Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Science Cognition and Technology

Ciclo XXIX

Settore Concorsuale di afferenza: 11/E1

Settore Scientifico disciplinare: M-PSI/01

TITOLO TESI

From Amodal to Grounded to Hybrid Accounts of Knowledge: New Evidence from the Investigation of the Modality-Switch Effect

Presentata da: Elisa Scerrati

Coordinatore Dottorato	**Relatore**
Marco Beretta	**Renata Galatolo**

Esame finale anno 2017

Acknowledgements

I wish to acknowledge the support of those I have collaborated with at the University of Bologna during the past three years: Giulia Baroni, Anna Maria Borghi, Stefania D’Ascenzo, Renata Galatolo, Luisa Lugli, and Roberto Nicoletti.

I would also like to thank people I have worked with at City University London: Albert Barque-Duran, James Hampton, Emmanuel Pothos, and James Yearsley.

I am grateful to reviewers of the papers that grew out of my dissertation research: Markus Kiefer, Ken McRae, Diane Pecher and an anonymous reviewer. Thanks also to Rutvik Desai for the insightful arguments he mentioned and discussed with me at the Emergent Meaning Workshop in Bethlehem, USA.

Thanks to Marco Beretta, Giovanna Corsi, and Giuliano Pancaldi from the University of Bologna and to all my colleagues at both the University of Bologna and City University London.

Finally, this dissertation is dedicated to Nicoletta Caramelli, who first introduced me to the wonders of the mind.

To my parents and Mickey for their support and endless love.

TABLE OF CONTENTS

PART 1: THE DEBATE ON THE FORMAT OF CONCEPTS:
THEORETICAL POSITIONS

PART 2: THE DEBATE ON THE FORMAT OF CONCEPTS:
EXPERIMENTAL EVIDENCE

IV. STUDY 1 EXPERIMENTS 1 & 2

The Modality-Switch Effect: Visually and Aurally Presented Prime Sentences Activate Our Senses

METHOD

V. STUDY 2 EXPERIMENTS 3 & 4

The Multilevel Modality-Switch Effect: What Happens when We See the Bees Buzzing and Hear the Diamonds Glistening

METHOD

INTRODUCTION

The debate on the representational format of concepts is more alive than ever as witnessed by the recent cutting-edge 15 articles in the special issue of the Psychonomic Bulletin and Review entitled The Representation of Concepts. Contributors dispute on the format of concepts. An example[1] will help me to pinpoint what is meant by “representational format of concepts”. Consider the binary and decimal coding of numbers. Within the decimal numeral system, the number “10” is represented as 10, whereas within the binary numeral system the same number is represented as 1010. These numeral systems exploit different representational codes or formats to encode the same content. So do the contending theories in this debate. Amodal theorists argue that concepts are represented in an amodal symbolic semantic system detached from the sensory and motor systems, whereas supporters of the grounded accounts of knowledge claim that concepts are represented in several different modality-specific brain areas.

In Part 1 of this dissertation I illustrate each view in detail and then discuss viable hybrid models of knowledge that combine aspects of both classes of theories. Part 2 is entirely devoted to testing predictions coming from amodal and grounded accounts of knowledge. Specifically, it is aimed at verifying the scope of the assumption that modality-specific representations underlie concepts and conceptual processing through the investigation of

---

[1] The example was taken from Machery (2016).

the Modality-Switch Effect, a cost for performance in terms of speed and accuracy occurring when two different sensory modality properties for concepts alternate (e.g., leaves *rustle* - diamond *glistens*) compared to when the same sensory modality properties are being presented (e.g., leaves *rustle* – bee *buzzes*).

PART 1

THE DEBATE ON THE FORMAT OF CONCEPTS: THEORETICAL POSITIONS

I. Amodal symbolic accounts of knowledge

According to amodal symbolic accounts of knowledge, the mind is a symbol system and cognition is symbols manipulation. Semantic and conceptual[2] processes are being attributed to a dedicated symbolic level, also known as the mental level (i.e., the mind), that is a rule-governed functional level independent of the physical substrate through which it realizes its functions. Such an autonomous symbolic level would thus be functionally detached from sensory and motor systems. In order to be manipulated, sensory and motor information coming from the environment would need to be transduced into a different format that is symbolic, amodal, and arbitrary.

