we wanted to address simultaneously several of these possibilities. We focused on the processing of time-related single words in tasks requiring left and right responses (thereby testing the activation of the left-right mental time line) in both time-relevant and time-irrelevant conditions. We created linguistic stimuli acquirable in single fixations designed to assure the processing of temporal reference even when it is not task relevant. Participants performed either a temporal reference judgment task (decide whether the stimulus refers to the past or the future) or a lexical decision task (decide whether the stimulus is a word or a nonword). All stimuli (both words and nonwords) were presented in each task, which allowed a subsequent trustworthy statistical comparison of the two tasks results. Finally, we measured participants reaction times in Experiment 1, as in previous studies, and also mouse trajectories (using MouseTracker software; Freeman & Ambady, 2010) in Experiment 2. Indeed, we reasoned that mouse-tracking has the potential of capturing the conceptual congruency effect during the real-time unfolding of the response, as it has been shown in several studies (e.g., Barca & Pezzulo, 2012). If the left-right space-time mapping can be activated automatically, at least one of the experiments will capture the conceptual congruency effect in both the time-relevant and time-irrelevant tasks. Otherwise, this effect should arise only in the time-relevant task.

8.2. Experiment 1

Experiment 1 used centrally presented Spanish tensed verbs (technically corresponding to full sentences as Spanish is a PRO-DROP language) and nonverbs in both an explicit temporal judgment task, where temporal reference was task relevant, and a lexical decision task, for which temporal reference was irrelevant. The design of the experimental materials made sure that temporal reference information was equally present, salient, and could be acquired in a single fixation, in both the words and the nonwords. Responses were given by means of bimanual lateralized keypresses, and participants' response latencies and accuracy were measured.

8.2.1. Method

8.2.1.1. Participants

Forty-eight Psychology students from the University of Granada (6 males; age range 19-26 y.; 6 left-handed by self-report) participated for course credit. All participants had normal or corrected-to-normal vision and were naive as to the purposes of the experiment. Participants were randomly assigned to two groups: 24 participants carried out the temporal judgment task, and the rest carried out the lexical decision task.

8.2.1.2. Materials

We selected 148 Spanish intransitive verbs (or with at least one very common intransitive use). Such a kind of verbs was chosen because Spanish is a PRO-DROP language, so the subject of a verb can be dropped from the sentence. Thus, single conjugated intransitive verbs as used here can stand as full, grammatically correct sentences. In order to create the nonword set, each verb was modified by changing one letter in their morphological stem, with the only constraint of resulting in sequences of pronounceable phonemes in Spanish. Therefore, the nonverbs did not pop out as such (e.g., “dormir” was changed to “dorpir”).

The 148 verbs and 148 nonverbs were then conjugated in both the simple past perfect indicative and the simple future indicative, and all six possible grammatical persons were almost equally represented over the whole set (the choice of grammatical person avoided ambiguous forms such as “amamos” which means both “we love” and “we loved”). The nonverbs thus also carried past and future inflections (e.g., “durpió”). Therefore, in order to distinguish the verbs from the non-verbs, participants had to pay close attention, and deeply elaborate the stimulus also in the lexical decision task. This resulted in 592 experimental stimuli of four types: past and future verbs, and past and future nonverbs. This total set was randomly divided into four lists of 148 stimuli each, avoiding item repetition. For example, from the item “faltar/falbar” the following four third person singular tensed versions were created: “faltó” (“he failed” - past verb), “faltará” (“he will fail” - future verb), “falbó” (past nonverb), “falbará” (future nonverb). Each of these different versions of the original item was randomly assigned to one of the four different lists with no item repetition. Each list was composed of 37 items from each of the four stimulus type conditions. Direct control for factors already known to affect word recognition times, such as frequency, length, age of acquisition, and so on, was not required because the theoretically relevant effect we were looking for was the interaction between temporal reference and response hand when participants processed the very same list of stimuli using the two possible response-key mappings.

8.2.1.3. Procedure

Stimuli were presented at the center of a LCD computer screen (Courier New font, 38 points, lower case), black printed on white background. Participants sat at a distance of 60 cm from the computer screen, and placed their left index finger on the Q key and their right index finger on the 9 key of the numerical keyboard in a standard Spanish QWERTY keyboard. The distance between response keys was 32 cm. Each trial began with the presentation of a central fixation cross (500 ms) followed by the target verb, that remained on the screen until a response was made. Incorrect trials were followed by a 500 ms red uppercase “X” at the same location of the stimulus. Each incorrect trial was then followed by a 1000 ms blank screen. Correct trials were followed by a 1500 ms blank screen. Participants in the time-relevant condition were instructed to decide whether the presented verb or nonverb referred to either the past or the future. Participants in the time-irrelevant condition decided whether the stimuli were real Spanish verbs or not.

Each task was composed by two experimental blocks of 148 trials (separated by a two minutes break) in which the same list of stimuli were responded to using a different mappings of responses (past/future or correct/incorrect) to keys (left/right). The order of presentation of the two mappings was counterbalanced over participants.

Each block was preceded by a short four trials training with a different set of stimuli to familiarize participants with the procedure. Overall, each subject responded to 296 experimental trials, plus 8 training trials. The experiment was programmed and run using E-prime 2.0 (Psychological Software Tools, Inc., Pittsburgh, PA).

8.2.1.4. Design and Analysis

Data were analyzed using a mixed factorial ANOVA with the between-participants factor Task (temporal judgment vs. lexical decision) and the within-participant factors Lexical status (word vs. nonword) x Temporal reference (past vs. future) x Key (left vs. right). Counterbalance was included in the design as a between-subjects factor in order to reduce noise, but it is not reported further because its main effect or interactions had no theoretical relevance.

8.2.2. Results

There were errors in 6.43% of trials in the temporal judgment task and 5.19% of trials in the lexical decision task. Errors and latency in correct trials were analyzed independently. Reaction times (RTs) were trimmed by mean of fixed cut-offs that left out a pre-established proportion of 2% of trials. In temporal judgement, cut-offs were set at 250 ms and 2400 ms, what left out 2.13% of the trials. In lexical decision, RTs out of the temporal window between 250 ms and 1750 ms were cut off, leaving out 1.97% of trials.

The ANOVA on RTs revealed three significant main effects, and showed also five significant interactions. We report the results starting from the more important for our hypotheses. First of all, there was an overall significant interaction between Temporal reference and Key (*F*(2, 92) = 4.13, *MSe* = 10857.24, *p* <.05), which showed faster responses to past verbs and nonverbs with the left than with the right key (left key *M* = 959 ms vs. right key *M* = 963 ms) and to future verbs and nonverbs with the right than with the left key (left key *M* = 986 ms vs. right key *M* = 947 ms). Of central importance, the interaction of Temporal reference and Key was modulated by Task (*F*(3, 184) = 4.39, *MSe* = 10857.24, *p* <.05; Temporal judgment: Past verbs - left key *M* = 1069 ms, right key *M* = 1095 ms; Future verbs - left key *M* = 1084 ms, right key *M* = 1022 ms - Newman-Keuls *all ps* <.05. Lexical decision: Past verbs - left key *M* = 848 ms, right key *M* = 833 ms; Future verbs - left key *M* = 887 ms, right key *M* = 873 ms) (see Fig. 1). This three-way interaction is due to a null effect of space-time conceptual congruency in lexical decision, whereas the effect was clear in time judgement. This is in line with the findings of Ulrich & Maienborn (2010).

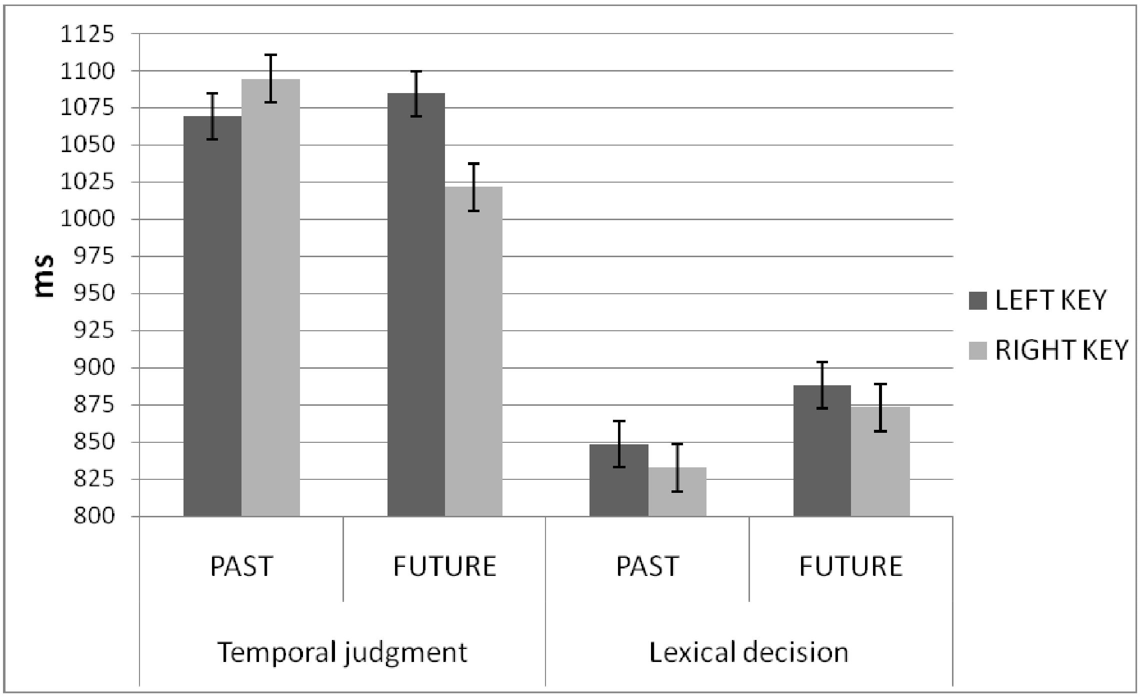


Figure 1. Experiment 1 - Mean latencies for the interaction Task x Temporal reference x Key.

To better understand this crucial interaction, we decided to collapse the two levels Temporal reference and Key in a Compatibility factor (congruent vs. Incongruent trials) and to report also the Task x Compatibility interaction (*F*(2, 92) = 4.39, *MSe* = 5428.62, *p* <.05; Temporal judgment: congruent *M* = 1046 ms, incongruent *M* = 1089 ms; Newman-Keuls *p* <.01. Lexical decision: congruent *M* = 861 ms, incongruent *M* = 860 ms) in Figure 2.

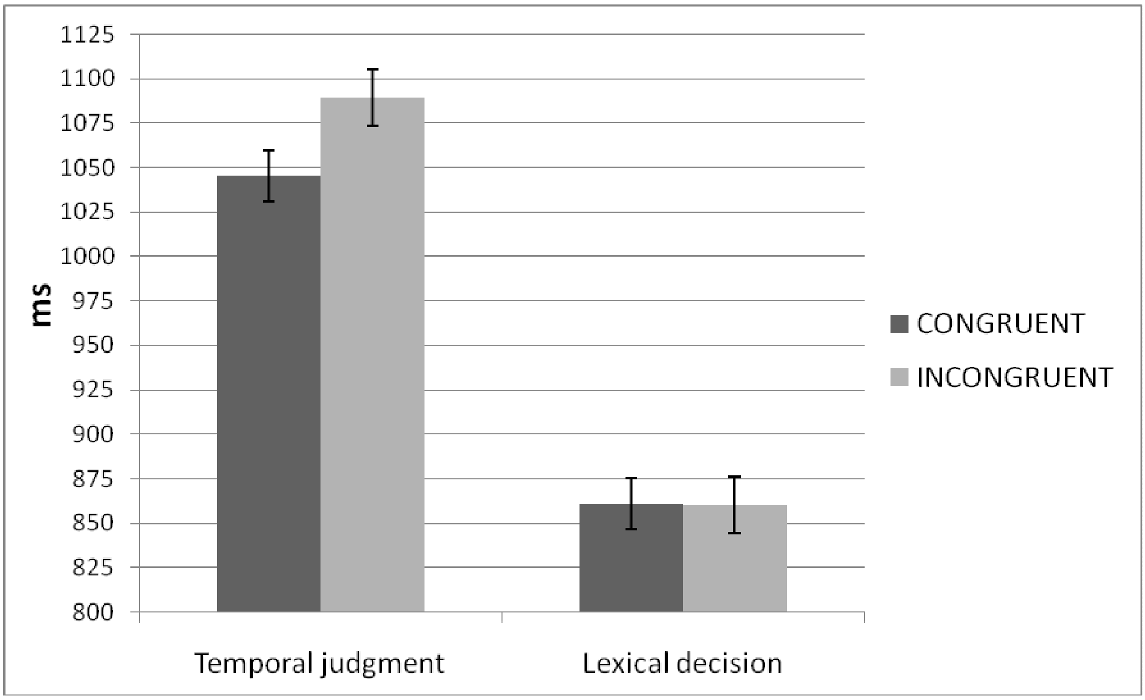


Figure 2. Experiment 1 - Mean latencies for the interaction Task x Congruency.

Less importantly for our current purposes, there was also a clear effect of Lexical status (*F*(1, 46) = 122.04, *MSe* = 8172.35, *p* <.001), due to longer latencies being observed for nonwords (*M* = 1015 ms) than words (*M* = 913 ms). This is in line with the psycholinguistic literature about the lexicality effect (e.g., Kinoshita el al., 2004; Pagliuca et al., 2010). The main effect of Task was significant too (*F*(1, 46) = 17.72, *MSe* = 232148.58, *p* <.001), showing longer mean latencies for the temporal judgment (*M* = 1067 ms) than lexical decision (*M* = 861 ms). The final significant main effect was Key (*F*(1, 46) = 6.80, *MSe* = 4123.65, *p* <.05), with longer latencies for the left key (*M* = 972 ms) than the right key (*M* = 955 ms).

Additionally, other three interactions were reliable. First, the Task x Lexical status interaction (*F*(2, 92) = 5.01, *MSe* = 8172.35, *p* <.05), due to the lexicality effect being smaller in the lexical decision task (word *M* = 820 ms vs. nonword *M* = 901 ms) than in the temporal judgment task (word *M* = 1006 ms vs. nonword *M* = 1129 ms). This interaction supports that participants paid close attention to lexical status in the time judgement task. Second, the Task x Temporal reference interaction (*F*(2, 92) = 22.79, *MSe* = 5020.66, *p* <.001), due to responses to past stimuli being faster than to future stimuli in lexical decision (past *M* = 841 ms vs. future *M* = 881 ms), whereas the opposite occurred in temporal judgment (past *M* = 1082 ms vs. future *M* = 1053 ms). Finally, the interaction of Task, Lexical status and Temporal reference was significant as well (*F*(3, 184) = 6.45, *MSe* = 3192.59, *p* <.05; Temporal judgment: Word - past verbs *M* = 1011 ms, future verbs *M* = 1001 ms; Nonword - past verbs *M* = 1153 ms, future verbs *M* = 1105 ms - Newman-Keuls *p* <.001. Lexical decision: Word - past verbs *M* = 805 ms, future verbs *M* = 835 ms - Newman-Keuls *p* <.05; Nonword - past verbs *M* = 876 ms, future verbs *M* = 926 ms - Newman-Keuls *p* <.001).

No other main effects nor interactions were significant.

The analysis of accuracy revealed only a lexicality main effect (*F*(1, 46) = 15.85, *MSe* = 0.01, *p* <.001), showing the greater easiness of processing words than nonwords (95% vs. 93% correct responses, respectively). No other main effects nor interactions were significant.

8.2.3. Discussion

In Experiment 1 we found a significant interaction between the temporal reference of the stimulus (past vs. future) and the side of response (left key vs. right key) in the time-relevant temporal judgement task, taking the form of a standard left-right space-time congruency effect: responses to past stimuli were faster with the left finger and responses to future stimuli were faster with the right finger, independently of their lexicality. This was not observed in the time-irrelevant lexical decision task.

Thus, Experiment 1 replicates prior findings in the literature (Torralbo el al., 2006; Santiago el al., 2007; Ulrich & Maienborn, 2010) suggesting, firstly, that a left-to-right mental time-line is activated in the temporal judgment task; and secondly, that this only occurs when the processing of temporal reference is task relevant (i.e., when it is required to select the correct response). This occurred in spite of both tasks being based on exactly the same set of stimuli, short single words and nonwords that could be apprehended in one fixation and which all of them carried morphological markers of tense. In the temporal judgment task, the emergence of the effect was independent from the lexicality of the stimuli, thereby confirming that the design of the materials made sure that the information about temporal reference was equally present and salient in both words/nonwords. Present data relieve several concerns raised in the introduction regarding the automaticity of the activation of the left-right mental timeline. The possibility remains, however, that a more sensitive measure may reveal a space-time congruency effect in the time-irrelevant lexical decision task. In order to test this possibility, Experiment 2 collected mouse trajectories instead of reaction times.

8.3. Experiment 2

Experiment 2 was almost identical to Experiment 1, changing only the data collection technique. Instead of reaction times, we chose a kinematic measure (mouse movements), which was collected using the MouseTracker Software Package (Freeman & Ambady, 2010). This software records the stream of x-y coordinates of participants' mouse trajectories, allowing a precise characterization of both the spatial and temporal dynamics of the responses. Thus, we were able to track continuous hand movements during the task with the aim of observing the graded effects of competition between items attracting the trajectory of the mouse. This technique, which has been successfully adopted in several psychological studies (Freeman et al., 2011), could be considered as a novel complement for more traditional techniques such as the measurement of reaction times. The use of MouseTracker implicated some procedural differences. Participants were instructed to click with a standard mouse used with their dominant hand a start button located at the bottom center of the screen. The stimulus was then presented at the center of the screen, with participants being instructed to respond by mouse-clicking one of two squares displayed at the top corners of the screen. All other methodological aspects of Experiment 2, chiefly the set of materials, the design, and the two types of task, were the same of Experiment 1. As MouseTracker tracks continuous manual reaching movements, the technique allows the study of the online dynamics of the decision process, which is realized over multiple competing possibilities. Thus, it can reveal the uncertainty during response unfolding and its time course. At a theoretical level, deciding between the two response choices (past vs. future, and word vs. nonword) could be described as a continuous and dynamic competition between attractors (Spivey, 2007). These possibilities can push or pull the movement of the hand (so affecting the recorded mouse trajectory) in different directions inside a workspace defined by the locations of the response squares, exactly like how real magnetic attractors would do to the needle of a compass. MouseTracker allows the collection of several measures of continuous performance. There are two temporal indexes. First, there is Initiation Time, which measures the time from the click on the start button and the beginning of mouse movement. Second, there is the Trajectory time, which is analogous to a standard reaction time, as it measures the total time elapsed from the click on the start button and the click on the response box. Third, there is Maximum Deviation Time, which indexes the time between the beginning of mouse movement and the moment of greatest divergence between the mouse trajectory and an ideal straight line between the start and response locations. There are also a variety of spatial measures of decisional conflict. The most sensitive, and the only one we focused on in this study, is the Area Under the Curve (AUC). It is calculated as the geometric area between mouse trajectory and the ideal straight line from the start button to the correct response box (Freeman & Ambady, 2010). Any attraction toward the alternative response box will induce a deviation of the mouse trajectory from this ideal straight line, generating an area between the actual and the ideal trajectory. The greater the attraction, the greater the area. A second, related spatial measure is Maximum Deviation, which is the point of the actual trajectory that is maximally distant from the ideal straight trajectory. Finally, there are the measures of x-flips and y-flips numbers, which refer to the number of direction switches along the horizontal and the vertical axis, respectively. They describe the complexity of the trajectory, offering a further index of indecision in the process of responding.

8.3.1. Method

8.3.1.1. Participants

Forty-eight Psychology students from the University of Granada (6 male; age range 20-29 y.; 4 left-handed by self-report) participated for course credit. All participants had normal or corrected-to-normal vision and were naive as to the purposes of the experiment. As in Experiment 1, participants were assigned to two groups: 24 of them performed the temporal judgment task, the other 24 the lexical decision task.

8.3.1.2. Materials and procedure

Everything was kept as identical as possible to Experiment 1 MouseTracker Software Package (Freeman & Ambady, 2010) was used to collect the stream of x-y coordinates of the mouse trajectories of participants’ manual responses. Thus, precise characterizations of both temporal and spatial dynamics of the mouse trajectories were available for analysis. Individual trajectories were first rescaled to a standard coordinate space, then normalized into 101 time steps (Freeman & Ambady, 2010). Data were exported to Microsoft Office Excel using the utilities included in the MouseTracker package, where they were trimmed (see details below).

Participants sat 60 cm from the computer screen, with their right hand placed over the mouse they found in front of them, centrally positioned with respect to the screen. Two grey squares were always displayed at the top corners of the screen as response alternatives. Each trial began with the appearance of the start button (a grey rectangle at the bottom center of the screen) that remained on the screen until clicked. Then the target stimulus appeared at the center of the screen. As in Experiment 1, stimuli were presented in Courier New font, 38 points, lower case, black printed on white background, remaining on the screen until a response was made by mouse-clicking one of the response alternatives. As in Experiment 1, incorrect trials were followed by a 500 ms red uppercase "X" at the same location of the stimulus. Each incorrect trial was then followed by a 1000 ms blank screen. Correct trials were followed only by a 1500 ms blank screen. Participants were instructed to decide whether the presented verb or nonverb referred to either the past or the future in the temporal judgment task or whether the stimuli were real Spanish verbs or not in the lexical decision task. Also as in Experiment 1, each task comprised two experimental blocks of 148 trials (separated by a two minutes break) in which the same list of stimuli were responded to using the two possible different mappings of responses (past/future or word/nonword) to response boxes (left/right). The order of presentation of the two mappings was counterbalanced over participants.

8.3.1.3. Design and Analysis

The data were analyzed using a mixed factorial ANOVA with the same factors as in Experiment 1: Task (temporal judgment vs. lexical decision) x Lexical status (word vs. nonword) x Temporal reference (past vs. future) x Key (left vs. right) x Counterbalance (to reduce noise, but not reported further). Notice that the factor Key actually refers to the location of the response box in either the left or right upper corners of the screen, but we keep the same name as in Experiment 1 for the sake of easing comparisons.

8.3.2. Results

Accuracy, Initiation Time and Trajectory Time

Errors occurred in 4.73% of trials of the temporal judgement task, and in 1.82% of trials in the lexical decision task. Errors were analyzed independently. Any trial in which Trajectory time, Initiation time or Maximum deviation exceeded 2.5 standard deviations from each participant's mean were excluded from the analysis. This process led to the removal of an additional 3.84% of data in the temporal judgment task and 3.17% in the lexical decision task.

The ANOVA on Initiation times showed a significant interaction between Task and Lexical status ($F(2, 92) = 5.28$, $MSe = 361.86$, $p < .05$): starting a movement of the mouse to a word vs. nonword stimulus took the same amount of time in the temporal judgment task (word $M = 135$ ms vs. nonword $M = 136$ ms), whereas it took longer for words than nonwords in the lexical decision task (word $M = 163$ ms vs. nonword $M = 155$ ms).