1. The symbolic model of mind. The *symbolic model of mind*, also known as the *representational* or *computational theory*[3] *of mind* (Fodor, 1975; 1987; Newell, 1980; Pylyshyn, 1980, 1984) describes the mind as being a symbol system or a system of representation. Such a system is characterized by a set of arbitrary physical tokens manipulated on the basis of explicit rules. Phisical tokens can be atomic symbol tokens (e.g., *ravens*) or composite symbol-tokens strings (e.g., *feathered ravens*).

---

[2] Much of the literature use the terms "semantic" and "conceptual" as synonyms. I do the same here, though it is worth mentioning that Murphy (2002, p. 385) proposes an interesting view on the relation between meanings and concepts known as the *conceptual view*. On this view, a word gets its significance by being connected to a concept. In other words, the meaning is built out of concepts.

[3] It is worth noting that while a "theory" attempts to explain phenomena, for example suggesting the mechanisms involved, a "model" is aimed at representing phenomena, for example describing the components and operations involved.

These symbols are manipulated on the basis of their shape[4] (not their meaning). Symbols or representations of the system have a combinatorial syntax and semantics. That is, structurally complex (molecular) representations are systematically built up out of structurally simple (atomic) constituents, and the semantic content of a complex representation is a function of the semantic content of its atomic constituents together with their mode of combination (see also Fodor & Pylyshyn, 1988). The symbol system and all its parts are semantically interpretable, namely the syntax can be systematically assigned a meaning, for example as describing states of affairs (e.g., *John loves Mary*).

Most of the arguments supporting the *representational theory of mind* derive their strength from their ability to explain certain empirical phenomena such as the productivity and systematicity of thought and thinking. Productivity refers to the ability of building and understanding a potentially infinite number of linguistic expressions starting from a finite number of linguistic elements. Systematicity refers to the ability of building and understanding recurring defined and predictable patterns such as *John loves Mary - Mary loves John*.

For the purposes of the debate on the format of concepts, it is worth emphasizing that symbols in a symbol system are conceived as amodal, that is, they are inherently nonperceptual. Amodal symbolic semantic systems assume that cognitive and

---

[4] For this reason, symbol systems are formal systems. Formalisms such as predicate calculus, probability theory, and programming languages inspired many new representational languages in cognitive science (e.g., feature lists, frames, schemata, connectionism, etc.) as we shall see later in this chapter.

perceptual information constitute separate systems that work following different rules and use different representational formats. Figure 1 illustrates this assumption.

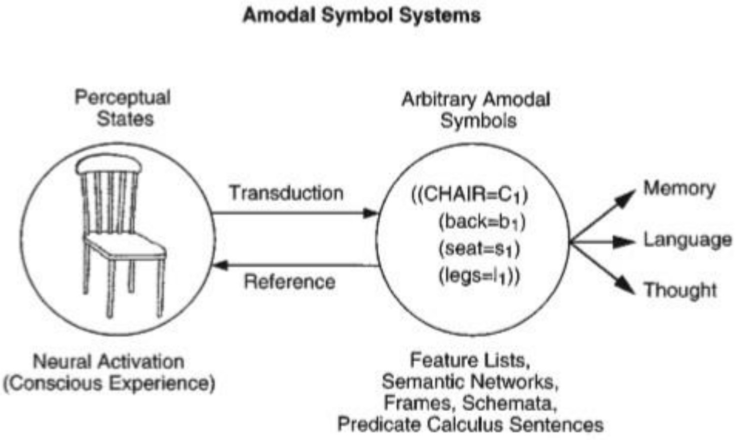

Fig. 1 The basic assumption underlying amodal symbol systems is that perceptual information is transduced into a new representational format that is completely amodal. As a result, the internal structure of these symbols is unrelated to the perceptual information that produced them and arbitrary, conventional associations establish reference.

Reference: Barsalou [1999]

Perceptual information coming from the environment is captured by sensory-motor systems and transduced into a completely new representation language that is inherently nonperceptual: the amodal system. Amodal symbols would become organized into larger representational structures, such as feature lists (Smith and Medin, 1981; Smith, Shoben, & Rips, 1974), frames (Barsalou & Hale, 1993), semantic networks (Collins & Loftus, 1975; Collins & Quillian, 1969), etc.. Each of these structures constitute a fully functional symbolic system with a combinatorial syntax and semantics that supports all of the higher cognitive functions, including memory, knowledge, language, and thought. Symbols in these systems are amodal because they do not correspond to the perceptual states that produced them. The amodal symbols that represent the colors of objects, for example, would be located in a completely different neural system from the one designated for perception of colors.

As a consequence of being amodal, symbols in a symbol system are arbitrarily linked to perceptual information. As Barsalou (1999, pp. 578-579) explains: “Similarly to how words typically have arbitrary relations to entities in the world, amodal symbols have arbitrary relations to perceptual states. Just as the word “chair” has no systematic similarity to physical chairs, the amodal symbol for chair has no systematic similarity to perceived chairs”.

Amodal symbols are usually represented as linguistic forms. In feature lists (e.g., Smith et al., 1974), words represent features. For example, for the concept *bird*, the words *paw*, *beak*, *feathers*, *wings*, *tail*, etc. represent its features. Similarly, relations, arguments, and values are represented as words in frames (Barsalou & Hale, 1993). For example, the relation *kick* involves an agent whose argument might be the word *kid* and an object whose argument might be the word *ball*. Although being usually represented as linguistic forms, amodal symbols’ content are not words. Rather, it is assumed that close amodal counterparts of words constitute the content of amodal symbols. However, as Barsalou (1999, p. 579) points out, “symbolic thought is assumed to be analogous in many important ways to language”.

Neuropsychological studies have shown results supporting the amodal format of concepts. In Semantic Dementia (SD), a neurodegenerative condition, a brain damage in the temporal pole and adjacent areas results in an impairment of conceptual processing (Patterson, Nestor, & Rogers, 2007). Patients with SD show “a progressive deterioration of expressive and receptive vocabulary and of knowledge about the properties of everyday objects” (Patterson et al., 2007, p. 978). Degraded knowledge extends across all conceptual domains (including animal, tools etc.) and conceptual modalities (e.g., visual, auditory, action-related etc.). Interestingly, it has been shown that atypical instances of a category are more impaired than typical ones (Rogers, Lambon Ralph, Garrad, Bozeat, McClelland, Hodges, & Patterson, 2004). Thus, for example, knowledge of *penguins* or *ducks* is more degraded than knowledge of *sparrows* or *robins*, the latter being more typical instances of the category *birds* than the former. It is important to point out that while typical instances of a category tend to share many properties with other exemplars of the same category, atypical instances have just few properties in common with other category members. For example, typical instances of *birds* such as *sparrow* and *robin* share many properties such as *feathers*, *nest in trees*, and *feed on the ground*. By contrast, atypical instances of *birds* such as *penguins* have idiosyncratic and distinguishing properties such as *being flightless* and *aquatic*. Rogers et al. (2004) pointed out that as damage accumulates, the system becomes increasingly unable to retrieve idiosyncratic and distinguishing information about objects because distinctive properties of individual items are not shared by other category members. Thus, while it is likely that small distortion of the penguin representation will prevent the retrieval of the penguin's specific name or, more generally, to identify a certain entity as a penguin, small distortion of the robin representation could still allow to retrieve the robin's name or to identify a certain entity as a robin given that many of its properties are shared by other category members. If one of these property is damaged, other properties of the concept's schema can help retrieve it, or stand in for it. Sharing a high number of properties ensures a considerable bundle of relations between concepts. Relations among concepts belonging to the same category ensure that semantically related items (for example, various different birds) are coded with similar patterns across neurons. Therefore, the more pronounced impairment for atypical rather than typical instances of a category would suggest a sensitivity of SD to abstract relations between concepts (Rogers et al., 2004; Patterson et al., 2007). Abstract relations or semantic generalizations are believed to require a single amodal hub that would be located in the anterior temporal lobe (ATL). Indeed, early symptomps of SD seem to emerge in conjunction with lesions in the temporal poles. In addition, ATL has been shown to be functionally relevant for conceptual processing in healthy people: transcranial magnetic stimulation (TMS) of anterior temporal areas brought about a deteriorated performance in semantic tasks for pictures and words similar to the impairment seen in SD (Pobric, Jefferies, & Lambon Ralph, 2010).