The ANOVA on total Trajectory times, which are to be considered as the RTs of the mouse responses, revealed a significant main effect of the factor Lexical status, ($F(1, 46) = 65.34$, $MSe = 12630.59$, $p < .001$), with longer latencies for nonwords ($M = 1416$ ms) than words ($M = 1323$ ms), again in line with the lexicality effect literature. Two interactions were significant as well. First, the Task x Temporal reference ($F(2, 92) = 20.59$, $MSe = 6751.90$, $p < .001$), due to past tense items being faster than future tense items in lexical decision (past M

= 1298 ms vs. future *M* = 1329 ms; Newman-Keuls *p* <.05), while the reverse was true in temporal judgment (past *M* = 1470 ms, future *M* = 1424 ms; Newman-Keuls *p* <.05). Second, the interaction of the factors Task, Lexical status and Temporal reference (*F*(3, 184) = 5.91, *MSe* = 3168.76, *p* <.05; Temporal judgment: Word - past *M* = 1407 ms, future *M* = 1372 ms; Nonword - past *M* = 1531 ms, future *M* = 1477 ms. Lexical decision: Word - past *M* = 1264 ms, future *M* = 1277 ms; Nonword - past *M* = 1331 ms, future *M* = 1381 ms).

The analysis of accuracy revealed two main effects: Task (*F*(1, 46) = 13.07, *MSe* = 7.97, *p* <.001), due to less error in lexical decision than temporal judgment (98.08% vs. 95.27% of correct responses, respectively), as in Experiment 1; and Lexical status (*F*(1, 46) = 15.12, *MSe* = 2.65, *p* <.001), which showed a greater accuracy for words than nonwords (97.55% vs. 95.81% of correct responses, respectively).

Trajectory spatial analysis

The ANOVA on Area Under the Curve (AUC) mean values revealed that, as in Experiment 1, the interaction of Temporal reference and Key was significant (*F*(2, 92) = 7.86, *MSe* = 0.09, *p* <.01): future verbs showed smaller AUC values when responded at the right corner (*M* = 0.71) than at the left corner (*M* = 1.02) (Newman-Keuls *p* <.001), and the opposite was true for past verbs (left *M* = 0.98, right *M* = 0.85; Newman-Keuls *p* <.01). As in Experiment 1, this interaction was modulated by the Task factor (*F*(3, 184) = 12.87, *MSe* = 0.09, *p* <.001; Temporal judgment: Past verbs - left key *M* = 1.16, right key *M* = 1.14; Future verbs - left key *M* = 1.30, right key *M* = 0.88; Newman-Keuls *p* <.001; . Lexical decision: Past verbs - left key *M* = 0.79, right key *M* = 0.56, Newman-Keuls *p* <.001; Future verbs - left key *M* = 0.73, right key *M* = 0.54, Newman-Keuls *p* <.001) (see Fig. 3).

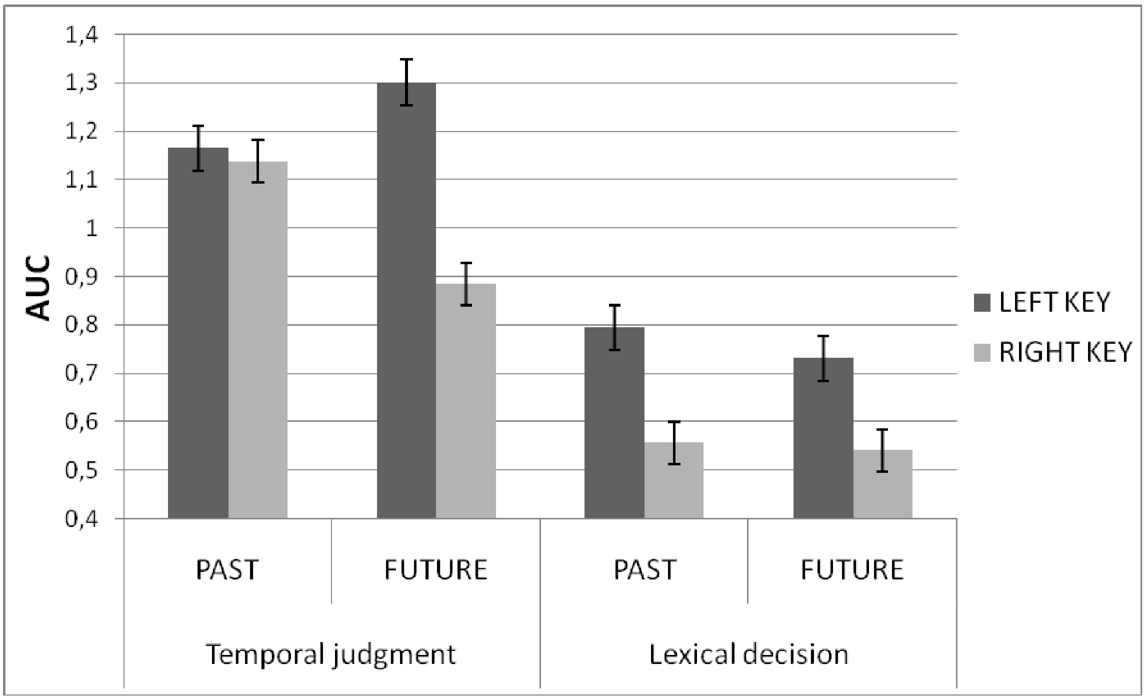


Figure 3. Experiment 2 - AUC mean values for the interaction Task x Temporal reference x Key.

Thus, the congruency effect was found again only in the temporal judgment task. As in Experiment 1, to better clarify this crucial result, we report also the interaction considering the factors Task and Compatibility (*F*(2, 92) = 12.87, *MSe* = 0.04, *p* <.001; Temporal judgment: congruent *M* = 1.02, incongruent *M* = 1.22 - Newman-Keuls *p* <.001. Lexical decision: congruent *M* = 0.67, incongruent *M* = 0.64) (see Fig. 4).

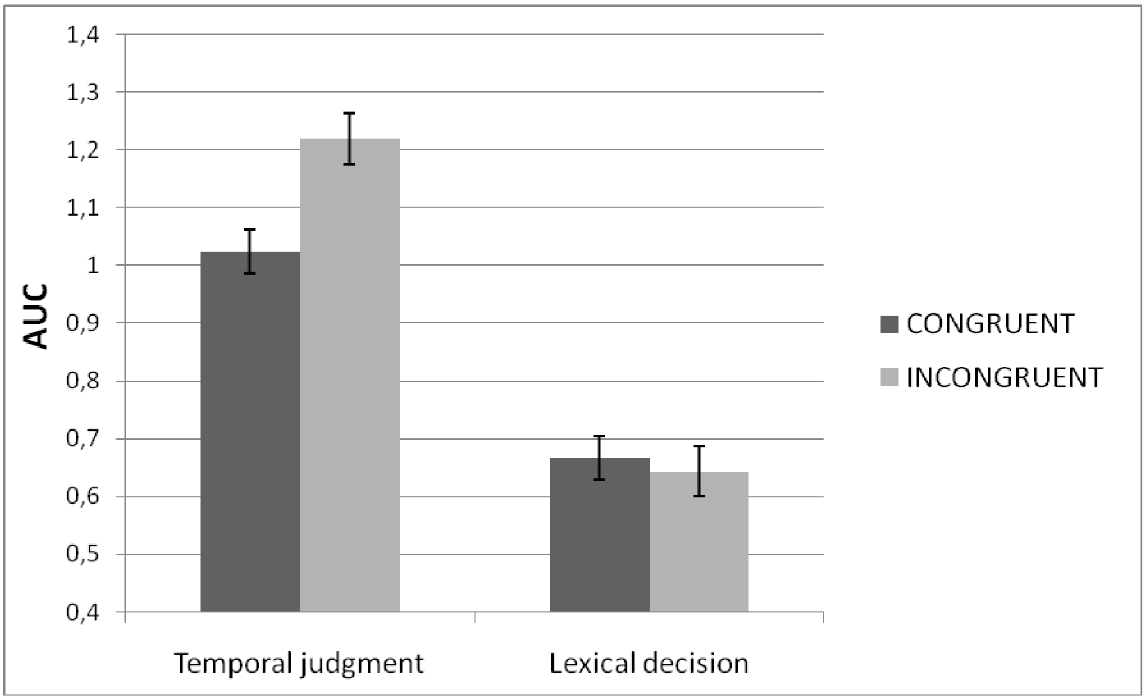


Figure 4. Experiment 2 - AUC mean values for the interaction Task x Compatibility.

As the same interaction was observed also on Trajectory Time, the results of Experiment 2 are perfectly consistent with those in Experiment 1 as well as with the findings of Ulrich & Maienborn (2010). This supports the conclusion that the space-time congruency effect does not arise automatically, but only when the temporal dimension is task relevant. A plot of the mean trajectories of each task are reported in Figure 5 and 6.

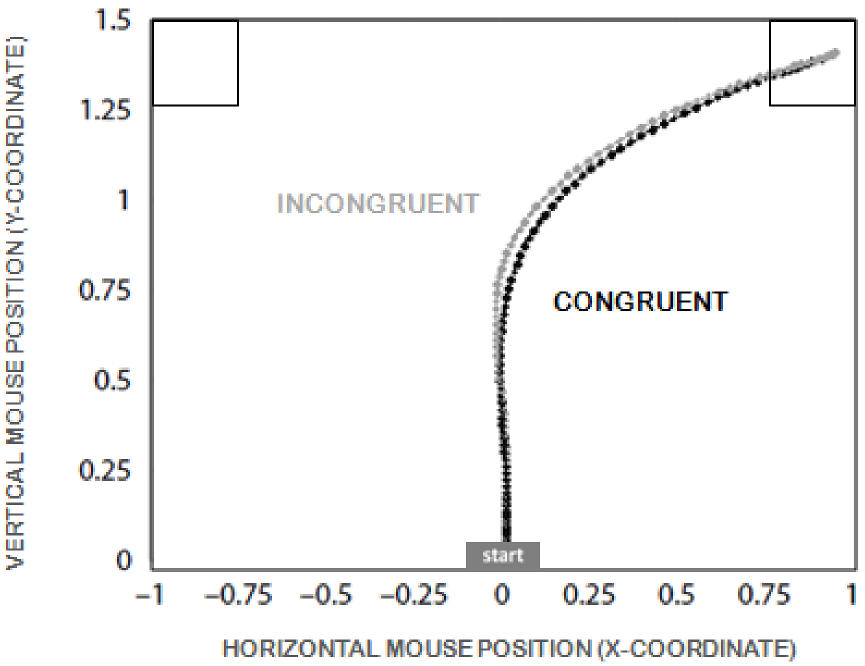


Figure 5. Mean trajectories for Congruent vs. Incongruent trials in Temporal judgment task.

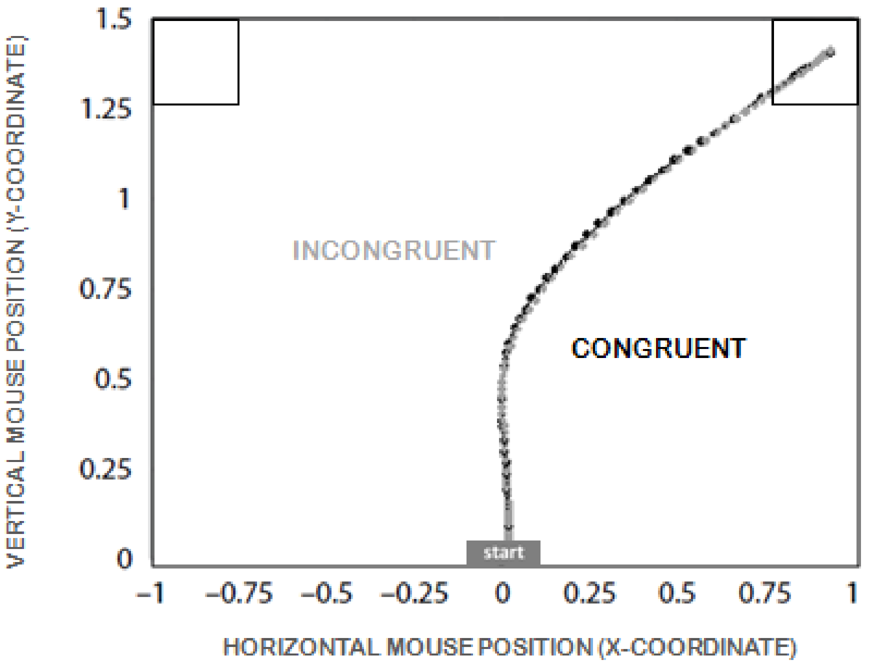


Figure 6. Mean trajectories for Congruent vs. Incongruent trials in Lexical decision task.

There were other significant findings of less relevance to current purposes. First, the factor Task was significant ($F(1, 46) = 11.23$, $MSe = 1.86$, $p < .01$), due to lower AUC mean values in lexical decision (*M* = 0.66) than in temporal judgment (*M* = 1.12). Second, the factor Key (*F*(1, 46) = 10.42, *MSe* = 0.44, *p* <.01) indicated higher trajectory curvatures when participants responded going to the left (*M* = 1.00) than to the right side of the screen (*M* = 0.78). Third, the factor Lexical status approached significance (*F*(1, 46) = 3.06, *MSe* = 0.12, *p* =.087), due to a bigger curvature - greater indecision - when participants responded to nonwords (*M* = 0.92) than to words (*M* = 0.86). Finally, there was an interaction between Task and Lexical status (*F*(2, 92) = 9.86, *MSe* = 0.12, *p* <.01), due to smaller areas in the lexical decision for words (*M* = 0.57) than nonwords (*M* = 0.74; Newman-Keuls *p* <.01), while in temporal judgment no difference was observed (words 1.15 vs. nonwords 1.10).

No other main effects or interactions were significant.

8.3.3. Discussion

Experiment 2 found a significant interaction between left-right responses and temporal reference in the temporal judgment task. Therefore, it confirmed and extended prior findings in the literature (e.g., Torralbo el al., 2006; Santiago el al., 2007; Ulrich & Maienborn, 2010), showing that the space-time congruency effect can also be observed on mouse trajectories. Experiment 2 also confirmed the absence of a congruency effect in the time-irrelevant lexical decision. This null interaction in the time-irrelevant task occurred in the context of a very clear interaction in the time-relevant task, using the same stimuli, procedures and participant population. This replicated the findings of Experiment 1 using a different and probably more sensitive measure. Therefore, , it seems that even when specially designed stimuli are used to make sure that temporal reference is processed, the emergence of a congruency effect between left-right space and temporal content of the stimuli is *strongly mediated by the task context*, specifically by the explicit goal of the task. The present pattern of results strongly suggests that the space-time congruency effect is not automatic.

8.4. General Discussion

The present study was aimed at addressing the issue of automaticity in the activation of the left-right mental time-line. In line with prior findings, we observed flexibility, not automaticity, in the activation of the mental time-line (e.g., Torralbo et al., 2006; Ulrich et al., 2012). The results we observed perfectly reproduced the already reported findings in a situation in which the potential methodological confounds present in the previous studies about the issue of automaticity were avoided (e.g., Ulrich & Maienborn 2010). Indeed, firstly short single words and nonwords especially designed to secure a deep processing generated a strong space-time congruency effect when participants judged their temporal reference, but failed to do it in a lexical decision task. Secondly, this happened not only when response latencies were used for measuring the effect (Experiment 1) as in the previous studies, but also when we recorded the mouse trajectories of the hand movements performed by participants to decide over concurrent response possibility (Experiment 2). Finally, we performed both the tasks following an identical procedure and using the very same pool of stimuli, thus allowing a real comparison of the results of the two tasks. But at the end, our findings on Spanish tensed verbs agree well with the conclusions obtained by Ulrich et al. (2012) regarding the front-back mental time line, and corroborates those by Ulrich and Maienborn (2010) regarding the left-right time line with full sentences in German.

Present results are consistent with the view that, all other factors being equal, only the conceptual mappings that are required to carry out the task are set up in working memory. This is predicted by the coherent working models theory proposed by Santiago and colleagues (2011), which fundamentally rests on the assumption that mental models of any situation are constrained to be the most internally coherent and simple as possible. So, given the requirements of the task, it would be very likely that only the necessary information is retrieved from the long-term memory system and recalled in the mental working space. The authors explicitly indicated this characteristic of mental models as the central factor for accounting the flexibility of conceptual congruency effects which has been observed in the literature in a wide variety of tasks, and not only in linguistic ones (i.e., the flexibility and context-relatedness of affordance-based compatibility effects, see for example Borghi et al., 2012; Kalénine et al., 2013; Natraj et al., 2013; see chapter 4 and 5). Thus, at a theoretical level, such a model seems to be able to explain the fitness of human behaviour: our life would not be so adapted to the environment, if stimuli would be so preponderant and invasive in respect to subjective intentions to determine cognitive processes with their mere presence.

However, it is obvious that present results leave open many future lines of inquiry, and even if they stand in favor of a flexible view of the effect, the issue of the automaticity (or lack thereof) in the activation of the mental time line is still not closed. Indeed, an important open question is whether it could be possible to observe the activation of the mental time line in time-irrelevant tasks when using different conditions. There are several directions that future research might attempt. A first, interesting possibility has to do with the use of temporal stimuli which have a more direct link to temporal reference, such as dates, months or weekdays (e.g., Gevers et al. 2004). A second possibility is that more sensitive and subtle measures (e.g., eye tracking) in respect of the ones already used might be able to find the effects with the same experimental paradigm (i.e., lexical decision). Finally, it is possible that lexical decision requires a too superficial processing of the stimuli: it is clearly showed by response latencies that this is true in respect of time judgments. Thus, future researches may be able to design tasks that would able to activate the spatial dimension even if requiring in-depth cognitive processes different from space-time mapping. Anyway, we do not propend for this possibility, as in the literature lexical decision has demonstrated to be a good setting for observing implicit conceptual effects (e.g., in the study on positive-negative valence items and forward-backward manual movements of Wentura et al., 2000).

On the other side, if the activation of the mental time line needs the participants attention to be directly drawn on temporal aspects of the stimuli and really remains goal-dependent, idea that as said above makes much sense to us, at least this raises the question of why. Indeed, in the literature other conceptual mappings which project onto the spatial dimension have been shown to be activated automatically, at least under certain conditions (e.g., number magnitude, Fischer et al., 2003; evaluation with approach-avoidance responses, Chen & Barg, 1999). Space and time seem to be intrinsically linked from the initial stages in development (Piaget, 1969), whereas the influence of space on temporal judgments in psychophysics tasks remains until the adult age (e.g., Casasanto & Boroditsky, 2008; Casasanto et al., 2010). Future research needs to definitively address the question of why participants do not activate the spatial dimension when processing linguistic stimuli even if they exhibit an evident temporal reference.

In conclusion, the present results corroborate the Flexible foundations view of abstract concepts (e.g., Santiago et al., 2011): the left-right space-time congruency effect was observed only in the case in which participants were explicitly required to judge temporal reference of the stimuli. Thus, this conceptual effect seems to be strongly mediated by the context that was determined by the goals of the experimental task.

9. Evidence of sound-symbolism with every-day objects figures

In par. 3.6, a discussion about the possible evolution of human language highlighted that a basic idea shared by many embodied accounts is the supposed gestural origin of contemporary verbal languages. Following the assumption that speech evolved from gestures it was proposed that, during the gradual phylogenetic shift to verbal language, it is very likely that the vocal system initially reproduced the characteristics of the gestural medium it emerged from. Thus, referring to the literature on actually observable gestural languages, that are contemporary signed languages, it was suggested that the principal property they exhibit is *iconicity*, intended as the resemblance relation between the sign meaning and its gestural expression. This led to the identification of a possible counterpart of this phenomenon in vocal language: if visual iconicity sustained the grounding of meaning in gestural languages, sound-symbolism - intended as the natural, non-arbitrary link between the sound and the meaning of a word – did it for speech. Several scholars have already suggested that sound-symbolism played an important role in the evolution of contemporary vocal languages, underlying the existence in modern lexicon of sound-symbolic words referring to sound-to-meaning correspondences about information from various modality. In line with this view, recent research has started to emphasize the role and importance of iconicity in language, as a possible mean to connect and ground the linguistic-communicative form with the sensory-motor characteristics of words referents. A compelling part of the evidence produced by prior experimental research on sound-symbolic correspondences has focused on sound-shape mappings (i.e. sonorant words associated to rounded shapes, strident words associated to jagged shapes). This literature, however, exhibits some evident methodological limitations. Indeed, the setting was usually a double forced-choice paradigm with simultaneous presentation of both words and shapes pairs. In addition, the totality of the studies used not only invented words, but also ad-hoc figures, created especially for such experiments. This led sound-symbolic skeptics to affirm that the reported effects may be due to the properties of the figures, or to the structure of the task. It is obvious that assuming sound-symbolism as playing a role in natural language evolution implies the necessity of observing it in more ecological experimental settings. In order to demonstrate the reality of sound-symbolism, the present study was guided by the hypothesis that sound-shape correspondences would be observed when participants had to choose, between two invented words, the name which better suited an image representing a common object/entity. Considering that stimuli reproduced every-day entities, a following hypothesis was that this effect would be modulated by the entity category too (e.g., natural objects would be represented with smoother shapes compared to artifacts). The results confirmed the "classic" sound-shape correspondences with this more ecological stimuli, showing the effect both in Experiment 1, when participants chose a name for figures of natural objects (e.g., leaf) and artifacts (e.g., fork), and in Experiment 2, when participants chose a name for figures of natural (e.g., animals) and artificial agents (e.g., robots). Furthermore, a modulation of the category emerged when participants had to name agents: sound-shape correspondences were not observed with robots, which were associated more often to strident words despite of their actual shape. On the whole, these results confirm the validity of sound-symbolism in more ecological situations, bolstering the hypothesis of a natural line of evolution from gestures to iconically grounded forms of speech.

9.1. Introduction

In his *Institutiones* (VI cent. C.E.) Justinian affirmed that *Nomina sunt consequentia rerum* to indicate that verbal language has its origin from the things it denotes. The very beginning of the philosophical debate on the origin of language precedes at least of some centuries this Latin laconism, as in Plato's *Cratylus* (IV cent. B.C.E) there is a talk about the possibility of the existence of a *resemblance relation* between the structure of words and of what they denote. Even if this is not Plato's position, this dialog suggests the emergence of a naturalistic vision of language as opposed to the alternative sophists' view, according to which the word-referent relation is totally arbitrary.

The principle of arbitrariness of human language, however, remains nowadays widely accepted among linguists, philosophers and psychologists (Kovic et al., 2009; Nielsen & Rendall, 2011; Nygaard et al., 2009a, b; for a different position, see Reilly et al., 2012). Since the *Course in General Linguistic* (De Saussure, 1916), contemporary sciences of language have followed the perspective which affirms that a label is always arbitrarily assigned to a referent (e.g. object, event, relation etc.), being each assignation grounded only on socio-cultural conventions (Nielsen & Rendall, 2011; Pietrandrea, 2002). This view considered as accidentally assigned even those names which entertain with their referents (e.g., sounds) a transparent iconic relation. With iconicity we refer to the similarity between certain properties of the words and the sensory-motor characteristics of their referents, as it happens in onomatopoeias, in which words evoke acoustic experiences (e.g. *buzz*, *hiss*; see chapter 3).

In spite of the wide acceptance of the principle of arbitrariness of language, some dissonant voices are starting to emerge. Research on language evolution has suggested that the emergence of lexicon conventionality could be a belated stage in the evolution of human language. Indeed, from a phylogenetic point of view, linguistic conventions might be settled on originally iconic linguistic forms (Merleau-Ponty, 1945; Steels, 2011), and so considered as the outcome of a process which has recursively and efficiently maximized communication (Burling, 1999; Corballis, 2009; Fay et al., 2010; Garrod et al., 2007; Zipf, 1949; for modeling work in this direction see Baronchelli et al., 2011).