In sum, Semantic Dementia has been shown to be sensitive to structural relations between concepts. Relational knowledge is best explained throught the existence of a single amodal hub. It has been proposed that areas within the anterior temporal cortex are the neural substrate of an amodal conceptual system (Patterson et al., 2007; McClelland & Rogers, 2003; Rogers et al., 2004). I will discuss this issue further in chapter III when presenting hybrid models of cognition.

I will now turn to discuss an important distinction within amodal symbolic systems: the local versus distributed distinction.

2. Localist versus distributed systems. According to a localist amodal account of conceptual representations, concepts are represented as nodes in a semantic network (Collins & Loftus, 1975; Collins & Quillian, 1969). Figure 2 depicts a schematic illustration of such a network.

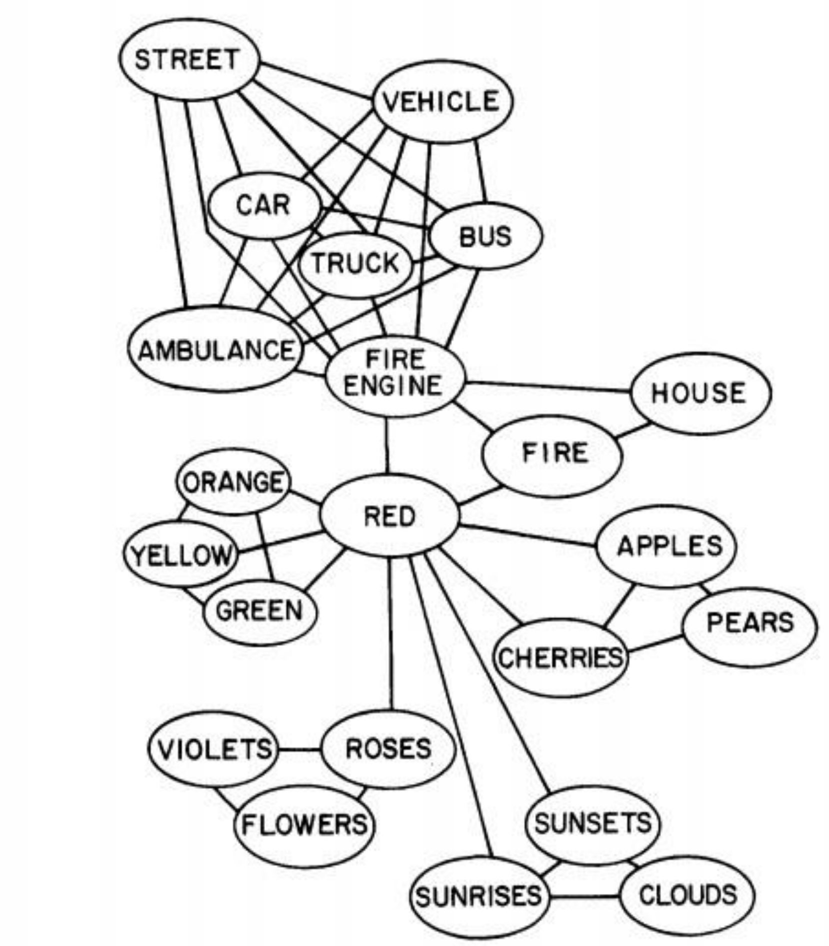

Fig. 2 A schematic representation of concepts in a semantic network (shorter line represents greater relatedness between concepts).

Reference: Collins & Loftus [1975]

Each node is related to a number of other nodes in the network on the basis of different types of relations (taxonomic: e.g., *car*-

vechicles; perceptual: e.g., *apples-red*; thematic or situational: e.g., *car-street*).

The localist view on concepts implies that the organization of conceptual knowledge in our minds reflects statistical information available in the environment. The network of interconnected nodes can be conceived as a large representational structure that provides propositional knowledge about a concept (e.g., that apples are red) in an explicit symbolic fashion. To illustrate, each node is identified by a label (i.e., a word) which is arbitrarily related to a specific content. Each concept is a specific node, which is distinct at both the neuroanatomical and the functional level from sensory and motor representations.

Neuroanatomically speaking, localist views of conceptual knowledge assume that concepts are single neuronal units (Barlow, 1972). Such *grandmother cell*[5] assumption, found support from recent work in neuropsychology. Studies using single cell recordings in patients found neurons firing in a highly specific manner to single objects, faces, words or persons, suggesting an at least plausible localist coding of information by grandmother cells (for a review see Bowers, 2009). Neurons in the lateral temporal lobe preferentially fired when single, specific words were presented (Creutzfeldt, Ojemann, & Lettich, 1989) although it is still not very clear whether neural responses

---

[5] The label “grandmother cell” designates a hypothetical neuron that represents a complex but specific concept or object that activates when a person sees, hears, or otherwise sensibly discriminates a specific entity such as his or her grandmother, that is where the label comes from.

were driven by perceptual or conceptual variables (see Kiefer & Pulvermüller, 2012). Similarly, neurons in the medial temporal lobe were selectively activated by highly different pictures of a given person, landmark or object and in some cases even by person names[6] (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Although such data seem quite consistent with the *grandmother cell* assumption, Kiefer and Pulvermüller (2012, p. 813) pointed out that in the above mentioned-study "the specificity or the individual cells' response patterns to stimulus type can only be compared to a relatively small number of control stimuli and comparison cells, so that the bold statement of absolute specificity can never be supported convincingly".

Although localist concepts are largely meant to be single neuronal units, localist representations do not necessarily imply a one-neuron-one-concept correspondence. On the contrary they may consist of larger neurons populations (Bowers, 2009). For example, it has been argued that a cell assembly can act as one single functional unit, it can have an activation threshold, and it can be activated as a whole when this threshold is reached (Garagnani, Wennekers, & Pulvermüller, 2008; Wennekers, Garagnani, & Pulvermüller, 2006). The *cell assembly* assumption allows to keep the critical aspects of localist models unchanged in the context of distributed neural networks.

---

[6] For instance, one neuron fired not only when a range of quite different pictures of Halle Barry's face were shown, but also to her written name, while it did not fire when other different stimuli were shown (Quiroga et al., 2005).

Feature list models (e.g., Smith et al., 1974) can be considered as an early example of distributed theories. In feature list models, a concept consists of a set of semantic features that code its basic different characteristics in an explicit fashion. For instance, the concept *apple* is constituted by the features *red*, *round*, *smooth*, *juicy*, *sweet*, *has stalk*, and so on. Subsequent models based on distributed theories such as the parallel distributed processing (PDP) or connectionist framework of cognition (see Rumelhart & McClelland, 1986), conceived concepts as dynamic patterns of activity in a multilayered network of units with weighted positive and negative interconnections (see also McClelland & Rogers, 2003; Tyler & Moss, 2001). Thus, in these models conceptual knowledge is not explicitly represented in form of symbolic features or single nodes. Rather, it is represented as propagation of activation among connected processing units in the network. The connection weights or strengths between the processing units within the network are learned through exposure and feedback, that is, they are shaped by experience and adjusted according to new inputs through a mechanism of backpropagation.

Both feature list and distributed network models assume that conceptual knowledge is represented in an amodal format within a unitary conceptual system that stores all kind of information independently of knowledge modality (e.g., visual, auditory, action-related, etc.) or category (e.g., animals or tools). Such a unitary conceptual system is assumed to be distinct from the perception and action brain systems. However, it is worth noting that feature list and PDP models can also work in the context of modality-specific conceptual representations as shown by a number of studies (Farah & McClelland, 1991; Plaut, 2002; Pulvermüller, 1999; Simmons & Barsalou, 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004) some of which will be discussed in chapter III.