A further important source of inspiration for the research aimed at stressing the fact that language is not always arbitrary, in particular when we consider face-to-face interactions, comes from the embodied and grounded approach to cognition (reviews in Barsalou, 2008; Borghi & Pecher, 2011). This approach solves the so-called symbol-grounding problem (Harnad, 1990) by proposing that symbols are grounded in the same systems used by perception, action and emotion (see chapter 1 and 3). According to this perspective, during the processing of words we would re-activate previous experiences with their referent. For example, the word "cat" would re-enact the experience of seeing a cat, caressing it, feeding it, and so on. In line with this view, recent research has started to emphasize the role and importance of iconicity in language, as a possible mean to connect and ground the linguistic-communicative form with the sensory-motor characteristics of words referents. This research has underlined the fact that arbitrariness might not be the rule during face-to-face communication, where gestures and iconic words might be consistently used. For example, Perniss & Vigliocco (in press) have highlighted the importance of iconicity in both spoken and signed language. In their perspective, the role of iconicity is crucial in three important aspects of language processing: phylogenesis, ontogenesis and processing. In phylogenesis iconicity facilitates displacement, i.e. the ability to refer about things distant in time and space, and in ontogenesis it provides a mechanism for establishing referentiality, linking linguistic form and meaning and hence facilitating word learning. Finally, in language processing iconicity facilitates grounding of words in sensory-motor and emotional systems, thereby determining embodiment.

In this paper we investigate the word-referent relationship focusing on the direct bindings between the word sound and certain aspects of the referent appearance (e.g., shape). With word sound we refer to a multimodal experience, including both the acoustic experience during language comprehension and the phono-articulatory experience during word production. The idea underlying this investigation is that words can entertain a non-arbitrary relation with their referents. This process has been identified in speech as sound-symbolism or phonosemantics (Hinton et al., 1994; Parise & Pavani, 2011), and it is something that works in a way similar to iconicity in sign languages (e.g. Corballis, 2002, 2009; Gentilucci et al., 2004; Gentilucci & Corballis, 2006; Pietrandrea, 2002; Perniss et al., 2010; Pizzuto & Volterra, 2000; Thompson et al., 2009, 2010).

The psychological literature on sound-symbolism is longstanding, dating at least the first half of the last century. Indeed, Edward Sapir already suggested in his most known essay on language that verbal labels are able to catch aspects of what they refer to. As evidence of this, Sapir (1929) reported results of an experiment where almost the totality of the participants - English mother tongues - assigned the invented names *mil* or *mal* to a small or a big table respectively, thus intuitively coupling the sound of the words with the size of the objects. Similarly, Wolfgang Köhler (1929; 1947) revealed the existence of correspondences between word sounds and visual shapes: Spanish participants intuitively assigned an invented word with rounded vowels (*baluma*, or by 1947 *maluma*) to rounded invented figures and an invented word with unrounded vowels (*takete*) to jagged invented figures. Finally, the cross-linguistic ability to guess foreign words meaning, registered with samples of different mother tongues (e.g. Brown et al., 1955; Gebels, 1969; Hinton et al., 1994; Koriat & Levy, 1979; Kunihira, 1971), drove some authors to explicitly affirm that speech may have emerged from universal imitative connections between sounds and meanings (Kovic et al., 2009).

In the last ten years, research conducted on speakers of different languages has gathered results which support the idea of sound-shape correspondences (e.g. Akita et al., 2008; Arata et al., 2010; Asano et al., 2011; Kovic et al., 2009; Iwasaki et al., 2007; Nielsen & Rendall, 2011; Nygaard et al., 2009a, b; Parault, 2006; Ramachandran & Hubbard, 2001; Spector & Maurer, 2008; Westbury, 2005). In particular, Maurer et al. (2006) investigated the takete-maluma phenomenon in 2.5-years-old children and adults to assess whether children reliably map words with rounded vowels to rounded shapes, and words with unrounded vowels to jagged shapes (for a different interpretation, see Nielsen & Rendall, 2011). Participants were shown simultaneously two shapes, a rounded and a jagged one, and had to name each shape choosing the name within two alternatives, one with rounded vowels and the other with unrounded ones. Results showed that children, as adults, matched names with rounded vowels to rounded shapes and names with unrounded vowels to the jagged ones, indicating that sound-shape correspondences are at work at the earliest stages of language acquisition.

As Maurer et al. (2006), several studies on sound-shape correspondences have adopted naming tasks with a two alternative forced-choice design. Although results of studies adopting such a kind of paradigm support the hypothesis of a non-arbitrary relation between words and their referents, there are some methodological issues which might bring a sound-symbolism skeptic to say that results can be due to some sort of confounds. For example, as far as our knowledge is concerned, the most part of experiments adopting the naming paradigm has proposed forced-choice tasks where two words, one sonorant and one strident, were simultaneously presented together with the stimuli figures pairs. The risk with such a design is twofold, as it was clearly highlighted by Nielsen and Rendall (2011). First, the purpose of the experiment can become very transparent to subjects, which are required to compare both figures and names in each trial. Second, the simultaneous presentation doesn't allow disentangling when the results are due to two matches, one for strident sounds/jagged shapes and another for sonorant sounds/rounded shapes, and when there is only one match in one of the two directions. Indeed, in each trial the subjects second-choice is automatically defined by the image and the name they have coupled first. Another problem is that these experiments might be poorly ecological and might not reflect what happens in real life, since both the words and the stimulus figures are created ad-hoc for the experiment. Moreover, these ad-hoc figures emphasize the features under investigation (e.g., roundness, jaggedness), thus possibly inducing an enhancement of the reported behavioral effects (see Nielsen & Rendall, 2011 for a similar critique).

For these reasons, in the present study we tested the correspondences between sounds and shapes avoiding a double forced-choice paradigm, and using visual stimuli representing every-day objects and entities. Participants were required to choose between two words the suitable name for an image which represented a well known object/entity, common in every-day life.

Using every-day stimuli allowed us to overcome one further limitation of current studies on sound-symbolic correspondences. To our knowledge, no study so far has taken into account the possible effects of categorical differences on sound-symbolic correspondences. In contrast, the literature on concepts and categorization has highlighted profound differences in the representation of artifacts and natural objects and of living and not living entities, as indicated by neural studies on brain activation (for a review, see Martin, 2007), by neuropsychological studies on categorical deficits (e.g. Humphreys & Forde, 2001; Gainotti, 2000) and by behavioral studies on categorization in children (e.g. Mandler, 1992; Rakison & Oakes, 2003) and adults (e.g. Borghi et al., 2007; Laurence & Margolis, 2007; Roversi et al., 2013). Some authors have underlined that categorical distinctions in infants might be based on perceptual cues as well as on motion cues (e.g. Mandler, 1992, 2004). Motion cues differ for animals and artifacts: animals are characterized by self-propelled movements and by nonlinear and smooth motion paths, while artifacts are characterized by induced motion and by linear motion path (Mandler, 1992; Rakison & Paulin-Dubois, 2001, 2002). In a similar vein, recent research has shown that some features of words, such as their grammatical gender, are not arbitrarily related to the characteristics of the referents belonging to different categories. Sera et al. (2002) asked English, Spanish, French and German children and adults speakers to attribute male or female voices to inanimate objects. Result showed that for gendered languages the relation between grammatical gender and the perception of items as being male- or female-like was not arbitrary: speakers of Spanish and French, unlike speakers of German, relied on grammatical gender in their assignment of male and female voices to inanimate objects. The different results obtained with French and Spanish compared to German speakers suggest that the effect is present with grammatical systems characterized by two gender categories (German has 3), where grammatical and natural gender are highly correlated. More crucially for us, the gender assignment interacted with the category of the objects, showing that artificial objects were more often perceived as male-like than natural ones, which were instead considered more female-like. Thus, the authors pointed out that the natural–artificial distinction may be correlated to several factors, such as item shape (rounded/jagged), density (light/heavy), or common use (typically used by females/males), which map onto the grammatical gender assignments in a phonesemantic-like manner.

On the basis of the reviewed evidence, we reasoned that manipulating the category would allow us to get some hints on how categories are represented with a paradigm never used in this context. As suggested by Sera et al. (2002) it is possible, indeed, that one further feature distinguishing natural objects and artifacts pertains their shape: the shape of natural objects could be mentally represented as smoother compared to artifacts' shape. In light of these considerations, we designed a paradigm that allowed us to study the development of the sound-symbolic correspondence with every-day objects/entities, which belonged to different categories. We will expose below our hypotheses, based on the main manipulations we decided to introduce.

First, we hypothesize that the sound-shape correspondence effect would be conserved if the figures represent every-day entities and, second, if they are one-by-one presented. If the hypothesis is confirmed, this will show that the effect is due neither to the structure of the task nor to the properties of the shown figures.

Third, we hypothesize that the sound-shape effect is modulated by objects/entities category. To this aim in Experiment 1 the stimuli figures represented every-day objects, which could be natural objects or artifacts, and in Experiment 2 the stimuli were figures representing natural (i.e., animals) or artificial agents (i.e., robots), to explore whether natural entities are represented with smoother shapes as compared to artificial entities.

9.2. Experiment 1

9.2.1. Method

9.2.1.1. Participants

Twenty-four undergraduate students from the University of Bologna participated in the experiment for course credits (9 males; mean age = 20.79 (2.23); 2 left-handed by self-report). All participants had normal or corrected-to-normal vision and were naive as to the purposes of the experiment.

9.2.1.2. Materials

The task consisted in choosing one word within a pair as the preferred name for the picture stimulus. Words pair was simultaneously displayed on the screen. Stimuli consisted of 24 black-and-white line figures chosen from the graphic database by Lotto, Dell'Acqua and Job (2001). Twelve figures referred to natural objects and twelve figures referred to artifacts, and each set was composed by 6 rounded-shaped and 6 angular-shaped figures. Each participant rated each picture after the experimental session on a 7-point Likert scale for sharpness/roundness (1: “very sharp” – 7: “very rounded”). Ratings were analyzed in a 2 x 2 ANOVA with the within factors Figure Type (Natural vs. Artificial) and Figure Shape (Rounded vs. Jagged). The ANOVA with items as random factor revealed a main effect of Figure Type, $F(1, 5) = 24.98$, $MSe = 0.17$, $p < .01$ (Natural $M = 4.21$, Artificial $M = 3.37$), and of Figure Shape, $F(1, 5) = 134.23$, $MSe = 0.41$, $p < .001$ (Rounded $M = 5.31$, Jagged $M = 2.26$). The interaction between Figure Type and Figure Shape was not significant.

The ANOVA with subjects as random factor showed the main effects of Figure Type, $F(1, 23) = 68.11$, $MSe = 0.25$, $p < .001$ (Natural $M = 4.21$, Artificial $M = 3.37$), and of Figure Shape, $F(1, 23) = 718.69$, $MSe = 0.31$, $p < .001$ (Rounded $M = 5.31$, Jagged $M = 2.26$). The interaction between Figure Type and Figure Shape was significant as well, $F(2, 46) = 11.26$, $MSe = 0.24$, $p < .01$ (Natural: Rounded shape $M = 5.89$, Jagged shape $M = 2.51$,; Artificial: Rounded shape $M = 4.72$, Jagged shape $M = 2.01$) (LSD, all $ps < .001$).

To control if participating in the experiment exerted an influence on participants’ attitudes about object shapes, an additional independent group was asked to make an identical evaluation of the pictures. The independent group was composed by twenty-four students from the University of Bologna, participating for course credits (7 males; mean age = 23.42 (3.51); 5 left-handed by self-report). They had normal or corrected-to-normal vision and were naive as to the purposes of the questionnaire. The independent ratings were again analyzed with the 2 x 2 ANOVA, with the within factors Figure Type (Natural vs. Artificial) and Figure Shape (Rounded vs. Jagged). The ANOVA by-item revealed a main effect of Figure Type, $F(1, 5) = 11.26$, $MSe = 0.30$, $p < .05$ (Natural $M = 4.37$, Artificial $M = 3.62$), and of Figure Shape, $F(1, 5) = 258.17$, $MSe= 0.27$, $p < .001$ (Rounded $M = 5.69$, Jagged $M = 2.30$). The interaction between Figure Type and Figure Shape was not significant.

The ANOVA by-participants showed the main effects of Figure Type, $F(1, 23) = 68.54$, $MSe = 0.20$, $p < .001$ (Natural $M = 4.37$, Artificial $M = 3.62$), and Figure Shape, $F(1,$

23) = 718.69, *MSe* = 0.31, $p < .001$ Rounded $M = 5.69$, Jagged $M = 2.30$). The interaction between Figure Type and Figure Shape was significant as well, $F(2, 46) = 11.37$, *MSe* = 0.14, $p < .01$ (Natural: Rounded shape $M = 6.19$, Jagged shape $M = 2.55$; Artificial: Rounded shape $M = 5.18$, Jagged shape $M = 2.06$) (LSD, all *ps* < .001). Thus, the ratings of the independent group showed a pattern identical to the ratings of the experimental group. Indeed, ratings predict both the sound-symbolic correspondence of names and shapes, and an effect of the category, which should emerge in the experimental results in a form similar to the standard sound-symbolic effect.

The 8 words, used as names for the 24 pictures, were taken from the study by Maurer et al. (2006) and manipulated in the way they were written to obtain in Italian the same sound they have in English (e.g., the English *bouba* was transformed in the Italian *boba*, see the Appendix for the complete list of stimuli). The 8 names were coupled in the same four pairs of the experiment by Maurer et al. (2006), always composed by one sonorant, round-sounding word (e.g. *maluma*) and one strident, sharp-sounding word (e.g. *takete*). Each pair of words was visually presented on a computer screen under the picture to name (see Figure 1). Thus, depending on the object appearance in the figure (Figure Shape: Rounded vs. Jagged) and on the phonological characteristics of the name (Response Type: Rounded vs. Jagged), in each trial it was possible to obtain as response a sound-symbolic combination (e.g. choosing *maluma* as the name of a rounded-shaped figure) or a non-sound-symbolic combination (e.g. choosing *maluma* as the name of a jagged-shaped figure). For sake of simplicity we defined the two levels of both Figure Shape and Response Type factors as Rounded vs. Jagged.

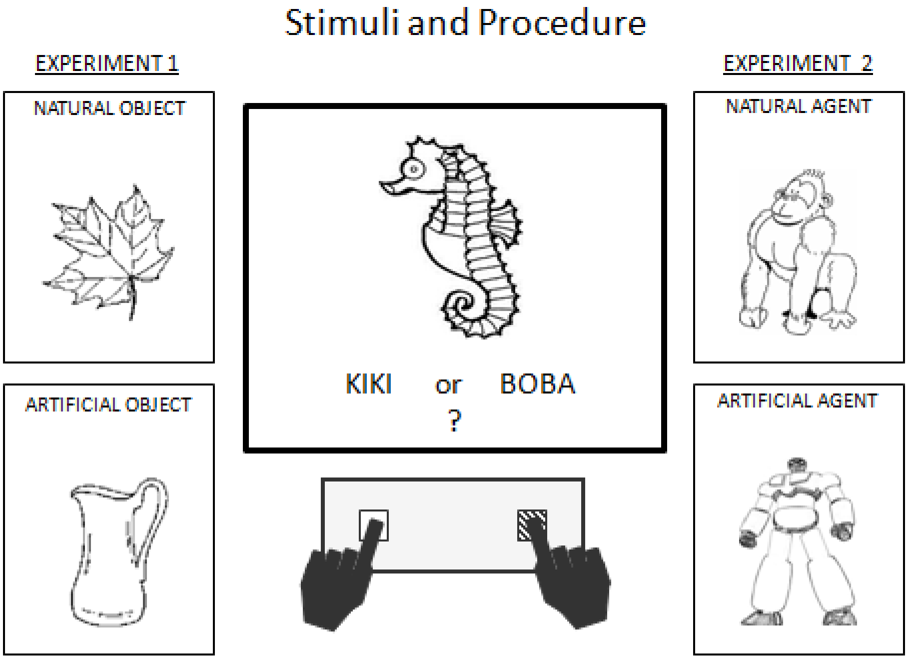


Figure 1. Experimental stimuli and procedure

9.2.1.3. Design and procedure

Participants sat 50 cm from the computer screen. Each trial began with a fixation point (+) lasting for 500 ms. Then the stimulus picture was displayed centrally, remaining on the screen for 5 seconds or until a response was made. The two names were simultaneously presented under the picture, one on the left and the other on the right (the names order was counterbalanced between subjects). Participants were required to decide which of the two names was more suitable for the picture displayed above them by pressing two keys on an Italian QWERTY keyboard. The keyboard was positioned close to the screen so that each of the two keys was located perfectly below the name to which it corresponded: participants pressed the 5 key to choose the name on the left and the 9 key for the name on the right (see Figure 1). At the beginning of the experiment they were instructed to respond as quickly as they could. They did not receive any feedback about the accuracy of their responses, as they were told no correct/incorrect values were settled for the trials. Considering that each of the total 24 pictures was presented once with each of the 4 word pairs, the experiment consisted overall in 96 experimental trials, preceded by 8 training trials to familiarize with the procedure.

9.2.2. Results

Missing responses (i.e. responses that required more than 5 seconds to be given) were removed. Their very low rate (0.17%) testified the task was easy to perform. All the remaining responses were transformed in percentage of choosing a rounded response (the percentage of rounded and jagged responses sum up to 100%, thus they cannot be considered as independent from each other) and entered in a 2 x 2 ANOVA with the within factors Figure Type (Natural vs. Artificial) and Figure Shape (Rounded vs. Jagged). Fisher's LSD *post hoc* tests were conducted on significant interactions. The ANOVA on the percentage of rounded responses showed only the expected main effect of Figure Shape, $F(1, 23) = 20.43$, $MSe = 341.23$, $p <. 001$, due to rounded sounding names being more frequently assigned as label to Rounded shapes ($M = 61.86\%$) than to Jagged shapes ($M = 44.82\%$), as reported in Figure 2.

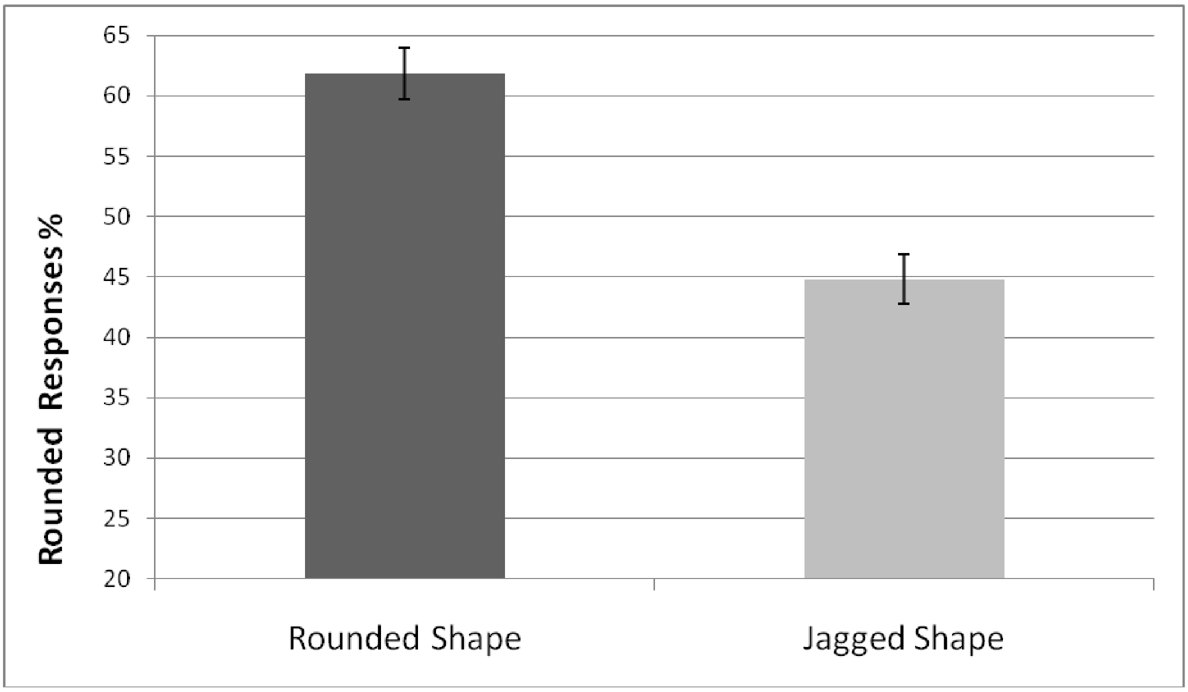


Figure 2. Experiment 1 - Main effect of Figure Shape

In this task, each picture was repeated 4 times. To control if picture repetition could have affected the results, we repeated the analysis on the data collected in Experiment 1 considering only the trials of first presentation of each picture. The percentage of rounded responses to the first presentation were entered in a 2 x 2 ANOVA with the within factors Figure Type (Natural vs. Artificial) and Figure Shape (Rounded vs. Jagged). Fisher's LSD *post hoc* tests were conducted on significant interactions.

The ANOVA on the percentage of rounded responses showed a marginally significant main effect of Figure Shape, $F(1, 23) = 3.95$, $MSe = 525.59$, $p <. 06$, due to rounded sounding names being more frequently associated to Rounded shapes ($M = 55.90\%$) than to Jagged shapes ($M = 46.60\%$).

Finally, to control if the experimental results were related to the pictures ratings, we correlated subjects' mean rating to each picture and the mean percentage of rounded response then assigned to that picture. The correlation was positive and highly significant ($r = .24$, $p < .001$), indicating that the more the picture was subjectively perceived as round, the more the participant tended to label it with a rounded word.

Importantly, we confirmed the results of the ANOVA and of correlations by mixed effects models, through which we simultaneously took into account the fine-grained effect of perceiving roundness (i.e., ratings values) and the variance due to random factors such as Item, or Names Pair. We defined the first model as the most complete one, i.e., with Figure Type and Ratings as fixed effects, and Item (i.e., each picture), Names Pair (i.e., the two names within to choose) and Participant in interaction with Figure Type and Rating as random effects. Then, at each consecutive step we tried another model by firstly eliminating random effects one by one, then interaction between fixed effects, and finally fixed effects one by one. Each time we performed a log-likelihood test to determine if one model was worse than the previous in fitting the data: if no significant difference was observed, or if the last model was significantly better than the previous one (i.e., log-likelihood value closer to 0) then the less complete model was chosen. The final model resulting from this procedure consisted of all random effects, and only Rating as fixed effect (AIC = 2888, BIC = 2945, Log-likelihood =-1434).

9.2.3. Discussion

As predicted, participants more frequently chose rounded words (e.g. maluma) as names for rounded shaped objects figures (e.g. orange) and jagged words (e.g. takete) as names for jagged shaped objects figures (e.g. fork), showing a high sensitivity to the correspondence between words sounds and visual shapes even if the figures to name represented familiar objects. This was true also when we considered only the first presentation of the visual stimuli. Importantly, the results of the correlational analysis between ratings and performance and of mixed effects models confirmed this finding, and indicated that sound-shape correspondence is a continuous, fine-grained effect. This result confirms evidence on the sound-shape correspondence effect. Furthermore, it suggests that attaching labels to external entities, and specifically to every-day objects which already have a common and conventionalized label, is not necessarily an arbitrary activity. Finally, we were able to obtain the sound-shape correspondence effect by presenting the stimuli one-by-one, thus avoiding potential limits of previous studies, such as the transparency of the experimental aim and the enhancement of the observed effects (see Nielsen &Rendall, 2011).

However, despite the fact that results from all the ratings performed on Experiment 1's pictures showed that artifacts were perceived as sharper than natural objects, the predicted effect of the object category (natural vs. artificial) was not found in the choice of the name.

9.3. Experiment 2

In Experiment 1 we asked participants to choose a name for pictures of already known objects. We found the predicted sound-symbolic correspondence between names and shapes, but no effect of the object category. One possible cause of the absence of a category effect is that very different items were compared – for example, the category of artifacts included both very simple tools (e.g. spoon) and more complex ones (e.g. compass). In Experiment 2 we investigated whether the effect would be found using a more compact sub-category within the artificial and natural entities, i.e. the category of *agents*. We define an *agent* as an entity perceivable as having the ability to autonomously *act* or *move*, and endowed with features typically linked to animacy (Landau et al., 1988; Backscheider et al., 1999), i.e., eyes. In this sense we consider as agents both animals (i.e. natural agents) and robots (i.e. artificial agents). In contemporary cultures robots have become a quite credible kind of agent, in part for common sense beliefs about Turing machines and Artificial Intelligence, and mostly for the part they play in popular culture (e.g. science fiction books, comics, movies).

Stimuli of Experiment 2 were animals and anthropomorphous robots figures. This choice allowed us to compare two categories the members of which are more similar than the members of the previously used artifacts and natural objects categories. Lastly, considering that an ontogenetic continuity of sound-symbolism has been already shown in literature with ad-hoc stimuli (e.g. Maurer et al., 2006), here we tested a sample composed by adults and children, to investigate the sound-symbolic phenomenon related to every-day categories in function of age. The reasons why we decided to test participants of different ages are multifold. First, typically studies on sound-symbolism are realized with children too, and considering that Experiment 1 confirms that sound-symbolic correspondences are able to affect the labeling of every-day objects, we think it is important to verify whether different processes characterize sound-symbolism with common objects in adults and children. Second, and more crucially, we are interested in whether the relationship between shapes and sounds emerges earlier in development than the relationship between categories and sounds. We start namely from the assumption that language influences categorization (e.g. Lupyan, 2012), and that a prolonged use of a given label may render the categories less malleable and more stable than they initially are. On this basis, it is very likely that children are more flexible than adults in forming categories as those of natural and artificial agents, as the features that result to be more salient depend from developmental processes where language plays a fundamental role (Sera et al., 2002). While the formation of categories of natural and artificial agents should be profoundly influenced by language, this should be less the case for shape categories. The interest for the different developmental pattern of the relation between sound and shape compared to that between sound and category derives from these considerations.

Thus, we predicted that, using the more specific subcategory of agents, we would find a modulation of the sound-symbolic effect in function of both category and age. In particular, we expected a more marked effect of category on the choice of name in adults, as they may have a more clear distinction between natural and artificial agents due to experience and prolonged use of verbal labels (Sera et al., 2002), as well as because in children the category of animated entities might be broader and the representations of natural and artificial agents might overlap. In addition, if language influences categorization (e.g., Lupyan, 2012) we expect that children's categories are more flexible and malleable compared to the categories of adults (see also Sera et al., 2002, for a similar conclusion pertaining gender). Furthermore, we expect that adults' categories will be less permeable to learning effects occurring during the experiment (e.g., effects due to figure repetitions).

9.3.1. Method

9.3.1.1. Participants

Twenty-four children (15 males; mean age = 8.79 (1.06); all right-handed) participated to the experiment as volunteers, and twenty-four students from the University of Bologna (10 males; mean age = 21.04 (2.91); 3 left-handed by self-report) participated for course credits. All participants had normal or corrected-to-normal vision and were naive as to the purposes of the experiment.

9.3.1.2. Materials

The materials consisted of twenty-four black-and-white pictures of manmade drawings, of which 12 represented animals (6 rounded and 6 jagged-shaped) and 12 robots (6 rounded and 6 jagged-shaped), and of the same eight words used in Experiment 1 (see the Appendix for the complete list of stimuli).

As in Experiment 1, after the experimental session, each subject rated pictures on a 7-point Likert scale for roundness/sharpness. A mixed 2 x 2 x 2 ANOVA with the between factor Group (Children vs. Adults), and the within factors Figure Type (Animal vs. Robot) and Figure Shape (Rounded vs. Jagged) was performed. The ANOVA by-item revealed a main effect of Figure Type, $F(1, 10) = 31.14$, $MSe = 0.40$, $p < .01$ (Animal $M = 4.23$, Robot $M = 3.21$), and of Figure Shape, $F(1, 10) = 331.59$, $MSe = 0.24$, $p < .001$ (Rounded $M = 5.01$, Jagged $M = 2.42$). The main effect of Group almost reached significance, $F(1, 10) = 3.72$, $MSe = 0.25$, $p < .09$ (Children $M = 3.86$, Adults $M = 3.58$), as well as the interaction of Group and Figure Type, $F(2, 20) = 3.61$, $MSe = 0.40$, $p < .09$ (Children: Animal $M = 4.19$, Robot $M = 3.52$; Adults: Animal $M = 4.26$, Robot $M = 2.90$). No other interactions reached significance.

A by-participants ANOVA with the same factors was performed as well. All the main effects were significant: the factor Group, $F(1, 10) = 5.46$, $MSe = 0.67$, $p < .05$ (Children $M =$

3.86, Adults *M* = 3.58), the factor Figure Type, *F*(1, 10) = 98.33, *MSe* = 0.50, *p* < .001 (Animal *M* = 4.23, Robot *M* = 3.21), and the factor Figure Shape, *F*(1, 10) = 316.51, *MSe* = 1.02, *p* < .001 (Rounded *M* = 5.01, Jagged *M* = 2.42). An interaction was reliable too, the Group x Figure Type, *F*(2, 20) = 11.39, *MSe* = 0.50, *p* < .01 (Children: Animal *M* = 4.19, Robot *M* = 3.52 - LSD *p* < .05; Adults: Animal *M* = 4.26, Robot *M* = 2.90 - LSD *p* < .001). No other interaction reached significance.

As in Experiment 1, we wanted to control the eventual influence of the repeated exposure to the labeling task on the attitudes about visual stimuli shapes, so an additional independent group performed the ratings on the pictures of Experiment 2. The independent group was composed of twenty-four children (11 males; mean age = 9.13 (0.45); 2 left-handed by self-report) who participated as volunteers, and twenty-four students from the University of Bologna (12 males; mean age = 23.54 (3.37); 3 left-handed by self-report) who participated for course credits. All participants had normal or corrected-to-normal vision and were naive as to the purposes of the questionnaire. The independent ratings were analyzed using a 2 x 2 x 2 ANOVA with the between factor Group (Children vs. Adults) and the within factors Figure Type (Animal vs. Robot), Figure Shape (Rounded vs. Jagged). The ANOVA by-item revealed the main effect of Figure Type, *F*(1, 10) = 7.43, *MSe* = 0.99, *p* < .05 (Animal *M* = 4.08, Robot *M* = 3.30), and Figure Shape, *F*(1, 10) = 404.36, *MSe* = 0.20, *p* < .001 (Rounded *M* = 4.99, Jagged *M* = 2.39). No other main effect or interactions reached significance.

An ANOVA by-subject with the same factors were performed as well, showing as significant the three main effects: the factor Group, *F*(1, 46) = 6.74, *MSe* = 1.06, *p* < .05 (Children *M* = 3.88, Adults *M* = 3.49), the factor Figure Type, *F*(1, 46) = 67.99, *MSe* = 0.44, *p* < .001 (Animal *M* = 4.08, Robot *M* = 3.30), and the factor Figure Shape, *F*(1, 46) = 589.67, *MSe* = 0.55, *p* < .001 (Rounded *M* = 4.99, Jagged *M* = 2.39). Three interactions almost reached significance, starting from the Group x Figure Shape, *F*(2, 92) = 3.42, *MSe*= 0.55, *p* < .08 (Children: Rounded shape *M* = 5.08, Jagged shape *M* = 2.68; Adults: Rounded shape *M* = 4.89, Jagged shape *M* = 2.10). Then, the Figure Type x Figure Shape, *F*(2, 92) = 3.45, *MSe*= 0.37, *p* < .07 (Animal: Rounded shape *M* = 5.46, Jagged shape *M* = 2.71; Robot: Rounded shape *M* = 4.51, Jagged shape *M* = 2.08). Finally, the three-way interaction of the factors Group, Figure Type and Figure Shape was almost significant, *F*(3, 184) = 4.01, *MSe*= 0.37, *p* < .06 (Children - Animal: Rounded shape *M* = 5.47, Jagged shape *M* = 3.08; Robot: Rounded shape *M* = 4.69, Jagged shape *M* = 2.29. Adults - Animal: Rounded shape *M* = 5.45, Jagged shape *M* = 2.33; Robot: Rounded shape *M* = 4.33, Jagged shape *M* = 1.88). No other effects were observed.

Overall, the independent group ratings partially confirmed the pattern of the experimental group, as both ratings predict the "classic" sound-symbolic correspondence of sounds and shapes, and also an effect of the category (by itself, or in interaction with the stimuli shape) for both experimental groups.

9.3.1.3. Design and procedure

The design and the procedure were exactly the same of Experiment 1, except for the fact that the stimuli used, instead of pictures of natural objects and artifacts, were pictures of animals and robots.

9.3.2. Results

Missing responses were removed (1.28%), and the remaining responses were entered as percentages in a mixed 2 x 2 x 2 ANOVA with the between factor Group (Children vs. Adults) and the within factors Figure Type (Animal vs. Robot), Figure Shape (Rounded vs. Jagged). Fisher's LSD *post hoc* tests were conducted on significant interactions.

The analysis showed, as expected, the significant main effect of the factor Figure Shape, *F*(1, 46) = 11.54, *MSe* = 132.49, *p* < .01, due to Rounded shapes (*M* = 52.45%) being more often labeled with a rounded sounding name than Jagged shapes (*M* = 46.81%) (see Fig. 3).

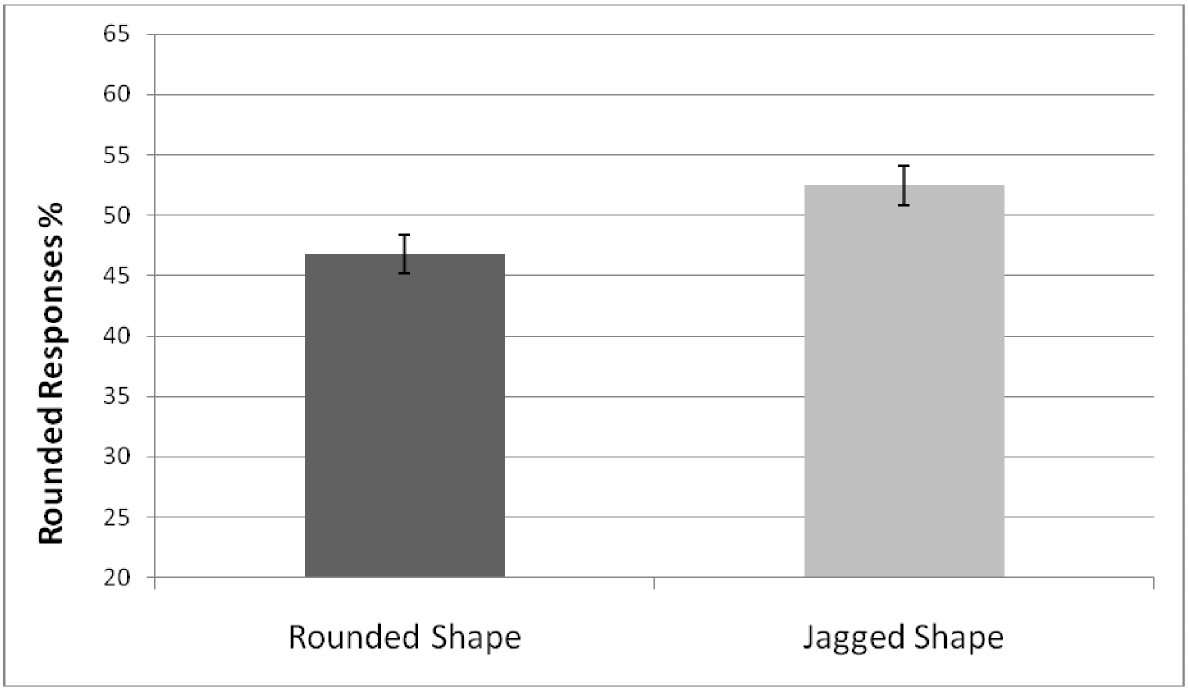


Figure 3. Experiment 2 - Main effect of Figure Shape

The main effect of Figure Type, *F*(1, 46) = 3.72, *MSe* = 1079.55, *p* < .06, almost reached significance, due to Animals (*M* = 54.21) being more often labeled with rounded names than Robots (*M* = 45.06), as reported in Fig. 4.

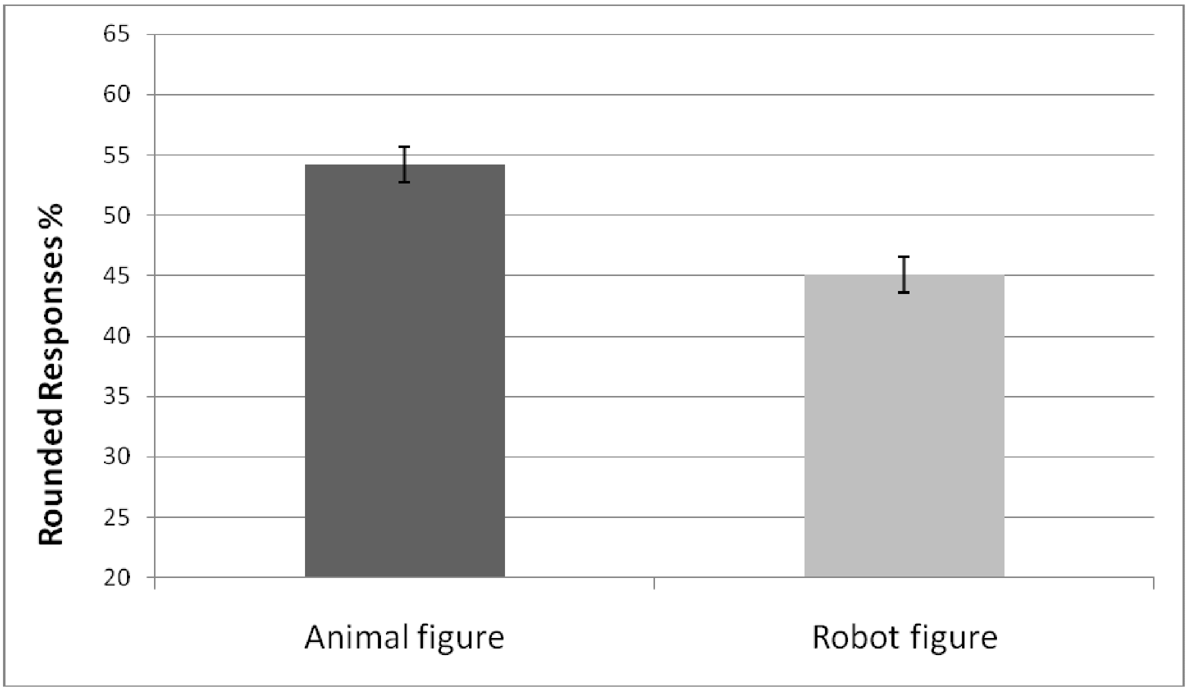


Figure 4. Experiment 2 - Main effect of Figure Type

The interaction Group x Figure Type was significant, $F(2, 92) = 5.85$, $MSe = 1079.55$, $p < .05$. If an undifferentiated pattern in respect of Figure Type was observed in Children responses (Animal $M = 47.92\%$, Robot $M = 50.24\%$), the category of the stimuli exerted a clear effect on the responses of Adults (Animal $M = 60.49\%$, Robot $M = 39.88\%$; LSD $p < .01$) (see Fig. 5).

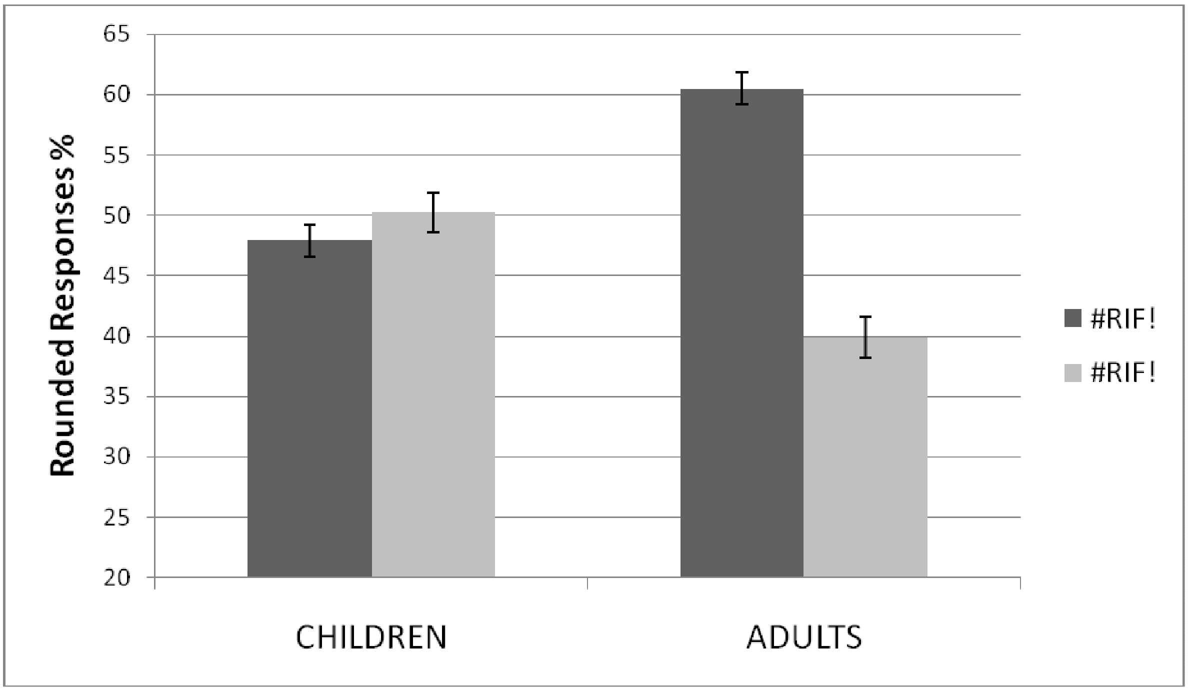


Figure 5. Experiment 2 - Interaction Group x Figure Type

The interaction Figure Type x Figure Shape was significant as well, *F*(2, 92) = 23.70, *MSe* = 90.55, *p* < .001, showing that Animal Rounded shape (*M* = 60.37%) received significantly more rounded responses than all other conditions (Animal Jagged shape *M* = 48.04%; Robot Rounded shape *M* = 44.54%, Robot Jagged shape *M* = 45.58%) (LSD, all *ps* < .001), with the category effect suppressing the sound-shape correspondence for natural jagged shape and especially for artificial rounded shape (see Fig. 6).

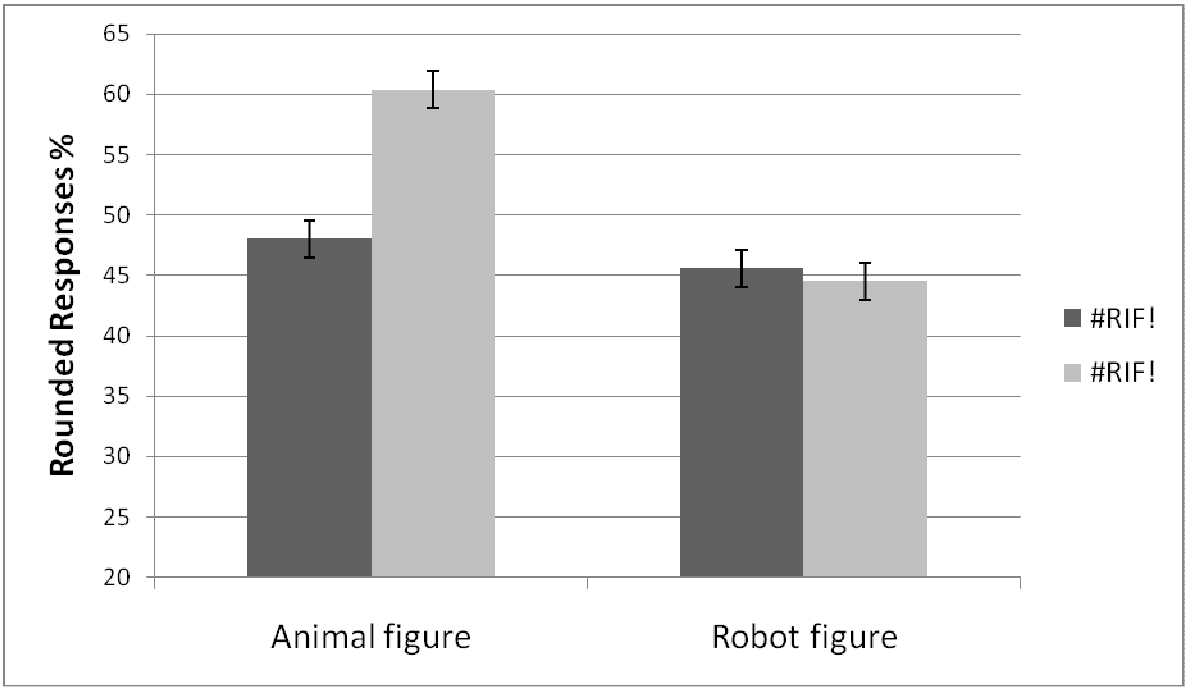


Figure 6. Experiment 2 - Interaction Figure Type x Figure Shape

No other effects or interactions were significant.

As in Experiment 1, a further analysis considering only the first presentation of each picture was performed to control for the effect of figure repetition. The rounded responses to the first presentation were entered as percentages in the mixed 2 x 2 x 2 ANOVA with the between factor Group (Children vs. Adults) and the within factors Figure Type (Animal vs. Robot), Figure Shape (Rounded vs. Jagged). Fisher's LSD *post hoc* tests were conducted on significant interactions. The analysis showed as significant the interaction Figure Type x Figure Shape, $F(2, 92) = 23.70$, $MSe = 90.55$, $p < .001$, indicating a modulation of Animal Rounded shape ($M = 59.55\%$), which received significantly more rounded responses than all other conditions (Animal Jagged shape $M = 50.42\%$; Robot Rounded shape $M = 48.30\%$, Robot Jagged shape $M = 49.86\%$) (LSD $p < .001$). No other effects or interactions were significant.

Given the different results of the analysis on the first presentation of each picture and the overall results of Experiment 2, it might be useful to look at the two groups individually, so we performed two separated 2 x 2 ANOVA, with the within factors Figure Type (Animal vs. Robot) and Figure Shape (Rounded vs. Jagged), on the data collected at the first appearance of each picture. Fisher's LSD *post hoc* tests were conducted on significant interactions. The analysis of the Children responses to the first appearance of each picture did not show main effects or interactions. The analysis of the Adults, instead, showed that a main effect almost reached significance, while the interaction was reliable. First, the expected main effect of Figure Shape was marginally significant, $F(1, 23) = 3.33$, $MSe = 354.57$, $p < .09$, due to Rounded shapes ($M = 54.65\%$) being more often assigned to a rounded sounding name than Jagged shapes ($M = 47.64\%$). The interaction between Figure Type and Figure Shape was reliable, $F(2, 46) = 6.03$, $MSe = 339.62$, $p < .05$, showing a modulation of the category especially for Animal Rounded shape ($M = 65.56\%$), which received significantly more rounded responses than all other conditions (Animal Jagged shape $M = 49.31\%$; Robot Rounded shape $M = 43.75\%$, Robot Jagged shape $M = 45.97\%$) (LSD, all *ps* < .001). No other main effects or interactions were significant.