Support for distributed systems comes from empirical evidence. The pattern of deficits in neurogenerative diseases such as Semantic Dementia (SD) and Alzheimer Disease (AD) can be much better accounted for by distributed than local representations at both the functional and neuroanatomical level. For instance, in SD knowledge of a single concept is not entirely impaired as would be predicted by a *grandmother cell* hypothesis. On the contrary, what has been observed is a progressive degradation of knowledge beginning with specific properties of an object concept (e.g., *doves are white*) that spreads to more general central properties shared by many exemplars (e.g., *doves have wings*; Rogers et al., 2004). Similarly, superordinate information (e.g., *canary is an animal*) is typically relatively preserved in SD, whereas more specific conceptual information (e.g., *canary is a bird*) suffers from a more severe impairment (Hodges, Graham, & Patterson, 1995; Rogers et al., 2004).

Moreover, neurophysiological studies show that many different parts of the brain are involved during conceptual tasks (e.g., Martin, 2007; Pulvermüller, Lutzenberger, & Preissl, 2009; Kiefer, Sim, Hernberger, Grothe, Hoenig, 2008) suggesting a distributed system is more plausible. In addition, it has been observed that the activation pattern in sensory and motor areas varies as a function of the task context (Hoenig, Sim, Bochev, Hernberger, & Kiefer, 2008). Further evidence coming from behavioral studies confirm this result (Barclay, Bransford, Franks, MCCarrell, Nitsch, 1974; Barsalou, 1982) as we shall see in the next section.

3. Stable versus flexible representations. The main and most important difference between localist and distributed views of concepts is conceptual flexibility (for a recent review see Yee & Thompson-Schill, 2016). That is, while localist concepts are conceived as stable mental knowledge entities that are situationally invariant, distributed views allows for the contribution of different units to the same concept. Moreover, different units are differently activated as a function of the context in which the concept is processed. For example, while on the localist view the meaning of *apple* is assumed to be the same across contexts, on the distributed view the same meaning is assumed to vary whether the apple is peeled or unpeeled, ripe or unripe, painted or cooked, etc..

As pointed out by Kiefer and Pulvermüller (2012, p. 807), "the stable-flexible distinction has its roots in modern analytical philosophy and linguistics". Theoretical positions focusing on normative aspects of meaning were concerned on whether words carry a core meaning, that is, a stable concept which is invariantly accessed each time the word is used. Processing of a particular concept would then be performed by an invariant pattern of activated brain areas irrespective of task demands. However, words such as *game* clearly shows that this cannot be so. Under the label *game* a number of instances is encompassed that do not share a fixed set of conceptual features. For example, chess, video games, football are all very different games. However, they are associated because of family resemblance. Wittgenstein (1953) pointed out that the various games relate to each other as the members of a big family do[7], with some pairs exhibiting great similarities (e.g., football-volleyball) while others varying considerably (e.g., football-chess).

A different yet relevant phenomenon for the stable-flexible distinction is lexical ambiguity. Context can affect the way in which an ambiguous word is encoded. Consider the word *jam*. In the compound *strawberry jam* it indicates the fruit conserve and it is thus related to the semantic domain of food, whereas in the compound *traffic jam* it indicates the vehicle congestion and it is thus related to the semantic domain of vehicles.

By focusing on relatively invariant features of words, normative theories of meaning also neglected the variability inherent in a word's interpretation. Indeed, uses and meanings of words are manifold. Under different circumstances, the same word can be interpreted very differently. Consider, for example, the way in which one's interpretation of the unambiguous word *piano* is affected by verb selection in the following five sentences: *the man lifted the piano*; *the man tuned the piano*; *the man smashed the piano*; *the man sat on the piano*; *the man*

---

[7] For this reason he calls such relationship "family resemblance".

photographed the piano (Barclay et al. 1974). Different properties of *piano* are differently emphasized as a function of the event described in each sentence.