To control if the results were related to the pictures ratings, we correlated subjects' mean rating to each picture and the mean percentage of rounded responses then assigned to that picture. While the correlation was not significant for children (r = -.03, $p > .4$), a positive significant correlation was observed for adults (r = .20, $p < .001$), indicating a fine-grained effect of the subjective perception of roundness on the choice of the name.

Mixed effects models were conducted to confirm and clarify results from ANOVA and correlations. Models started from the most complete one, with Figure Type, Rating and Group as fixed effects, and as random effects Item (i.e., each picture), Names Pair (i.e., the two names within to choose), and Participant in interaction with Figure Type, Ratings and Group. The procedure was the same as described in Experiment 1. The final model resulting from this procedure consisted of all fixed effects in interaction, and Item, Names Pair and

Participant in interaction with Figure Type and Group as random effects (AIC = 5840, BIC = 5943, Log-likelihood =-2904).

9.3.3. Discussion

The sound-symbolic correspondence between shapes and sounds was confirmed with figures of natural and artificial agents in Experiment 2. It was also observed in the adult group when considering only the first presentation of each visual stimulus. Thus, results found in Experiment 1 with natural objects and artifacts were replicated with natural and artificial agents.

Besides this “classic” sound-symbolic effect, in the overall analysis we found that label assignment was modulated by the category, thus confirming our hypothesis (see Fig. 4 and 6). As to the developmental pattern, in the adults group we found a clear interaction between sound and category, which was not present in children (see Fig. 5). Indeed, a sonorant, rounded-sounding word (e.g. maluma) was more frequently assigned to an animal, and a strident, sharp-sounding word (e.g. takete) was more frequently assigned to a robot. This result shows with a paradigm never used in studies on categorization that natural and artificial agents differ also for some general characteristics related to sounds: natural items are associated to smoother sounds in comparison to artificial ones. Results of Experiment 2 show also that the category effect interacted with the sound-shape correspondence. If the classic effect was present for animals, jagged responses were always preferred for robots, independently of their shape (as reported in Fig. 6). Thus, with artificial agents a sound-category correspondence suppressed the classic sound-shape correspondence. This interaction clearly indicated that the modulation of the takete-maluma effect due to the category appeared both in adults and children, as confirmed also by the marginally significant main effect of Figure Type (see Fig. 4) . However, it could be argued that the effect of category is due to a confound of the figures stimuli, which might be more rounded for animals than for robots. We do not believe that this is the case, for two main reasons. Firstly, neither an interaction between Figure Type and Figure Shape nor a triple interaction between Group, Figure Type and Figure Shape were observed in the experimental group's ratings. Thus, results of the ratings by the experimental groups are not predicting the interaction between Figure Type and Figure Shape, that we observed in the labeling task. Moreover, the difference between jagged and rounded robots in participants' ratings was significant (Robot: Rounded shape $M = 4.45$, Jagged shape $M = 1.97$; LSD $p < .001$), while the effect on the task performance with robots was not. This pattern of responses was not predicted by the corresponding independent groups' ratings (Robot: Rounded shape $M = 4.51$, Jagged shape $M = 2.08$; LSD $p < .001$). Secondly, and most important, mixed effects models results clearly showed that both Figure Type and Rating values (i.e., subjective perception of jaggedness), as well as Group, are crucial factors required to explain our data. While we cannot exclude that perceptual differences between the two kinds of figures we have chosen might have a contribution in performance, we believe that this cannot be the main factor underlying the resulting effects related to the Figure Type factor.

As to the developmental pattern, our results suggest that the effect emerges with age, confirming the hypothesis that categories of young children are more malleable, since the effects of language on them is not yet marked as in adults categories (Sera et al., 2002). This higher malleability of children's categories is confirmed by the fact that in children the sound-symbolic correspondence was not present during the first presentation of the stimuli, while for adults it was. This suggests that children were extracting the sound-symbolic relations and learning how to use them during the labeling task. Furthermore, the results of the correlational analysis between children's ratings and performance indicated that also the link between the perception of external entities' features related to shape and the ability to implicitly use them in cognitive tasks is still developing.

9.4. General Discussion

In Experiment 1 we asked a group of adults to choose a name between two alternatives for figures of every-day objects, i.e., natural objects (e.g., fruit) and artifacts (e.g., kitchen tools). The results showed the "classic" takete-maluma effect: we found a reliable perceptual correspondence between names sound and shapes appearance. This result reveals that the effect can be found also with every-day objects. However no effect of the objects category (natural vs. artificial) was found, even if it was predicted by the ratings of both experimental and independent groups.

In Experiment 2 we selected a subcategory among the natural and artificial categories, the category of agents, with the aim to render the two contrasting categories of natural and artificial agents more compact and better comparable. Stimuli were figures depicting animals (natural agents) or robots (artificial agents). In order to investigate the development of the effect, both adults and children were tested. This second experiment confirmed the "classic" takete-maluma effect. In addition, an effect of the category (natural vs. artificial) was observed too. First, adults only more frequently assigned rounded sounding names to animals than to robots, while for sharp sounding names the opposite was true (Figure 5). Second, the category also interacted with the takete-maluma effect, as this effect was present only with animals, while with robots a jagged response was always preferred independently on shape (Figure 6). Importantly, this interaction indicated that this modulation of the takete-maluma effect due to the category appeared both in adults and children.

Our results allow us to address the predictions made. First, we demonstrated that the sound-shape correspondence effect emerges with figures representing every-day objects or agents, that is with more ecological stimuli compared to those typically used in the literature. Second, the pairs of names used (Maurer et al., 2006) showed the predicted sound-to-shape symbolic mapping even when the figures were one-by-one presented (Nielsen & Rendall, 2011). Thus, the effect was found with a paradigm which minimized both the risk of an enhancement of the results as the risk that participants could understand the aim of the study. These two considerations confirm and strengthen the effect observed in previous studies (e.g. Maurer et al., 2006).

Third, modulations of the effect by the stimulus category were found in Experiment 2. Specifically, sound-shape correspondences were not observed with robots, which were associated more often to jagged responses despite of their actual shape. Importantly, this effect was not predicted by ratings, where adults and children both judged jagged robots as sharp, and rounded ones as round. One possible reason why we found the effect in Experiment 2, with the more compact and apparently less differentiated category of "agents", and not in Experiment 1, can depend on the special "naming habit" used by children and adults in their interactions with biological agents (e.g. animals), and with any entity presenting animacy cues (e.g. eyes, mouth) and perceived as able to autonomously act (e.g. robots). Indeed, entities perceived as agents are usually renamed during the interactions with them: children and adults typically use a special name for their pets (e.g. their cat is not called just "cat") and for their favorite teddy-bear, or robot toy. In contrast, it might seem more difficult to associate novel names to entities endowed only with generic names and not with proper names, as the objects of Experiment 1. In support of this explanation, research on the mutual exclusivity or lexical contrast constraint (Clark, 1987; Markman, 1989, 1992) has shown that during language acquisition we experience difficulties in using more names, for example a basic and a superordinate one (e.g. “dog” and “animal”), to indicate the same referent.

Fourth, while the takete-maluma effect and its modulation due to the category were stable across ages, the interaction between the produced sound and the category changed with development. Indeed, only adults showed the tendency to associate a jagged name to a robot and a rounded name to an animal *independently from the shape*. Children, differently from adults, did not show a sound-symbolic correspondence between word and category independently from shape. One could speculate that the emergence of sound-symbolic correspondences at the semantic level requires the acquisition of linguistic and cultural aspects related to categories. We explain the result referring to the linguistic experience of adults: it is possible that adults have more experience in listening to or actively associating more nouns to agents such as pets and toys. This experience might have led to associations between sound features (i.e. strident, sonorant) and categories properties (i.e. animal more rounded, robot more jagged) which go beyond the “classic” sound-symbolic correspondence between shapes and names based only on perceptual aspects of the stimulus. This result confirms that children categories are more perceptually grounded than adults ones (for the importance of shape and perceptual grounding in children categorization, see the literature on the “shape bias”, showing that names are extended on the basis of shape similarity, e.g., Landau et al., 1988, 1992; Smith, 2005). This interpretation is in line with the idea that, once the mapping between perceptual aspects and linguistic aspects is established, no grounding is necessary, but that participants can use a shortcut relying on associative knowledge (Barsalou et al., 2008; Connell & Lynott, 2013; for a discussion see Borghi et al., 2011). Importantly, the effects of category we observed cannot be merely attributed to perceptual aspects of the selected stimuli, as they are not mirroring the subjective judgments on figures roundness observed in the ratings. Nonetheless, in the ratings interesting effects of category emerged both for objects and for agents: we suggest that these results might indicate an influence of factors such as category and age on the subjective perception of roundness.

One could argue that the difference between children and adults could be due to the fact that the distinction between animate and inanimate entities becomes better differentiated with age. We tend to exclude this interpretation for at least two reasons. The first is that we found that the sound-shape correspondence was modulated by category in both adults and children. Only the interaction between sound and category independently of shape was present in adults but not in children. Thus the difference between adults and children does not seem to pertain the development of categories, but rather the development of associations between categories and linguistic labels. The second is that the literature has shown the ability to distinguish between artifacts and natural objects as emerging rather early: some studies have demonstrated with habituation, preferential looking or other methods that already prelinguistic infants are able to differentiate these two macro-categories (e.g. Behl-Chadha, 1996; Mandler, 2004; Quinn & Johnson, 2000).

Taken together, these results extend and strengthen evidence on the takete-maluma effect indicating that sound-symbolic correspondences may arise at either perceptual or semantic levels. In general, our results bolster the hypothesis of a natural relation between the structure of words and the meanings they convey, extending previous findings of the sound-symbolic literature to entities taken from every-day life (see chapter 3). Moreover, they clearly suggest a mutual influence between the naturally biased sound-shape correspondences and the cultural and linguistic learning by which categorization is socially determined.

Indeed it has been confirmed as possible not only an iconic relation between the name and its referent, but also a sound-symbolic correspondence with semantic aspects of the referents categories too, in interaction with the previous knowledge about them. As our results have been collected presenting real stimuli and invented words, a future direction for our research could be conducting experiments in which presentation is reverse, that is invented rounded/jagged figures are presented with real words (strident/sonorant, and referring to natural objects/artifact). In such a setting it is very likely to observe sound-shape correspondences as facilitated or interfered according to word meaning.

Finally, our results might have implications for the classical question about the arbitrariness of verbal language discussed in the introduction. Indeed, they can offer suggestions to speculate about a possible origin of contemporary conventionalized lexicons from more iconic ones, in keeping with the perspectives on cognition which hypothesize a direct, natural line of evolution from gestures to speech (e.g. Corballis, 2002, 2009; Gallese, 2008; Gentilucci et al., 2004; Gentilucci & Corballis, 2006; Rizzolatti & Craighero, 2004).

GENERAL CONCLUSIONS

The research project of my thesis, performed in the theoretical framework of embodiment, was aimed at demonstrating the high dynamicity and flexibility of situated simulation processes, as introduced in chapter 1. Among the issues relevant for this aim in the literature of embodied and grounded cognition, my experimental research has been focused on affordances and verbal language. These two topics were separately introduced in chapter 2 and 3 of the thesis, where they were treated at both the theoretical and empirical level, with an extensive review of the related literature. In the second part of the thesis, the two topics were investigated in a series of original studies. The common starting point for all the reported experiments was that both object/action understanding and language comprehension are tied to the context in which they take place. In this direction, the reported studies attempted to clarify the factors able to modulate the simulation processes triggered by affordance and verbal language processing.

Chapter 4 opened the experimental section of the thesis. The reported behavioural study investigated affordance-based compatibility effects related to tools embedded in complex visual contexts, defined by the presence of the tool, another object and a hand in interaction with them. The presence of the second object allowed to create different visual contexts defined by the tool-object relation, as the two items could be either spatially, functionally or not related. The active object and the second object could be presented alone, with a hand in potential interaction with the tool, or with a hand in effective interaction with the tool, in a manipulative or functional posture. In this setting, the activation of the different affordances (i.e., manipulative/functional) evoked by the same active object displayed in different contexts (i.e, move/use) were tested in a categorization task (i.e., the objects are related?) with hand or foot responses (Experiment 1 and 2, respectively). The results indicated the advantage in response times for the functional over the spatial context in both experiments. This finding suggests that in complex scenes action knowledge was available earlier and accessed faster than thematic knowledge about objects. The cause of this different accessibility was related to the differences between functional and spatial relations. While functional relations between two objects are normally clear and, in some cases, socially established, spatial relations are more subject to individual differences and less conventionalized. This higher variability might explain why participants needed longer and produced more errors to verify a potential spatial relation than a functional relation between two objects.

In the results of Experiment 1, manual responses revealed also the predicted action-context compatibility effect: manipulative postures were favored by move contexts and inhibited by use contexts, while the reverse was true for functional postures. This confirmed the hypothesis that the emergence of specific object affordances is selectively modulated by the visual and social context displayed by the pictures. Additionally, the results of Experiment 2 on foot responses did not show the compatibility of hand actions and contexts, suggesting that the findings of Experiment 1 were due to an effector-specific motor simulation, rather than to simple associations between contexts and hand-postures.

The finding of a flexible activation modulated by the context have important theoretical implications for the debate on the supposed automaticity of the affordance effect. This automaticity was often inferred from the fact that, even if the task requires processing a given aspect of an object (for example assigning it to a given a category), affordances related to other aspects (e.g., grip, orientation) were activated. However, some evidence had already suggested factors that can flexibly modulate the activation of affordances. For example, Pellicano et al. (2010) and Tipper et al. (2006) have shown that affordance effects are not present with tasks implying simple perceptual processing of the stimuli, such as color discrimination tasks, whereas they emerge when the task implies deeper processing, as in categorizations or decisions on objects shape. In line with this evidence, recent neuroimaging findings suggested that ventral stream areas for awareness of correct/incorrect contexts for man-made artifacts were not active when subjects were not seeking functional relationships between objects (Mizelle & Wheaton, 2010a). This suggests that the activation of affordances is dependent on the whole experimental situation, and the results described in chapter 4 bolsters this conclusion as well, showing that also the (visual and social) context selectively modulates the activation of affordances. The present study suggests that our cognitive systems responds flexibly to changing contexts, and support the affordance competition hypothesis (Cisek, 2007), according to which a competition between different available action opportunities is activated. The evidence reported in chapter 4 demonstrates that contextual relations between objects, as well as the presence of an hand suggesting a specific action goal, are able to solve this competition.

The study reported in chapter 5 further investigates the influence of different contexts on affordance activation focusing on a particular class of tools called conflict objects (i.e., tools endowed with different affordances associated to specific actions). In the present experiment, participants were asked to categorize (i.e., natural/artifact) conflict object pictures presented in different complex visual scenes that evoked either move- or use-related actions (i.e., move/use context). Categorization judgments were performed on a manipulandum ad-hoc by executing a move- or use-related action (i.e., clench/pinch). The results confirmed the action-context compatibility effect found in chapter 4 with conflict objects as well, showing shorter latencies when the manual response movement and the visual context were compatible. The present data extends the findings reported in the previous chapters, demonstrating that the complex visual environment in which objects were embedded were able to guide affordance activation, modulating the move- and use-related gestures performed during semantic object processing. As the reported action-context compatibility effects were related to the meaning conveyed by the action intention evoked by the whole visual scene, these findings reinforce the conclusion advanced in chapter 4 that affordances are flexibly activated in dependence of the context. Additionally, the fact that the observed compatibility effects raised from exposition to very naturalistic visual contexts also provided a further ecological validation to action evocation phenomena during object processing. On the whole, the present study confirmed the predicted influence of visual scene on stimulus-response compatibility effects during semantic object processing. This finding brings additional support to action models that consider both action sub-types and contexts as key determinants for understanding interactions between visual objects and action processing (e.g., Buxbaum & Kalénine, 2010). Finally, the evidence reported in chapter 5 confirmed that the notion of automaticity often applied to affordance-based compatibility effects should at least be redefined. Indeed, the present results were observed in respect of properties of the motor responses that were totally irrelevant to successfully carry on the task. However, it is evident that they don't stand in favor of an idea of automatic emergence or constant activation of affordances. In this direction, further research should try to integrate the different findings in the literature investigating the time scale at which the observed context effects emerge across different experimental situations. Indeed, it is still unclear if context-relevant action modulation arose before conceptual object processing is completed, thus affecting semantic processing from earlier stages of perceptual processing, or if object-related actions were automatically evoked during early processing stages, with the modulation of the context being a late filter that enhanced relevant action features during conceptual processing, discarding irrelevant ones.

On the whole, the effects described in these two chapters clearly indicate a high degree of flexibility in the emergence of affordance effects. The results reported suggest overall that all action features are not systematically integrated in the simulation of object concepts, and that the context play a crucial, selective role for this integration. Indeed, the studies presented in chapter 4 and 5, along with some findings reviewed in chapter 2 of this thesis, indicate that the situated simulation of object affordances can be modulated by several factors, including the visual scene and the social situation in which objects are embedded, the specific goals of the tasks, the mappings evaluated and the motor responses performed. This evidence about the understanding of action-related cognition confirms the general hypothesis of this thesis that flexibility and contextuality are crucial properties in the occurrence of simulation processes.

The study in chapter 6 linked the two experimental topics of this thesis, affordances and verbal language, to take a further step in the investigation of simulation processes. The principal aim of the two experiments it reports was the direct comparison between the simulation triggered by visual objects processing and the simulation triggered by words referring to objects. Indeed, much research in the literature has been aimed at showing similarities in this processes, while more recent proposals have began to focus on their differences (e.g., Borghi & Riggio, 2009). In line with this, the present study tested the emergence of affordance-based compatibility effects in a categorization task on pictures of big and small manipulable objects and on the names of the same objects. Participants listened to a category-word (i.e., artificial/natural) and had to decide among two stimuli simultaneously presented on the screen which corresponded to it, reporting their choice using either a big mouse, requiring a power grip, or a small mouse, requiring a precision grip. The results of Experiment 1 showed a compatibility effect between the hand posture on the mouse and the grip elicited by pictures of objects in function of their size: mouse trajectories were more direct, revealing less uncertainty in the decisional process, when the dimension of the mouse and the object size were congruent. In addition, a clear influence of the distractor size on the response was observed, as in the case in which the dimension of the mouse matched that of the distractor, responses were more uncertain. Furthermore, the degree of uncertainty was higher when the object and the distractor size matched than when they did not. Finally, the finding that natural objects were processed faster than artifacts not only confirmed previous research (e.g., Borghi et al., 2007), but related also to what observed in chapter 4 and 5. Indeed, since artifacts are designed to be used, not only manipulation but functional information was activated as well, with the possibility that the two kinds of information competed slowing down responses. Overall, these results revealed that participants were sensitive to the static hand postures required to use the response devices. The results obtained in Experiment 2 with words (i.e., the names of the objects presented in Experiment 1) were quite different, as they did not show the predicted compatibility between the hand posture and the implied dimension of the target stimulus. However, an interaction of the object and the distractor size in dependence of the mouse dimension showed higher response uncertainty when big targets were matched with big than small distractors. This finding revealed that participants were in a certain measure sensitive to the size of object words referred to. Finally, the results showed an higher uncertainty for artifacts compared to natural objects, matching what observed in Experiment 1.

At first sight, the absence of the predicted compatibility effect might seem problematic for an embodied account of language processing. However, evidence of the activation of motor information with words was found as well, as the presence of modulations of size suggests that words elicited modal information part of an embodied re-enactment of the associated sensory-motor experience. The overall results can be read in terms of recent proposals emerging in the embodied cognition literature. Indeed, theories of reuse (or neural exploitation, see Gallese, 2008) suggest that if language recruits structures and mechanisms characterizing the motor system, it also modifies and builds on them (e.g., Anderson, 2010;

Borghi, 2012; Pezzulo & Castelfranchi, 2009). The present results suggest that, while the compatibility between the executed grip and the observed visual object occurs online, motor information on object size is processed offline as well, influencing language comprehension too. On the whole, the reported study further confirms with static hand postures the flexibility of affordance-based compatibility effects, extending the previous evidence on the simulation triggered by visual objects to linguistic stimuli referring to objects.

The three experimental studies reported in chapter 7, 8, and 9 are focused on the second topic of the thesis, verbal language. In particular, abstract language grounding was investigated in the experiments described in chapter 7 and 8, while chapter 9 reported an original research on sound-symbolism.

The study described in chapter 7 was aimed at testing the predictions advanced about language grounding by the Words As social Tools theory (WAT), according to which the differences between abstract and concrete words are related to their different modalities of acquisition (MoA). The hypothesis tested in the present study was that abstract words meaning would be more strongly grounded in socio-linguistic experiences, while concrete words meaning would rely more on sensory-motor experiences. In order to mimic the acquisition of novel concrete and abstract words/concepts, firstly participants were submitted to a training in which they manipulated invented objects (i.e., concrete entities) and observed groups of invented objects interacting in novel ways (i.e., abstract relations) on a screen. Thus, novel abstract categories did not refer to single objects, but to complex relationships between different objects, and they were not manipulable. Then, in a further training phase, this set of stimuli was then labeled with invented category-names, and in four experiments a variety of tasks were performed to study the acquisition of the novel concrete and abstract concepts/words. Experiment 1 controlled with a production task that the pattern of properties produced for the novel categories was similar to the one typically observed for existing concrete and abstract words. In Experiment 2 and 3, abstract words were also learned by half of the participants by associating an explanations of their meaning, while in Experiment 4 the explanation was provided along with the label for concrete categories as well. This situation should resemble the learning process of children, as showed in studies on MoA (e.g., Wauters et al., 2003). The results of Experiment 2, 3 and 4 showed that this learning process influenced a later color verification task: participants responded faster to concrete words with manual responses, while response latencies for abstract words were shorter with vocal responses. This finding confirmed previous evidence on action words that showed faster responses with the microphone than with the keyboard to sentences describing mouth-related actions (Scorolli & Borghi, 2007). In addition, in line with WAT prediction, the performance with abstract words improved when participants were provided with a verbal explanation along with the label (Experiment 2 and 3), while this effect was not observed in concrete words (Experiment 4). The fact that the advantage of the explanation was confined to abstract words revealed that their difference was not simply due to phono-articulatory aspects, but that for accessing the meaning of abstract words linguistic information played a crucial role. This was confirmed by the finding that when in Experiment 4 the verbal explanations were designed to contradict the perceptually-based category features, the advantage of the microphone over the keyboard for abstract words was strongly reduced compared to the other experiments. Finally, the results of categorical recognition tasks in Experiment 3 and 4 clearly revealed that language was relevant for both concrete and abstract categories, probably because the linguistic label is a further cue that allows a better differentiation between categories (e.g., Mirolli & Parisi, 2011). These findings open an interesting scenario. Language was relevant for both concrete and abstract words, but in tasks for which categorization was not relevant, such as the final color verification task, it was more accessible in representing abstract than concrete concepts. This might occur as more linguistic information was needed in order to define the boundaries of abstract concepts, being the members of abstract categories were not manipulable.

On the whole, the results of the present study confirmed WAT predictions. They confirmed that linguistic information was more crucial than sensory-motor information in the representation of abstract concepts. However, it was also demonstrated that all the novel concepts acquired were not only grounded in perception and action systems, because in developing their representation language was important too. This finding suggests that real abstract words would activate linguistic as well as sensory-motor areas in the brain, in line with embodied and grounded theories of language comprehension. Finally, these results - revealing that a different learning process flexibly led to differences in performance on a variety of tasks (e.g., property production, property verification) - showed that the novel conceptual representations were developed and positioned in the concrete-abstract continuum in direct function of the different contexts (e.g., sensory-motor, linguistic) in which each category was acquired.

Chapter 8 reported a study that further inquire into the grounding of abstract concepts. In this case, the metaphoric mapping of an abstract dimension on a more concrete one is investigated, in particular the TIME IS SPACE metaphor, with the aim of highlighting the flexible activation of such conceptual mappings. The TIME IS SPACE metaphor consists in the use of a mental timeline to organize and spatially represent the abstract contents of temporal expressions. As in the literature previous research on the issue of space-time compatibility effect in time-irrelevant task have produced unclear evidence, the present study tested the activation of the left-right mental timeline in both an explicit temporal reference judgment and a lexical decision task. The set of stimuli was composed by verbs and pseudo-verbs endowed with past and future tense markers, especially selected to secure a deep processing of each item. Both tasks, apart from the dimension evaluated, followed an identical procedure and used an identical pool of stimuli, in order to allow a trustworthy statistical comparison of their results. The results showed a strong space-time congruency effect when participants judged temporal reference, while no compatibility was observed for the time-irrelevant lexical decisions. This happened not only in Experiment 1, when response latencies were used to measure the effect as in previous studies, but also in Experiment 2, when the mouse trajectories performed by participants to decide over concurrent response possibilities were recorded. The findings of both experiments are consistent with the view that, all other factors being equal, only the conceptual mappings that are required to carry out the task are set up in working memory. This is predicted by the coherent working models theory proposed by Santiago and collegues (2011), which fundamentally rests on the assumption that mental models of any situation are constrained to be the most internally coherent and simple as possible. The authors explicitly indicated this characteristic of mental models as the central factor for accounting the flexibility of conceptual congruency effects which has been observed in the literature in a wide variety of tasks, and that this thesis highlighted for affordance-based compatibility effects too (e.g., Borghi et al., 2012; Kalénine et al., 2013; Natraj et al., 2013). Thus, only the strictly necessary information was retrieved from the long-term memory system to be included in participants' mental model of the current situation. However, even if this evidence clearly favors a flexible view of compatibility effects, the issue of the automaticity (or lack thereof) in the activation of the mental timeline could not be considered definitively closed. Indeed, there are several directions for modifying time-irrelevant tasks that future research still needs to explore, starting from the use of temporal stimuli which have a more direct link to temporal reference (e.g., weekdays, months). A second possibility is that more sensitive and subtle measures (e.g., eye tracking) might be able to find the effects in the same experimental setting. Finally, it is also possible that future research may design tasks that in respect to lexical decisions are able to activate the spatial dimension even if requiring cognitive processes different from space-time mapping.

On the whole, the present results support flexible accounts of the activation of conceptual mapping such as the TIME IS SPACE metaphor, confirming that space-time congruency effects emerge only in the case in which participants are explicitly required to judge temporal reference of the stimuli. This evidence suggests that conceptual effects are strongly mediated by the experimental context, as it was determined by the different goals of the tasks performed.

Chapter 9 reported the last experimental study of the thesis, which focused on sound-symbolism. Following the assumption that speech evolved from gestures (e.g., Corballis, 2002), several scholars suggested that sound-symbolism played an important role in the evolution of contemporary vocal languages, sustaining grounding and embodiment of verbal languages (see par. 3.6). In line with this view, recent research emphasized the role and importance of iconicity in language as the possible mean for connecting and grounding the linguistic-communicative form in the sensory-motor characteristics of words referents. These hypotheses are thought to be supported by a variety of findings, among others the evidence of iconic correspondences between the sound of invented words (i.e., strident, sonorant) and information from the visual modality (e.g., roundness, jaggedness). However, the literature on sound-to-meaning mappings was confined in the realm of invented entities, being limited to figures created ad-hoc for the experiments. If sound-symbolism is assumed as playing a role in both natural language evolution and ontogenetic grounding, this implies the necessity of observing it in more ecological experimental settings. In order to extend the literature to existing categories, the hypothesis guiding the present study was that sound-shape correspondences would be observed when participants had to choose, between two invented words, the name which better suited an image representing an every-day entity. The results confirmed the “classic” sound-shape correspondence in this more ecological settings, showing the effect both in Experiment 1, when participants chose a name for figures of natural objects (e.g., leaf) and artifacts (e.g., fork), and in Experiment 2, when participants chose a name for figures of natural (i.e., animals) and artificial agents (i.e., robots). Furthermore, a modulation of the category emerged when participants had to name agents. Specifically, sound-shape correspondences were not observed with robots, which were associated more often to strident words despite of their actual shape. One possible reason for observing this effect with the category of agents can depend on the special “naming habit” characterizing the interactions of humans with biological agents (e.g., animals), and with any entity presenting animacy cues (e.g., eyes), or perceived as able to autonomously act (e.g., robot). Indeed, entities identified as agents are usually renamed during the interactions with them: adults typically use a special name for their pets, and children do it also for their favorite teddy-bear or robot toy. In contrast, it might seem more difficult to associate novel names to entities endowed only with generic names and not with proper names as the objects of Experiment 1. In support of this explanation, research on the mutual exclusivity or lexical contrast constraint (e.g., Markman, 1992) showed that during language acquisition we experience difficulties in using more names, for example a basic (e.g., dog) and a superordinate one (e.g., animal) to denote the same referent. In general, the results bolster the hypothesis of a natural relation between the structure of words and the meanings they convey, extending previous findings in the sound-symbolic literature to entities taken from every-day life. Moreover, they clearly suggest a mutual influence between the naturally biased sound-shape correspondences and the cultural and linguistic learning by which categorization is socially determined. Indeed, it has been confirmed as possible not only an iconic relation between the name and its referent, but also a sound-symbolic correspondence with semantic aspects of the referents categories too, in interaction with the previous knowledge about them. Finally, the present results have implications for the classical question about the arbitrariness of verbal language. Indeed, they may offer a base to speculate about a possible origin of contemporary conventionalized lexicons from more iconic ones, in keeping with the perspectives on cognition which hypothesize a direct, natural line of evolution from gestures to speech (e.g., Arbib, 2005; Corballis, 2002; Gallese, 2008; Gentilucci & Corballis, 2006; Rizzolatti & Craighero, 2004).

On the whole, this thesis describes evidence on affordances and verbal language that clearly stay in favor of embodied and grounded views of cognition. Crucially, the results collected on the investigated conceptual effects support the original hypothesis that guided this research: the high flexibility of situated simulation processes. Indeed, the evidence reported confirms that the context is able to play a crucial role in affordances emergence, metaphoric mappings activation and language grounding. The perception of the whole situation itself is flexibly shaped and determined by a number of factors, which in an experimental setting can include the goals of the task, the kind of stimuli used, the physical and social scene in which stimuli might be embedded, the mappings evaluated, the training performed, the responses executed. In other words, the identification of the experimental situation as being *that specific situation, which calls for certain appropriate behaviours,* is itself a matter of contextuality.

In conclusion, this thesis highlights that in an embodied perspective cognition is necessarily situated and anchored to a specific context, as it is sustained by the existence of a specific body immersed in a specific environment. This suggests that in future the experimental research on embodied cognition should be aimed at investigating simulation processes – and, in general, the human mind and behaviour - in the most ecological settings, if not in natural, every-day situations.

REFERENCES

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature neuroscience*, *11*(9), 1109-1116.

Akita, K., Matsumoto, Y., & Ohara, K. H. (2008). How motion sounds looks in Japanese and English: Mimetics in the typology of motion expressions. *Kobe Papers in Linguistics, 6*, 1-19. Hyogo: Kobe University.

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). NJ: Hillsdale: Erlbaum.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, *31*, 578-602.

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence, 149*, 91–130.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences, 33*(04), 245–266. doi:10.1017/S0140525X10000853.

Anderson, S. E., & Spivey, M. J. (2009). The enactment of language: Decades of interactions between linguistic and motor processes. *Language and Cognition*, *1*(1), 87-111.

Anelli, F., Nicoletti, R., & Borghi, A.M. (2010). Categorization and action. What about object consistence? *Acta Psychologica, 133*, 203-211.

Apel JK, Cangelosi A, Ellis R, Goslin J, Fischer MH. (2012)Object affordance influences instruction span. *Experimental Brain Research, 223*(2), 199-206.

Arata, M., Imai, M., Sotaro, K., Guillaume, T., & Okada, H. (2010). Perception of sound symbolism in 12 month-old infants: An ERP study. *Neuroscience Research, 68S*, 223-334.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences*, *28*(02), 105-124.

Asano, M., Kitajo, K., Arata, M., Guillaume, T., Sotaro, K., Okada, H., & Imai, M. (2011). Preverbal infants' sensitivity to sound symbolism: An EEG study. *Neuroscience Research, 71S*, 408-416.

Austin, J. (1962), *How to Do Things with Words*, Oxford, Clarendon Press.

Avant, L. L. (1965). Vision in the Ganzfeld. *Psychological Bulletin*, *64*(4), 246.

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology-Paris*, *102*(1), 35-39.

Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2006). Lateralization of the human mirror neuron system. *The Journal of neuroscience*, *26*(11), 2964-2970.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current biology*, *16*(18), 1818-1823.

Backscheider, A. G., Gelman, S. A., Martinez, I., & Kowieski, J. (1999). Children's use of different information types when learning homophones and nonce words. *Cognitive Development, 14*(4), 515-530.doi:10.1016/S0885-2014(99)00017-9

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*, 617–629.

Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PloS one*, *7*(4), e35932.

Baronchelli, A., Gong, T., Puglisi, A., & Loreto, V. (2010). Modeling the emergence of universality in color naming patterns. *Proceedings of the National Academy of Science, 107*(6), 2403-7.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(04), 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281-1289.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. *Grounding cognition: The role of perception and action in memory, language, and thought*, 129-163.

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. 245–283). Oxford, U.K.: Oxford University Press.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in cognitive sciences*, *7*(2), 84-91.

Barsalou, L.W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher and R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought*, pp. 129-163 . New York: Cambridge University Press.

Becchio, C., Cavallo, A., Begliomini, C., Sartori, L., Feltrin, G., & Castiello, U. (2012). Social grasping: From mirroring to mentalizing. *NeuroImage, 61*, 240–248.

Behl-Chadha, G. (1996).Basic-level and superordinate-like categorical representations in early infancy. *Cognition, 60*, 105-141.

Bekkering, H., & Neggers, S. F. W. (2002). Visual search is modulated by action intentions. *Psychological science, 13*(4), 370–4.

Bennet, K. M. B., Thomas, J. I., Jervis, C., & Castiello, U. (1998). Upper limb movement differentiation according to taxonomic semantic category. *NeuroReport, 9*, 255–262.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*(5), 733-764.

Berthoz, A. (1997). *The sense of movement*. Paris: Editions Odile Jacob.

Binkofski, F., & Buccino, G. (2004). Motor functions of the Broca's region. *Brain and language*, *89*(2), 362-369.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: evidence for a mental time line. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2257-2273.

Boot, I., & Pecher, D. (2010). Similarity is closeness: metaphorical mapping in a conceptual task. *The Quarterly Journal of Experimental Psychology*, *63*(5), 942-954.

Borghi, A. M. (2004). Object concepts and action: Extracting affordances from objects parts. *Acta Psychologica*, *115*(1), 69-96.

Borghi, A. M. (2005). Object concepts and action. In D. Pecher and R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought*, pp. 8-34. New York: Cambridge University Press.

. Borghi, A.M. (2012). Action language comprehension, affordances and goals. In Y. Coello, A. Bartolo (Eds). *Language and action in cognitive neuroscience. Contemporary topics in cognitive neuroscience series* (pp. 125-143). Psychology Press.

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. & Caramelli, N. (2001), Taxonomic relations and cognitive economy in conceptual organization, in Moore, J.D., & Stenning, K. (eds.), *Proceedings of 23rd Meeting of the Cognitive Science Society*, London: Erlbaum, 98-103.

Borghi, A. M., & Caramelli, N. (2003). Situation bounded conceptual organization in children: From action to spatial relations. *Cognitive Development, 18*, 49–60.

Borghi, A. M., Caramelli, N., & Setti, A. (2005). Conceptual information on objects' locations. *Brain and language*, *93*(2), 140-151.

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*(3), 763-773.

Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D. & Scorolli, C. (2011). Manipulating objects and telling words: A study on concrete and abstract words acquisition. *Frontiers in Cognition 2,* 15. doi: 10.3389/fpsyg.2011.00015.

Borghi, A. M., Flumini, A., Natraj, N., & Wheaton, L. A. (2012). One hand, two objects: Emergence of affordance in contexts. *Brain and Cognition, 80*(1), 64-73.

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, *32*(6), 863-873.

Borghi, A. M., & Pecher, D. (2011). Introduction to the special topic Embodied and Grounded Cognition. *Frontiers in Psychology, 2*, 187. doi: 10.3389/fpsyg.2011.00187.

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

Borghi, A. M., & Scorolli, C. (2009). Language comprehension and dominant hand motion simulation. *Human Movement Science*, *28*(1), 12-27.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*(1), 1-28.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive psychology*, *43*(1), 1-22.

Borreggine, K. L., & Kaschak, M. P. (2006). The action–sentence compatibility effect: It's all in the timing. *Cognitive Science*, *30*(6), 1097-1112.

Botvinick, M. M., Buxbaum, L. J., Bylsma, L. M., & Jax, S. a. (2009). Toward an integrated account of object and action selection: a computational analysis and empirical findings from reaching-to-grasp and tool-use. *Neuropsychologia, 47*(3), 671–83.

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of cognitive neuroscience*, *18*(10), 1607-1615.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica, 106,* 3–22.

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain and cognition*, *44*(2), 124-143.

Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *The journal of abnormal and social psychology*, *50*(3), 388-393.

Bruce, V. (2003). *Visual perception: Physiology, psychology, and ecology*. Psychology Press.

Bruzzo, A., Borghi, A. M., & Ghirlanda, S. (2008). Hand–object interaction in perspective. *Neuroscience Letters*, *441*(1), 61-65.

Bub, D. N., & Masson, M. E. (2010). Grasping beer mugs: on the dynamics of alignment effects induced by handled objects. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 341-358.

Bub, D. N., Masson, M. E., & Bukach, C. M. (2003). Gesturing and naming: the use of functional knowledge in object identification. *Psychological Science, 14,* 467-472.

Bub, D. N., Masson, M. E., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition*, *106*(1), 27-58.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European journal of neuroscience*,*13*(2), 400-404.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., ... & Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *Journal of Cognitive Neuroscience*, *16*(1), 114-126.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355-363.

Buccino, G., Sato, M., Cattaneo, L., Rodà, F., & Riggio, L. (2009). Broken affordances, broken objects: a TMS study. *Neuropsychologia*, *47*(14), 3074-3078.

Burgess, C., & Lund, K. (1997). Representing Abstract Words and Emotional Connotation in a High-dimensional Memory Space. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society: August 7-10, 1997, Stanford University* (Vol. 19, p. 61-65). Psychology Press.

Burling, R. (1999). Motivation, conventionalization, and arbitrariness in the origin of language. In B. J. King (Ed.), *The Origins of Language: What Nonhuman Primates Can Tell Us* (pp. 307–350). Santa Fe, NM: School of American Research Press.

Buxbaum, L. J., & Kalenine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. Annals of the New York Academy of Science, 1191, 201–218.

Buxbaum, L. J., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, *1191*(1), 201-218.

Buxbaum, L. J., Sirigu, A., Schwartz, M. F., & Klatzky, R. (2003). Cognitive representations of hand posture in ideomotor apraxia. *Neuropsychologia*, *41*(8), 1091-1113.

Caligiore, D., Borghi, A. M., Parisi, D., & Baldassarre, G. (2010). TRoPICALS: A computational embodied neuroscience model of compatibility effects. *Psychological Review*, *117*(4), 1188.

Caligiore, D., Borghi, A. M., Parisi, D., Ellis, R., Cangelosi, A., & Baldassarre, G. (2013). How affordances associated with a distractor object affect compatibility effects: A study with the computational model TRoPICALS. *Psychological Research, 77*(1), 7-19.

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral cortex*, *15*(8), 1243-1249.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(19), 1905-1910.

Cangelosi, A., Tikhanoff, V., Fontanari, J. F., & Hourdakis, E. (2007). Integrating language and cognition: A cognitive robotics approach. *Computational Intelligence Magazine, IEEE*, *2*(3), 65-70.

Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right-and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351.

Casasanto, D. J., & Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition, 106*, 579–93.

Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and time in the child's mind: Evidence for a cross-dimensional asymmetry. *Cognitive Science, 34*, 387–405.

Castiello, U. (2005). The neuroscience of grasping. *Nature Reviews Neuroscience*, *6*(9), 726-736.

Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., & Rizzolatti, G. (2007). Impairment of actions chains in autism and its possible role in intention understanding. *Proceedings of the National Academy of Sciences*, *104*(45), 17825-17830.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of experimental psychology: Learning, memory, and cognition*, *30*(3), 687.

Chambon, V., Domenech, P., Pacherie, E., Koechlin, E., Baraduc, P., & Farrer, C. (2011). What are they up to? The role of sensory evidence and prior knowledge in action understanding. *PloS one*, *6*(2), e17133.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, *12*(4), 478-484.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological psychology*, *15*(2), 181-195.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: MIT press.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*(2), 215-224.

Chersi, F., Thill, S., Ziemke, T., & Borghi, A. M. (2010). Sentence processing: linking language to motor chains. *Frontiers in neurorobotics*, *4*, 4.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1585-1599.

Clark, A. (1997). The dynamical challenge. *Cognitive Science*, *21*(4), 461-481.

Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought. Interdisciplinary Themes* (pp. 162 - 183). Oxford: Oxford University Press.

Clark, A. (1999). An embodied cognitive science? *Trends in cognitive sciences*, *3*(9), 345-351.

Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings. *Cognition*, *2*(2), 161-182.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1-33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H. (1969). Influence of language on solving three-term series problems. *Journal of Experimental Psychology*, *82*(2), 205.

Clark, H. H. (1973). Space, time, semantics, and the child. In: Moore, T. E. (Ed.). *Cognitive development and the acquisition of language*. New York: Academic Press.

Cole, M. (1996). *Cultural Psychology*. Harvard University Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240-247.

Condillac, E. B. de. (1746/1971). *An Essay on the Origin of Human Knowledge*. Gainesville, FL: Scholars Facsimiles and Reprints.

Connell, L., & Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychonomic bulletin & review*, *16*(3), 573-577.

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, 1-9.

Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton University Press.

Corballis, M. C. (2009). The evolution of language. *Annals of the New York Academy of Sciences*, *1156*(1), 19-43.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain and language*, *112*(1), 25-35.

Costantini, M., Ambrosini, E., Scorolli, C., & Borghi, A. M. (2011). When objects are close to me: Affordances in the peripersonal space. *Psychonomic Bulletin & Review, 18*(2), 32–38.

Costantini, M., & Sinigaglia, C. (2012). Grasping affordance: a window onto social cognition. In Seemann, A. (Ed.). (2011). *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*. MIT Press.

Craighero, L., Bello, A., Fadiga, L., & Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, *40*(5), 492–502.

Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1999). Action for perception: A motor-visual attentional effect. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1673-1692.

Craighero, L., Fadiga, L., Umiltà, C. A., & Rizzolatti, G. (1996). Evidence for visuomotor priming effect. *Neuroreport*, *8*(1), 347-349.

Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological review*, *99*(4), 633.

Creem, S. H., & Proffitt, D. R. (2001a). Defining the cortical visual systems: “what”, “where”, and “how”. *Acta Psychologica*, *107*(1), 43-68.

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

Creem-Regehr, S. H., & Lee, J. N. (2005). Neural representations of graspable objects: are tools special? *Cognitive Brain Research*, *22*(3), 457-469.

Cross, E. S., de C Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *Neuroimage, 31*(3), 1257–1267.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*, 615–627.

Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.

Darwin, C. (1871/2004). *The descent of man*. Digireads.com Publishing.

De Vega, M., Robertson, D. A., Glenberg, A. M., Kaschak, M. P., & Rinck, M. (2004). On doing two things at once: Temporal constraints on actions in language comprehension. *Memory & Cognition*, *32*(7), 1033-1043.

Deerwester, S., Dumais, G. W., Furnas, S. T., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*, 391–407.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371.

Devlin, J. T., Moore, C. J., Mummery, C. J., Gorno-Tempini, M. L., Phillips, J. A., Noppeney, U., ... & Price, C. J. (2002). Anatomic constraints on cognitive theories of category specificity. *Neuroimage*, *15*(3), 675-685.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, *91*(1), 176-180.

Dove, G. (2009). Beyond Conceptual Symbols. A Call for Representational Pluralism. *Cognition, 110*, 412-431.

Edelman, G. M. (1978). Group selection and phasic reentrant signaling: A theory of higher brain function. In G. M. Edelman & V. B. Mountcastle (Eds.), *The mindful brain: Cortical organisation and the group-selective theory of higher brain function.* Cambridge, MA: MIT Press.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection.* New York : Basic Books.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. Basic books.

Edwards, M. G., Humphreys, G. W., & Castiello, U. (2003). Motor facilitation following action observation: A behavioural study in prehensile action. *Brain and Cognition*, *53*(3), 495-502.

Ehrsson, H. H., Fagergren, A., Forssberg, H., Mosier, K., Lau, C., Wang, Y., Venkadesan, M., et al. (2001). Differential Fronto-Parietal Activation Depending on Force Used in a Precision Grip Task : An fMRI Study. *Journal of Neurophysiology, 85*(6), 2613–2623.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology, 91*(4), 451–471.

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*(3), 670.

Emmorey, K. (2001). *Language, cognition, and the brain: Insights from sign language research*. Psychology Press.

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic Thinking: The Apprehension and Consequences of Thematic Relations. *Psychology of Learning and Motivation-Advances in Research and Theory*, *54*, 249.

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head Up, Foot Down Object Words Orient Attention to the Objects' Typical Location. *Psychological Science*, *19*(2), 93-97.

Fagioli, S., Hommel, B., & Schubotz, R. I. (2007). Intentional control of attention: action planning primes action-related stimulus dimensions. *Psychological Research*, *71*(1), 22–29.

Fay N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science, 34*, 351–386.

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., ... & Fadiga, L. (2009). Encoding of human action in Broca's area. *Brain*, *132*(7), 1980-1988.

Ferrari, P. F., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience*, *17*(2), 212-226.

Ferri, F., Riggio, L., Gallese, V., & Costantini, M. (2011). Objects and their nouns in peripersonal space. *Neuropsychologia, 49*(13):3519-24.

Fias W., Lauwereyns J., & Lammertyn J. (2001). Irrelevant digits affect feature-based attention depending on the overlap of neural circuits. *Cognitive Brain Research, 12*, 415-423.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*(6), 825-850.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature neuroscience*, *6*(6).

Fischer, M. H., Prinz, J., & Lotz, K. (2008). Grasp cueing shows obligatory attention to action goals. *The Quarterly Journal of Experimental Psychology*,*61*(6), 860-868.

Fischer, M. H., Warlop, N., Hill, R. L., & Fias, W. (2004). Oculomotor bias induced by number perception. *Experimental Psychology*, *51*(2), 91.

Fischer, M.H., & Zwaan, R. (2008). Embodied Language: A Review of the Role of the Motor System in Language Comprehension. *The Quarterly Journal of Experimental Psychology, 61,* 825-850.

Flumini, A. (2009). La fenomenologia della parola di Maurice Merleau-Ponty: per un ritorno al «soggetto parlante». *RIFL - Rivista Italiana di Filosofia del Linguaggio, 1*, 100-121.

Flumini, A. (2010). Embodied Cognition, parole "astratte" e lingue. In G. Ferrari, P. Bouquet, M. Cruciani, F. Giardini (Eds.), *Pratiche della cognizione. Atti del 7° Convegno Nazionale dell'Associazione Italiana di Scienze Cognitive* (pp. 102-106). Trento: Università degli Studi di Trento.

Flumini, A. (2011). Iconicità, non arbitrarietà: simbolismo fonetico ed evoluzione del linguaggio verbale. In F. Rubinacci, A. Rega, N. Lettieri (Eds.), *Le Scienze Cognitive in Italia - AISC 2011. Atti dell'8° Convegno Nazionale dell'Associazione Italiana di Scienze Cognitive* (pp. 100-103). Milano.

Flumini, A., Barca, L., Borghi, A. M., & Pezzulo, G. (*submitted*). How do you hold your mouse? Mouse-tracking the compatibility effect between hand posture and stimulus size.

Flumini, A., Ranzini, M., & Borghi, A. M. (*submitted*). Nomina sunt consequentia rerum. Sound-shape correspondences with every-day objects figures.

Flumini, A., & Santiago, J. (*in preparation*). Flexibility in the activation of the left-right mental timeline.

Flumini, A., & Santiago, J. (2013). Time (also) flies from left to right... if it is needed! In M. Knauff, M. Paulen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society* (pp. 2315-2320). Austin, TX: Cognitive Science Society.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press: Cambridge.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Clarendon Press/Oxford University Press.

Fogassi L., Ferrari P.F., Gesierich B., Rozzi S., Chersi F., Rizzolatti G., 2005. Parietal lobe: from action organization to intention understanding. *Science, 308*, 662-667.

Fogassi, L., & Ferrari, P. F. (2007). Mirror neurons and the evolution of embodied language. *Current directions in psychological science*, *16*(3), 136-141.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding.*Science*, *308*(5722), 662-667.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research, 12*(3), 189–98.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavioural Research Methods*, *42*(1), 226–241.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*.

Freina, L., Baroni, G., Borghi, A. M., & Nicoletti, R. (2009). Emotive concept nouns and motor responses: Attraction or repulsion? *Memory & cognition*,*37*(4), 493-499.

Friederici, A. D. (2004). Event-related brain potential studies in language. *Current neurology and neuroscience reports*, *4*(6), 466-470.

Gainotti, G. (2000). What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: a review. *Cortex, 36*(4), 539-59.

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 PRPF*, *73*(4), 486-498.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, *22*(3-4), 455-479.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in cognitive sciences*, *15*(11), 512-519.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-609.

Gangitano, M., Mottaghy, F. M., & Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*,*12*(7), 1489-1492.

Garbarini, F., & Adenzato, M. (2004). At the root of embodied cognition: Cognitive science meets neurophysiology. *Brain and cognition*, *56*(1), 100-106.

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science*, *165*, 664–672.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007).Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science, 31*(6), 961–987.

Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. *Journal of Verbal Learning and Verbal Behavior, 8*, 310–312.

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews*, *30*(7), 949-960.

Gentilucci, M., & Volta, R. D. (2008). Spoken language and arm gestures are controlled by the same motor control system. *The Quarterly Journal of Experimental Psychology*, *61*(6), 944-957.

Gentilucci, M., Benuzzi, F., Bertolani, L., Daprati, E., & Gangitano, M. (2000). Language and motor control. *Experimental Brain Research*, *133*(4), 468-490.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology*, *86*(4), 1685-1699.

Gentilucci, M., Santunione, P., Roy, A. C., & Stefanini, S. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *European Journal of Neuroscience*, *19*(1), 190-202.

Gerlach, C., Law, I., & Paulson, O. B. (2002). When action turns into words. Activation of motor-based knowledge during categorization of manipulable objects. *Journal of Cognitive Neuroscience*, *14*(8), 1230-1239.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organized: evidence from days of the week. *Cortex, 40*, 171–172.

Gianelli, C., Dalla Volta, R., Barbieri, F., & Gentilucci, M. (2008). Automatic grasp imitation following action observation affects estimation of intrinsic object properties. *Brain research*, *1218*, 166-180.

Gibbs, R. W. J. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Mahwah, NJ: Erlbaum.

Gibson, J. J., & Dibble, F. N. (1952). Exploratory experiments on the stimulus conditions for the perception of a visual surface. *Journal of experimental psychology*, *43*(6), 414.

Gibson, J. J., & Waddell, D. (1952). Homogeneous retinal stimulation and visual perception. *The American Journal of Psychology*, *65*(2), 263-270.

Girardi, G., Lindemann, O., & Bekkering, H. (2010). Context effects on the processing of action-relevant object features. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 330.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and brain sciences*, *20*(01), 41-50.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905-922.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin & review*, *9*(3), 558-565.

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*(3), 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*(6), 905-919.

Glover, S., & Dixon, P. (2002). Semantics affect the planning but not control of grasping. *Experimental Brain Research*, *146*(3), 383-387.

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research*, *154*(1), 103-108.

Goslin, J., Dixon, T., Fischer, M. H., Cangelosi, A., & Ellis, R. (2012). Electrophysiological examination of embodiment in vision and action. *Psychological Science, 23*(2),152-7.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*, *6*(4), 231-236.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of Language* (73-113). Cambridge, MA.

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human brain mapping*, *12*(1), 1-19.

Grèzes, J., & Decety, J. (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia*, *40*(2), 212-222.

Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study. *Neuroimage*, *18*(4), 928-937.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, *20*(3), 371-388.

Haazebroek, P., Van Dantzig, S., & Hommel, B. (2011). A computational model of perception and action for cognitive robotics. *Cognitive processing*, *12*(4), 355-365.

Hamilton, A.F., & Grafton, S.T., (2008). The motor hierarchy: from kinematics to goals and intentions, in: Haggard, P., Rosetti, Y., Kawato, M., (Eds.), *Attention and performance XXII* (pp. 381–407). Oxford University Press, New York.

Hammond, M. (2010). What is an affordance and can it help us understand the use of ICT in education? *Education and information technologies*, *15*(3), 205-217.

Hannus, A., Cornelissen, F. W., Lindemann, O., & Bekkering, H. (2005). Selection-for-action in visual search. *Acta Psychologica*, *118*(1), 171-191.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335-346.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307.

Hinton, L., Nichols, J., & Ohala, J. J. (1994). *Sound symbolism.* New York: Cambridge University Press.

Hockett, C. (1978). In search of love's brow. *American Speech*, *53*, 243–315.

Holland, D., Lachicotte, W. S., Skinner, D., & Cain, C. (2001). *Identity and Agency in Cultural Worlds*. Cambridge, MA: Harvard University Press.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and brain sciences*, *24*(05), 849-878.

Hsu, N. S., Kraemer, D. J., Oliver, R. T., Schlichting, M. L., & Thompson-Schill, S. L. (2011). Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables. *Journal of Cognitive Neuroscience*, *23*(9), 2544-2557.

Humphreys, G. (2001). Objects, affordances… action! *The Psychologist.*

Humphreys, G.W., & Forde, E.M. (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behavioral and Brain Sciences, 24*(3), 453-76; discussion 476-509.

Hurford, J. R. (2011). *The origins of grammar: Language in the light of evolution II* (Vol. 2). Oxford University Press.

Hurley, S. L. (2002). *Consciousness in action.* Harvard University Press.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol, 3*(3), 529-535, e79.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? *Child Development*, *76*(2), 340-355.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism between a word and an action facilitates early verb learning. *Cognition, 109*, 54–65.

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2007). What do English speakers know about gera-gera and yota-yota?: A cross-linguistic investigation of mimetic words for laughing and walking. *Japanese-Language Education around the Globe*, *17*, 53-78.

Jax, S. A., & Buxbaum, L. J. (2010). Response interference between functional and structural actions linked to the same familiar object. *Cognition*, *115*(2), 350-355.

Jax, S. A., & Buxbaum, L. J. (2013). Response interference between functional and structural object-related actions is increased in patients with ideomotor apraxia. *Journal of neuropsychology, 7*(1), 12–8.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, *14*(1), S103-S109.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in neurosciences*, *18*(7), 314-320.

Jeannerod, M., Decety, J., & Michel, F. (1994). Impairment of grasping movements following a bilateral posterior parietal lesion. *Neuropsychologia*,*32*(4), 369-380.

Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., & Binkofski, F. (2010). Grasping language–A short story on embodiment. *Consciousness and cognition*, *19*(3), 711-720.

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in cognitive sciences*, *8*(2), 71-78.

Kalénine, S., Bonthoux, F., & Borghi, A. M. (2009). How action and context priming influence categorization: a developmental study. *British Journal of Developmental Psychology*, *27*(3), 717-730.

Kalénine, S., Bonthoux, F., & Borghi, A. M. (2009a). How action and context priming influence categorization: A developmental study. British Journal of Developmental Psychology, 27, 717–730.

Kalenine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of experimental psychology. Learning, memory, and cognition*, *38*(5), 1274–95.

Kalénine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., & Baciu, M. (2009). The sensory-motor specificity of taxonomic and thematic conceptual relations: A behavioral and fMRI study. *Neuroimage*, *44*(3), 1152-1162.

Kalénine, S., Shapiro, A. D., Flumini, A., Borghi, A. M., & Buxbaum, L. J. (2013). Visual context modulates potentiation of grasp types during semantic object categorization. *Psychonomic Bulletin & Review,* 1-7.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, *35*(3), 575-586.

Keil. F.C. (1989). *Concepts. kinds, and cognitive development*. Cambridge: Cambridge University Press.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166.

Kinoshita, S., Lupker, S.J., & Rastle, K. (2004). Modulation of regularity and lexicality effects in reading aloud. *Memory and Cognition, 32*, 1255–1264.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics, 35*, 379–415.

Kita, S. (2008). World-view of protolanguage speakers as inferred from semantics of sound symbolic words: A case of Japanese mimetics. In *The Origins of Language* (pp. 25-38). Springer Japan.

Kita, S., Kantartzis, K., & Imai, M. (2010). Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage. In *Evolution of language. Proceedings of the 8th international conference (Evolang8)* (pp. 206-213).

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, *297*(5582), 846-848.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Köhler, W. (1947/1992). *Gestalt psychology: An introduction to new concepts in modern psychology*. WW Norton & Company.

Koriat, A., & Levy, I. (1979). Figural symbolism in Chinese ideographs. *Journal of Psycholinguistic Research*, *8* (4),353–365.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19-28.

Kunihira, S. (1971).Effects of expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 10*, 427–429.

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 726.

Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General, 141*(3), 584–594.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York, NY: Basic Books.

Landau, B., Smith, L. B., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development 3*(3), 299-321.

Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language 31*(6), 807-825.

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*(2), 211.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, *25*(2-3), 259-284.

Landy, D. H., Jones, E. L., & Hummel, J. E. (2008). Why spatial-numeric associations aren't evidence for a mental number line. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 357-362).

Laurence, E., & Margolis, S. (Eds.) (2007). *Creations of the Minds. Theories of Artifacts and Their Representation*. Oxford: Oxford University Press.

Lee, C., Middleton, E., Mirman, D., Kalenine, S., & Buxbaum, L. J. (2013). Incidental and context-responsive activation of structure- and function-based action features during object identification. *Journal of experimental psychology. Human perception and performance*, *39*(1), 257–70.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts.*Journal of Experimental Psychology: General*, *130*(1), 3.

Liuzza, M. T., Setti, A., & Borghi, A. M. (2012). Kids observing other kids' hands: visuomotor priming in children. *Consciousness and cognition*, *21*(1), 383-392.

Lotto, L., Dell'Acqua, R., & Job, R. (2001). Le figure PD/DPSS. Norme di accordo sul nome, tipicità, familiarità, età di acquisizione e tempi di denominazione per 266 figure. *Giornale Italiano di Psicologia, 28*, 193-207.

Lucariello, J., Kyratzis, A., & Nelson, K. (1992). Taxonomic knowledge: What kind and when?. *Child development*, *63*(4), 978-998.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in psychology, 3.*

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99*, 587–604.

Mandler, J. M. (2004). *The foundations of mind: The origins of conceptual thought*. New York: Oxford

Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction*. Cambridge, MA: MIT Press.

Markman, E. M. (1992) Constraints on word learning: speculations about their nature, origins and domain specificity. In Gunnar, M. R. and Maratsos, M. P. (eds.) *Modularity and Constraints in Language and Cognition: The Minnesota Symposium on Child Psychology*, pp. 59–101, Erlbaum.

Marocco, D., Cangelosi, A., Fischer, K., & Belpaeme, T. (2010) Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Frontiers in Neurorobotics*, *4*.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* Henry Holt and Co. Inc., New York, NY.

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

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*(6566), 649-652.

Masson, M. E., Bub, D. N., & Breuer, A. T. (2011). Priming of reach and grasp actions by handled objects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1470.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental Science, 9*(3), 316–322.

Meier, B. P., & Robinson, M. D. (2004). Why the Sunny Side Is Up: Associations Between Affect and Vertical Position. *Psychological science*, *15*(4), 243-247.

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why Good Guys Wear White Automatic Inferences About Stimulus Valence Based on Brightness. *Psychological Science*, *15*(2), 82-87.

Merleau-Ponty, M. (1945/1962). *Phenomenology of Perception*. London: Routledge.

Merleau-Ponty, M. (2003). *The Nature: Course Notes from the Collège de France*. Northwestern University Press.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex, 48,* 788–804.

Michaels, C. F. (1988). SR compatibility between response position and destination of apparent motion: evidence of the detection of affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 231.

Michaels, C. F. (1993). Destination compatibility, affordances, and coding rules: A reply to Proctor, Van Zandt, Lu, and Weeks.

Michaels, C. F., Withagen, R., Jacobs, D. M., Zaal, F. T., & Bongers, R. M. (2001). Information, perception, and action: A reply to commentators. *Ecological Psychology*, *13*(3), 227-244.

Milner, A. D., & Goodale, M. A. (1995). The visual brain in action. *New York: Oxford*. Myachykov, A.,Ellis, R., Cangelosi, A., & Fischer, M.H. (2013). Visual and linguistic cues to graspable objects. *Experimental Brain Research, 229*(4), 545-59.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. New York: Oxford.

Minsky, N. (1975). *A framework for representing knowledge in the psychology of computer vision* (Winston O., ed) McGraw Hill.

Mirolli, M., & Parisi, D. (2011). Towards a Vygotskyan cognitive robotics: the role of language as a cognitive tool. *New Ideas in Psychology*, *29*(3), 298-311.

Mizelle, J. C., & Wheaton, L. A. (2010a). Neural activation for conceptual identification of correct versus incorrect tool–object pairs. *Brain research*, *1354*, 100-112.

Mizelle, J. C., & Wheaton, L. A. (2010b). Why is that Hammer in My Coffee? A Multimodal Imaging Investigation of Contextually Based Tool Understanding. *Frontiers in human neuroscience*, *4*, 233-233.

Mizelle, J. C., & Wheaton, L. A. (2011). Testing perceptual limits of functional units: Are there "automatic" tendencies to associate tools and objects?.*Neuroscience letters*, *488*(1), 92-96.

Molnar-Szakacs, I., Wu, A. D., Robles, F. J., & Iacoboni, M. (2007). Do you see what I mean? Corticospinal excitability during observation of culture-specific gestures. *PLoS One*, *2*(7), e626.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, *48*(2), 507-517.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current biology*, *20*(8), 750-756.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of neurophysiology*, *78*(4), 2226-2230.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge: MIT.

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: what a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 572.

Myung, J. Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition*, *98*(3), 223-243.

Myung, J., Blumstein, S. E., Yee, E., Sedivy, J. C., Thompson-Schill, S. L., & Buxbaum, L. J. (2010). Impaired access to manipulation features in Apraxia: evidence from eye-tracking and semantic judgment tasks. *Brain and language*, *112*(2), 101–12.

Natraj, N., Poole, V., Mizelle, J. C., Flumini, A., Borghi, A. M., & Wheaton, L. A. (2013). Context and hand posture modulate the neural dynamics of tool-object perception. *Neuropsychologia*, *51*(3), 506-519.

Neisser, U. (1977). Gibson's Ecological Optics: Consequences of a Different Stimulus Description. *Journal for the Theory of Social Behaviour*, *7*(1), 17-28.

Neisser, U. (1994). Multiple systems: A new approach to cognitive theory. *European Journal of Cognitive Psychology, 6*, 225–241.

Neisser, U. (1995). Criteria for an ecological self. *Advances in psychology*, *112*, 17-34.

Nielsen, A., & Rendall, D. (2011). The Sound of Round: Evaluating the Sound-Symbolic Role of Consonants in the Classic Takete-Maluma Phenomenon. *Canadian Journal of Experimental Psychology, 65*(2), 115–124.

Noë, A. (2004). *Action in perception*. MIT Press.

Núñez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive science*, *30*(3), 401-450.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009a). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33*, 127–146.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009b). Sound to meaning correspondences facilitate word learning. *Cognition, 112*, 181–186.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, *24*(5), 939-972.

Ouellet, M., Santiago, J., Israeli, Z., & Gabay, S. (2010). Is the future the right time? *Experimental psychology*, *57*(4), 308.

Oztop E., & Arbib M.A. (2002). Schema Design and Implementation of the Grasp-Related Mirror Neuron System. *Biological Cybernetics* 87: (2) 116-140

Pagliuca, G., Arduino, L.S., Barca, L., & Burani, C. (2008). Fully transparent orthography, yet lexical reading aloud: The lexicality effect in Italian. *Language and Cognitive Processes, 23*, 422–433.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255.

Paivio, A., Yuille, J.C., & Madigan, S.A. (1986). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology*, *Monograph Supplement*, *76*, 3, part 2.

Parault, S. J. (2006). Sound symbolic word learning in context. *Contemporary Educational Psychology, 31*, 228–252.

Parise, C., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research, 214*(3), 373-380.

Paulus, M., Lindemann, O., & Bekkering, H. (2009). Motor simulation in verbal knowledge acquisition. *The Quarterly Journal of Experimental Psychology*, *62*(12), 2298-2305.

Pavese, A., & Buxbaum, L. J. (2002). Action matters: The role of action plans and object affordances in selection for action. *Visual Cognition, 9*(4), 559–590.

Pecher, D., Boot, I., & Van Dantzig, S. (2011). Abstract concepts: sensory-motor grounding, metaphors, and beyond., in *The Psychology of Learning and Motivation*, Vol. 54, Ed. B. Ross (Burlington: Academic Press), 217–248.

Pecher, D., Van Dantzig, S., Boot, I., Zanolie, K., & Huber, D. E. (2010). Congruency between word position and meaning is caused by task-induced spatial attention. *Frontiers in psychology*, *1*, 30.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review*, *11*(1), 164-167.

Pecher, D., & Zwaan, R. A. (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press.

Pellicano, A., Iani, C., Borghi, A. M., Rubichi, S., & Nicoletti, R. (2010). Simon-like and functional affordance effects with tools: The effects of object perceptual discrimination and object action state. *The Quarterly Journal of Experimental Psychology*, *63*(11), 2190-2201.

Pellicano, A., Thill, S., Ziemke, T., & Binkofski, F. (2011). Affordances, adaptive tool use and grounded cognition. *Frontiers in psychology*, *2*.

Perani, D., Fazio, F., Borghese, N. A., Tettamanti, M., Ferrari, S., Decety, J., & Gilardi, M. C. (2001). Different brain correlates for watching real and virtual hand actions. *Neuroimage*, *14*(3), 749-758.

Perniss, P., & Vigliocco, G. (in press). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B.*

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227.

Pezzulo, G., & Castelfranchi, C. (2009). Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychological Research PRPF*, *73*(4), 559-577.

Pezzulo, G., Barca, L., Bocconi, A. L., & Borghi, A. M. (2010). When affordances climb into your mind: Advantages of motor simulation in a memory task performed by novice and expert rock climbers. *Brain and cognition*, *73*(1), 68-73.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2011). The mechanics of embodiment: a dialog on embodiment and computational modeling. *Frontiers in Psychology*, *2*. (5), 1–21.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2013). Computational Grounded Cognition: a new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, *3*, 612.

Pezzulo, G., Candidi, M., Dindo, H., & Barca, L. (2013). Action simulation in the human brain: twelve questions. *New Ideas in Psychology*, *31*(3), 270–290.

Phillips, J. C. & 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. (1969). *The child's conception of time*. London: Routledge & Keagan Paul. Previtali, P., de Hevia, M.D., & Girelli, L. (2010). Placing order in space: the SNARC effect in serial learning. *Experimental Brain Research, 201*, 599–605.

Pietrandrea, P. (2002). Iconicity and arbitrariness in Italian Sign Language. *Sign Language Studies, 2*, 296–321.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(04), 707-727.

Pizutto, E., & Volterra, V. (2000). Iconicity and transparency in sign languages: A cross-linguistic crosscultural view. In K. Emmorey & H. Lane (Eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima* (pp. 261–286), Mahwah, NJ: Lawrence Erlbaum Associates.

Preissl, H., Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Evoked potentials distinguish between nouns and verbs. *Neuroscience Letters*, *197*(1), 81-83.

Prinz, W. (1990). *A common coding approach to perception and action* (pp. 167-201). Springer Berlin Heidelberg.

Prinz, W. (1997). Perception and action planning. *European journal of cognitive psychology*, *9*(2), 129-154.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological bulletin*, *132*(3), 416.

Pulvermüller, F. (2003). *The neuroscience of language: on brain circuits of words and serial order*. Cambridge University Press.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576-582.

Pulvermüller, F. (2012). Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics*, *25*(5), 423-459.

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and language*, *78*(2), 143-168.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793-797.

Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. *Cerebral cortex*, *9*(5), 497-506.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*,*17*(6), 884-892.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological bulletin*, *80*(1), 1.

Pylyshyn, Z. W. (1986). *Computational cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and brain sciences*, *22*(3), 341-365.

Quinn, P. C., & Johnson, M. H. (2000). Global-Before-Basic Object Categorization in Connectionist Networks and 2-Month-Old Infants. *Infancy, 1*, 31–46.

Radden, G. (2004). The metaphor TIME AS SPACE across languages. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, *8*(2/3), 1-14.

Rakison, D. H., & Oakes, L. M. (Eds.) (2003). *Early Category and concept development: Making Sense of the Blooming Buzzing Confusion*. New York: Oxford University Press.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological bulletin, 127*(2), 209.

Rakison, D. H., & Poulin-Dubois, D. (2002). You go this way and I'll go that way: developmental changes in infants' detection of correlations among static and dynamic features in motion events. *Child Development, 73*, 682-699.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia - a window into perception, thought and language. *Journal of consciousness studies*, *8*(12), 3-34.

Raos, V., Umiltá, M. A., Murata, A., Fogassi, L., & Gallese, V. (2006). Functional properties of grasping-related neurons in the ventral premotor area F5 of the macaque monkey. *Journal of neurophysiology*, *95*(2), 709-729.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388-396.

Reilly, J., Westbury, C., Kean, J., & Peelle, J.E. (2012). Arbitrary Symbolism in Natural Language Revisited: When Word Forms Carry Meaning. *PLoS ONE, 7*(8): e42286. doi:10.1371/journal.pone.0042286

Riggio, L., Iani, C., Gherri, E., Benatti, F., Rubichi, S., & Nicoletti, R., (2008). The role of attention in the occurrence of the affordance effect. *Acta Psychologica, 127,* 449-458.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in neurosciences*, *21*(5), 188-194.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research*, *71*(3), 491-507.

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

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalography and clinical neurophysiology*, *106*(4), 283-296.

Rizzolatti, G., & Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Experimental Brain Research*, *153*(2), 146-157.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: How our minds share actions and emotions*. Oxford University Press.

Rizzolatti, G., & Sinigaglia, C., (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience, 11*(4)**,** 264-274.

Roth, W. M., Woszczyna, C., & Smith, G. (1996). Affordances and constraints of computers in science education. *Journal of Research in science Teaching*, *33*(9), 995-1017.

Rousseau, J.J. (1782/1969). *Essai sure l'origine des langues*. Paris: A. G. Nizet.

Roversi, C., Borghi, A. M., & Tummolini, L. (2013). A Marriage is an Artefact and not a Walk that We Take Together: An Experimental Study on the Categorization of Artefacts. *Review of Philosophy and Psychology*, 1-16.

Ruiz Fernández, S., Rahona, J. J., Hervás, G., Vázquez, C., & Ulrich, R. (2011). Number magnitude determines gaze direction: Spatial–numerical association in a free-choice task. *Cortex*, *47*(5), 617-620.

Rumiati, R. I., & Humphreys, G. W. (1998). Recognition by action: dissociating visual and semantic routes to action in normal observers. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 631.

Rüschemeyer, S. A., Brass, M., & Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor stems. *Journal of Cognitive Neuroscience*, *19*(5), 855-865.

Sabsevitz, D.S., Medler, D.A., Seidenberg, M.,& Binder, J.R. (2005). Modulation of the semantic system by word imageability. *Neuroimage*, *27*, 188-200.

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

Sakreida, K., Scorolli, C., Menz, M. M., Heim, S., Borghi, A. M., & Binkofski, F. (2013). Are abstract action words embodied? An fMRI investigation at the interface between language and motor cognition. *Frontiers in human neuroscience*, *7*.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2661-2665.

Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, *108*(1), 263-270.

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review, 14*, 512-516.

Santiago, J., Ouellet, M., Román, A., & Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cognitive science*, *36*(6), 1051-1077.

Santiago, J., Román, A., & Ouellet, M. (2011). Flexible foundations of abstract thought: A review and a theory. In A. Maass & T. W. Schubert (Eds.), *Spatial dimensions of social thought* (pp- 39-108). Berlin: Mouton de Gruyter.

Santiago, J., Román, A., Ouellet, M., Rodríguez, N., & Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychological Research PRPF*, *74*(1), 59-70.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225–239.

Saussure, F. (1916/1966). *Course in general linguistics*. New York, USA: McGraw Hill.

Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, Language, and the Human Mind*. New York: Oxford University Press.

Saygın, A. P., Wilson, S. M., Dronkers, N. F., & Bates, E. (2004). Action comprehension in aphasia: linguistic and non-linguistic deficits and their lesion correlates. *Neuropsychologia*, *42*(13), 1788-1804.

Schank, R. C., Abelson, R. P., & Scripts, P. (1977). *Goals and understanding*. Erlbanum: Eksevier Science.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *Journal of personality and social psychology*, *89*(1), 1.

Schwartz, M. F., Brecher, A. R., Whyte, J., & Klein, M. G. (2005). A patient registry for cognitive rehabilitation research: a strategy for balancing patients' privacy rights with researchers' need for access. *Archives of physical medicine and rehabilitation, 86*(9), 1807–14.

Scorolli, C., Binkofski, F., Buccino, G., Nicoletti, R., Riggio, L., & Borghi, A. M. (2011). Abstract and concrete sentences, embodiment, and languages. *Frontiers in psychology*, *2*.

Scorolli, C., & Borghi, A. M. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain research*, *1130*, 119-124.

Scorolli, C., Jacquet, P. O., Binkofski, F., Nicoletti, R., Tessari, A., & Borghi, A. M. (2012). Abstract and concrete phrases processing differentially modulates cortico-spinal excitability. *Brain research*, *1488*, 60-71.

Sell, A. J., & Kaschak, M. P. (2011). Processing time shifts affects the execution of motor responses. *Brain and language*, *117*(1), 39-44.

Semin, G. R., & Smith, E. R. (Eds.). (2008). *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. Cambridge University Press.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: an analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, *131*(3), 377.

Šetić, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, *22*(2), 297-312.

Shadmehr, R., Smith, M.A., & Krakauer, J.W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, *33*, 89–108. doi:10.1146/annurev-neuro-060909-153135

Shapiro, L. (2010). *Embodied cognition*. Taylor & Francis.

Shaw, R., Turvey, M. T., & Mace, W. (1982). Ecological psychology: The consequence of a commitment to realism. *Cognition and the symbolic processes*, *2*, 159-226.

Shen, K., & Pare, M. (2011). Neural basis of feature-based contextual effects on visual search behavior. *Frontiers in behavioral neuroscience*, *5*, 91.

Shepard, R. N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological review*, *91*(4), 417.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science (New York, NY)*, *171*(3972), 701.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802-2810.

Smith, L. B. (2005). Action alters shape categories. *Cognitive Science*, *29*, 665–679.

Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Science*, *13*(8), 360–366. doi:10.1016/j.tics.2009.04.009

Spector, F., & Maurer, D. (2008). The color of Os: Naturally biased associations between shape and color. *Perception, 37*, 841–847.

Spivey, M. (2007). *The continuity of mind*. Oxford University Press, USA.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological research*, *65*(4), 235-241.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological science*, *12*(2), 153-156.

Steels, L. (2011). Modeling the cultural evolution of language. *Physical Life Review, 8*(4), 339-56.

Steels, L., (2003) Evolving grounded communication for robots. *Trends in Cognitive Sciences*, *7*, 308-312.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American scientist*, *57*(4), 421-457.

Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: selective neural encoding of apparent human movements. *Neuroreport*, *11*(1), 109-115.

Sturmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study on imitation. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1746–1759.

Sugita, Y., & Tani, J. (2005) Learning Semantic Combinatoriality from the Interaction between Linguistic and Behavioral Processes. *Adaptive Behavior, 13*, 211 - 225.

Swinnen, S. P. (2002). Intermanual coordination: from behavioural principles to neural-network interactions. *Nature Reviews Neuroscience*, *3*(5), 348-359.

Swinnen, S. P., & Wenderoth, N. (2004). Two hands, one brain: cognitive neuroscience of bimanual skill. *Trends in cognitive sciences*, *8*(1), 18-25.

Tai, Y. F., Scherfler, C., Brooks, D. J., Sawamoto, N., & Castiello, U. (2004). The human premotor cortex is 'mirror' only for biological actions. *Current biology*, *14*(2), 117-120.

Talmy, L. (2000). *Toward a Cognitive Semantics. Volume I: Concept Structuring Systems*. Cambridge, MA: MIT Press.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of cognitive neuroscience*, *17*(2), 273-281.

Thompson, R. L., Vinson, D. P., & Vigliocco, G. (2009). The link between form and meaning in American Sign Language: Lexical processing effects. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *35*, 550–557.

Thompson, R. L., Vinson, D. P., & Vigliocco, G. (2010). The link between form and meaning in British Sign Language: Effects of iconicity for phonological decisions. *Journal of Experimental Psychology: Language, Memory, and Cognition, 36*, 1017–1027.

Thorpe, S. J., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature, 381*, 520-522.

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., & Nori, F. (2008) An open-source simulator for cognitive robotics research: The prototype of the iCub humanoid robot simulator. *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop* (PerMIS08).

Tipper, S. P., Howard, L. A., & Jackson, S. R. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition*, *4*(1), 1-38.

Tipper, S. P., Paul, M. A., & Hayes, A. E. (2006). Vision-for-action: The effects of object property discrimination and action state on affordance compatibility effects. *Psychonomic Bulletin & Review*, *13*(3), 493-498.

Torralbo, A., Santiago, J., & Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science*, *30*(4), 745-757.

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

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*(6), 769–800.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, *116*(2), 185-203.

Turner, P. (2005). Affordance as context. *Interacting with computers*, *17*(6), 787-800.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological psychology*, *4*(3), 173-187.

Ulrich, R., & Maienborn, C. (2010). Left–right coding of past and future in language: The mental timeline during sentence processing. *Cognition*, *117*(2), 126-138.

Ulrich, R., Eikmeier, V., Vega, I., Ruiz Fernández, S., Alex-Ruf, S., & Maienborn, C. (2012). With the past behind and the future ahead: Back-to-front representation of past and future sentences. *Memory & Cognition, 40*(3), 483–495.

Vainio, L., Ellis, R., & Tucker, M. (2007). The role of visuospatial attention in action priming. *The Quarterly Journal of Experimental Psychology*, *60*, 241-261.

Vainio, L., Ellis, R., Tucker, M., & Symes, E. (2006). Manual Asymmetries In Visually Primed Grasping. *Experimental Brain Research, 173*(3), 395–406.

Vainio, L., Symes, E., Ellis, R., Tucker, M., & Ottoboni, G. (2008). On the relations between action planning, object identification, and motor representations of observed actions and objects. *Cognition, 108*, 444–465.

van Dantzig, S. (2009). *Mind the body: grounding conceptual knowledge in perception and action*. Erasmus Universiteit Rotterdam.

Van Elk, M., Van Schie, H. T., & Bekkering, H. (2008). Conceptual knowledge for understanding other's actions is organized primarily around action goals. *Experimental Brain Research*, *189*(1), 99-107.

Vico, G. B. (1744/1953). *La Scienza Nova*. Bari: Laterza.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219-247.

Vogt, S., Taylor, P., & Hopkins, B. (2003). Visuomotor priming by pictures of hand postures: perspective matters. *Neuropsychologia*, *41*(8), 941-951.

Wauters, L. N., Tellings, A. E., Van Bon, W. H., & Van Haaften, A. W. (2003). Mode of acquisition of word meanings: The viability of a theoretical construct. *Applied Psycholinguistics*, *24*(03), 385-406.

Wenger, E. (1998) *Communities Of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: the attention-grabbing power of approach-and avoidance-related social information. *Journal of personality and social psychology*, *78*(6), 1024-1037.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language, 93*, 10–19.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68*(1), 77-94.

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, *29*(5), 719-736.

Wiemer-Hastings, K., Krug, J., & Xu, X. (2001). Imagery, context availability, contextual constraint and abstractness. In J. D. Moore and K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.* Edinburgh, Scotland, August 1-4, 2001 (p. 1106-1111). Mahwah, NJ: Lawrence Erlbaum Associates.

Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-specific representations of action verbs neural evidence from right-and left-handers. *Psychological Science*, *21*(1), 67-74.

Witt, J. K., Kemmerer, D., Linkenauger, S. A., & Culham, J. (2010). A functional role for motor simulation in identifying tools. *Psychological Science*,*21*(9), 1215-1219.

Wittgenstein, L. (1953/2001), *Philosophical Investigations. The German Text, with a Revised English Translation*. Oxford: Blackwell Publishing.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*(2), 173-189.

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Human brain mapping*, *32*(12), 2141-2150.

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2012). The context-object-manipulation triad: cross talk during action perception revealed by fMRI. *Journal of Cognitive Neuroscience, 24*(7), 1548-59.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (can) prime furiously. *Psychological science*, *23*(4), 364-369.

Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *The American journal of psychology*, 349-384.

Young Yoon, E., & Humphreys, G. W. (2007). Dissociative effects of viewpoint and semantic priming on action and semantic decisions: Evidence for dual routes to action from vision. *The Quarterly Journal of Experimental Psychology*, *60*(4), 601-623.

Yoon, E. Y., Humphreys, G. W., & Riddoch, M. J. (2010). The paired-object affordance effect. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 812.

Young, G. (2006). Are different affordances subserved by different neural pathways? *Brain and cognition*, *62*(2), 134-142.

Zipf, G. K. (1949) *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison–Wesley.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of learning and motivation*, *44*, 35-62.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954-958.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, *28*(4), 611-619.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological science*, *13*(2), 168-171.

APPENDIX

CHAPTER 4

SUPPLEMENTARY MATERIALS

	TARGET OBJECT	2nd OBJECT	RELATION
1	Spoon	Plate	Functional
2	Fork	Strawberries	Functional
3	Tennis racquet	Ball	Functional
4	Tea-strainer	Mug	Functional
5	Knife	Butter	Functional
6	Screw-driver	Screw	Functional
7	Bottle opener	Bottle	Functional
8	Paint-brush	Can of paint	Functional
9	Scissors	Paper	Functional
10	USB	Laptop	Functional
11	Wrench	Nut/Bolt	Functional
12	Syringe	Needle medicine	Functional
13	Iron	Crumpled Shirt	Functional
14	Coffee Pot	Mug	Functional
15	Hammer	Nail	Functional
16	Coffee measurer/scoop	Coffee	Functional
17	Shaving razor	Foam	Functional
18	Skillet	Vegetable	Functional
19	Ice cream scooper	Ice cream	Functional
20	Spatula	Skillet	Functional
21	Marker	Paper	Functional
22	Can opener	Can of vegetables	Functional
23	Wine bottle opener	Wine bottle	Functional
24	Spoon	Screw	No Relation
25	Fork	Tennis ball	No Relation
26	Tennis racquet	Book bag	No Relation
27	Tea-strainer	Laptop	No Relation
28	Knife	Nail	No Relation
29	Screw-driver	Paper	No Relation
30	Bottle opener	Mouse	No Relation
31	Paint-brush	Lamp	No Relation
32	Scissors	Bottle	No Relation
33	USB	Tennis ball	No Relation
34	Wrench	Paper	No Relation

35	Syringe	Spoon	No Relation
36	Iron	Tea cup	No Relation
37	Coffee Pot	Marker	No Relation
38	Hammer	Crumpled Shirt	No Relation
39	Coffee measurer/scoop	Paper	No Relation
40	Shaving razor	Paper	No Relation
41	Skillet	Pen	No Relation
42	Ice cream scooper	Paper	No Relation
43	Spatula	Mouse	No Relation
44	Marker	USB	No Relation
45	Can opener	USB	No Relation
46	Wine bottle opener	Mouse	No Relation
47	Spoon	Pot	Spatial
48	Fork	Glass	Spatial
49	Tennis racquet	Shoe	Spatial
50	Tea-strainer	Sugar	Spatial
51	Knife	Pot	Spatial
52	Screw-driver	Flash-light	Spatial
53	Bottle opener	Coffee cup	Spatial
54	Paint-brush	Marker	Spatial
55	Scissors	Stapler	Spatial
56	USB	Mouse	Spatial
57	Wrench	Catalogue	Spatial
58	Syringe	Cotton swab	Spatial
59	Iron	Detergent	Spatial
60	Coffee Pot	Milk	Spatial
61	Hammer	Wood	Spatial
62	Coffee measurer/scoop	Coffee cup	Spatial
63	Shaving razor	Soap	Spatial
64	Skillet	Plate	Spatial
65	Ice cream scooper	Plate	Spatial
66	Spatula	Fork	Spatial
67	Marker	Scissors	Spatial
68	Can opener	Plate	Spatial
69	Wine bottle opener	Glass	Spatial

Table a. Target-objects list with the corresponding paired objects

CHAPTER 5

SUPPLEMENTARY MATERIALS

Normative study

The 40 critical scene pictures were assessed for their ability to evoke a particular gesture, either pinch or clench, when the subject was explicitly asked to pantomime a gesture appropriate for the conflict object as depicted in the scene. 16 additional healthy subjects (5 males) volunteered for participation in the norming experiment. On each trial, the scene picture appeared on the screen. After 1000 ms, a red box appeared around the conflict item in each photograph lasting 750 ms, and then disappearing from the photograph. Participants task was to pantomime with their left hand how they would interact with the highlighted object in that particular context. Participants were always asked to respond with their left hand while their right hand in order to allow future comparisons with left hemisphere stroke patients. Thus, as left hemisphere stroke patients frequently have reduced right arm mobility, the mobility of the right limb was limited with an arm sling. Participants were given explicit instructions to take note of the context as it would inform them about how they might interact with that object in real life. They were also instructed to respond as quickly as possible so as to reflect the most natural and immediate gesture evoked by the object in that environment. A response was coded as a clench if pantomimed contact with an imaginary object included the palm of the hand, had a rounded aperture, and used more than 3 active fingers. A response was coded as a pinch if the subject pantomimed contact with the object with only the thumb and the index finger or only the thumb, index, and middle finger. It was also coded as a pinch if the subject gestured with more than 3 fingers but pantomimed contacting the object with the tips of the fingers and/or formed a hand posture with a flat aperture. Responses were recorded by video camera and coded offline by one of 2 experimenters, who demonstrated 90% inter-rater reliability. To prevent coding biases experimenters did not have knowledge of the scene or object to which the subject was pantomiming. The norming experiment contained 40 trials, presented randomly. There were two scenes for each of 20 conflict objects: one scene depicted the conflict object in a USE context and the other scene depicted the conflict object in a MOVE context. Norming data confirmed that each conflict object received more pinch gestures in the USE compared to the MOVE context while it received more clench gestures in the MOVE compared to the USE context. Overall, in the MOVE context there were 175 clench and 142 pinch gestures, whereas in the USE context there were 103 clench and 210 pinch gestures.

CONFLICT ITEM	**MOVE SCENE**	**USE SCENE**
Book	On bookshelf	Open on desk
Playing cards	Stacked in drawer	Stacked on card table
Binder clip	In supply drawer	Affixed to paper stack on desk
Cookie jar	On pantry shelf	On counter, lid slightly ajar
Corkscrew	In kitchen drawer	In corked wine bottle
Cheese grater	In kitchen drawer	In bowl on kitchen counter
Jewelry box	On bathroom shelf	On bathroom counter, slightly ajar
Keys	In desk drawer	Inserted to desk lock
Lamp	On supply shelf	On desk, angled toward magazine
Pencil sharpener	On bathroom shelf	On desk, pencil inserted
Pin cushion	On bathroom shelf	On bathroom counter
Post-it	On bathroom shelf	On desk, top note written on
Pot lid	On dish rack	On pot on stove
Soda can	On pantry shelf	On counter next to glass
Measuring spoons	In kitchen drawer	On kitchen counter, inserted in baking soda
Tape dispenser	In desk drawer	On desk, with gift wrapping supplies
Kitchen timer	In kitchen drawer	On kitchen counter
Tissue	On bathroom shelf	On bathroom counter, slightly ajar
Toilet paper	On bathroom shelf	On roll next to toilet
Tupperware	On dish rack	On kitchen counter with food inside, slightly ajar

Table a. Conflict item list with their corresponding MOVE and USE scenes

CHAPTER 9

SUPPLEMENTARY MATERIALS

WORDS PAIRS	
Jagged Names	Rounded Names
KIKI	BOBA
KUTI	BAMA
TITI	GOGA
TUKITI	MABUMA

Table 1. Word pairs

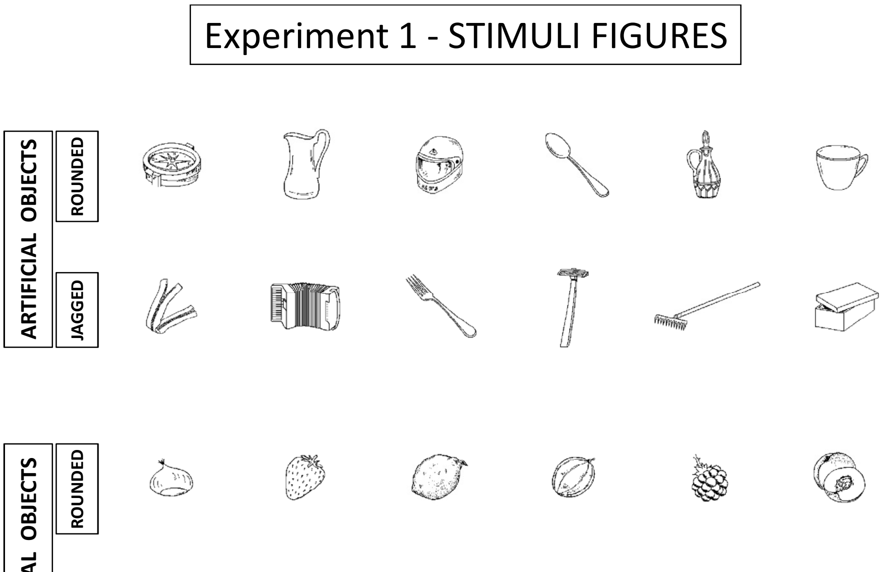


Figure a. Experiment 1 - Stimuli figures

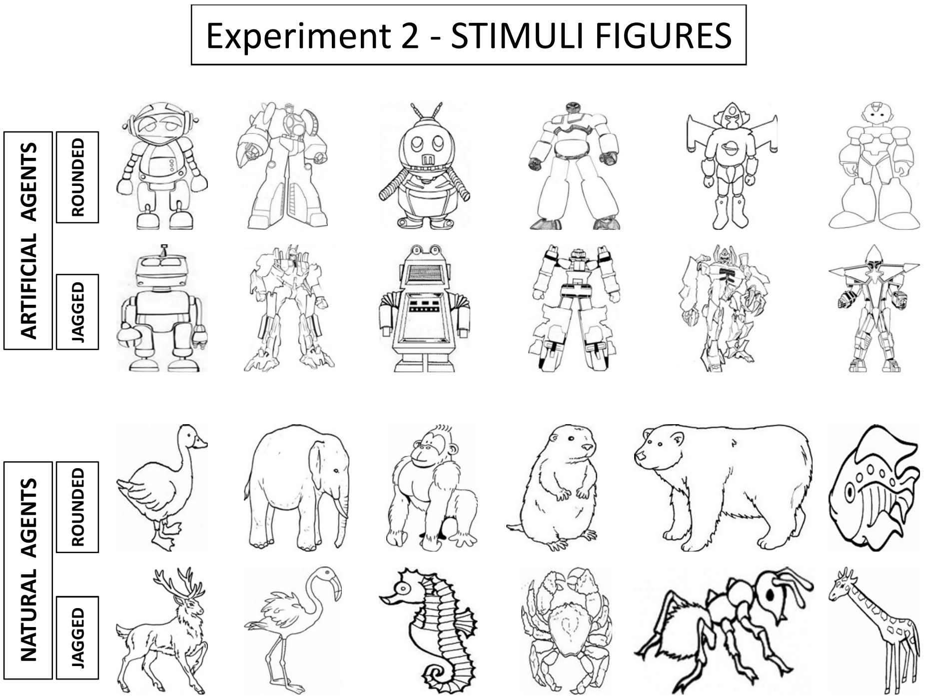


Figure b. Experiment 2 - Stimuli figures

Acknowledgements

This experimental work was supported by a Full scholarship granted by MIUR - Italian Board of Education, University and Research, and in part by the PRIN 2008-2010 project and by the FP7 project ROSSI (Emergence of communication in RObots through Sensorimotor and Social Interaction), Grant agreement no: 216125 to Professor Anna M. Borghi.

Thanks to Anna M. Borghi, Roberto Nicoletti, Julio Santiago and to all the people of EmCo Lab (Bologna) and Grounded Cognition Lab (Granada) for useful discussions on the hypotheses and results of this thesis. Thanks also to all the students of the University of Bologna and Granada for participating in the reported experiments.

I would like to thank Anna M. Borghi and Roberto Nicoletti for supervising me during my PhD, and for providing the funding and environment to perform the research reported in my thesis. I would like to thank also Julio Santiago for supervising my visiting period at the University of Granada, and for providing the environment to perform the study reported in chapter 8.

Finally, special thanks to Anna M. Borghi for her constant support throughout the years and for transimitting me the enthusiasm of doing research in cognitive science.