The phenomena of family resemblance, lexical ambiguity and variability of the word's interpretation can be best explained by distributed theories assuming conceptual flexibility. According to such theories, concepts (i.e., word meanings) are constituted of dynamically recruited features depending on the context (Barsalou, 1982; Kiefer, 2005). The activation of features contributing to a concept differs on the basis of weighting mechanisms and of contextual constraints. By weighting mechanisms is meant the contextually determined relevance of a word's semantic properties. For example, given the event *lifting* the heaviness of a piano is relevant, while the sound it can make is not. By contextual constraint is meant the interaction between linguistic knowledge and more general world knowledge. For instance, if one knows that a piano is being smashed then he also knows that no one will be able to play that piano before it will be fixed.

Behavioral literature shows that the contribution of features to a concept are context dependent. Barclay et al., (1974) produced evidence that the interpretation of familiar, unambiguous words varied with their sentential contexts. Cues mentioning some property of the target word's referent (e.g., piano) induced a better recall when the information they expressed was relevant (e.g., something heavy), rather than irrelevant (e.g., something with a nice sound), to the events described by previously shown sentences (e.g., the man lifted the piano). They replicated their result in a series of subsequent experiments with different experimental settings and materials.

Barsalou (1982, Experiment 1) demonstrated that the speed of property verification was affected by the context for some properties but not for others. Specifically, context-independent properties are those shared by common categories such as *birds*, *furniture*, *vegetables*. For example, the concepts *sparrow* and *robin* share the common property of *flight*. Shared properties of common categories were shown to be equally activated with and without a given context. On the contrary, context-dependent properties are those shared by ad hoc categories such as *things that float*, *things that have a smell*, *things that can be thrown*. For example, the concepts *basketball* and *log* share the common property of *floating*. Shared properties of ad hoc categories were shown to be normally inactive and got activated only when there was an available context. Barsalou (1982, Experiment 2) also showed that the similarity of two concepts was not rated as increased by presenting a context relevant to common categories, whereas the same measure increased when presenting a context relevant to ad hoc categories. For example, the similarity of the pair of concepts *robin-eagle* did not increase when the context word was *birds*, while the similarity of the pair *record album-necklace* did increase when the context was *possible gifts*.

The notion of flexible concepts was further tested in a combined functional magnetic resonance imaging (fMRI) and event-related potentials (ERP) study (Hoenig et al., 2008). Participants performed verifications of two property types (i.e., visual, action-related) for words referring to different categories, namely artefacts and natural objects. Functional imaging predominantly revealed cross-over interactions between category and property type in visual, motor and motion-related brain areas indicating that access to conceptual knowledge is strongly influenced by the type of property (visual, action-related). Activation in these modality-specific brain areas was increased when non-dominant conceptual features (i.e., visual features for artefacts and action-related features for natural kinds) had to be verified. ERPs in turn indicated that these cross-over interactions between category and property type emerged as early as 116 msec after stimulus onset suggesting that they reflect rapid access to conceptual features rather than post-conceptual processing. These results foster the hypothesis that concepts are flexible mental entities. Following this evidence, concepts and corresponding word meanings as well as their neurobiological underpinnings should therefore be viewed as context-dependent. Therefore, the use of a concept in different situations can be modeled as the context-specific firing of cell assemblies, which is constrained by both established connections between neurons that constitute conceptual long-term memory traces and the context-dependent influence, which primes different sets of neural populations (Hoenig et al., 2008; Kiefer, 2005; Pulvermüller, 1999).

In sum, concepts consist of semantic features which are recruited from distributed, yet localized semantic maps in modality-specific brain regions depending on contextual constraints. I will focus on modality specific brain activation and its interpretation in chapter II while I will now introduce the main problems affecting amodal symbolic accounts of knowledge.

4. The symbol grounding problem. In a pure symbolic model the crucial connection between the symbols and their referents is missing. In other words, an autonomous symbol system is ungrounded. This is known as the symbol grounding problem. The symbol grounding problem (Searle, 1980; Harnad, 1990) refers to how amodal symbols would be mapped to perceptual states and entities in the world. Two examples will help us understand the problem. The first is Searle's "Chinese room argument", in which the symbol grounding problem is referred to as the problem of intrinsic meaning or "intentionality". According to the *computational theory of mind* (Fodor, 1975; 1987; Newell, 1980; Pylyshyn, 1980, 1984), if a computer could respond to all Chinese symbol strings it receives as input with Chinese symbol strings that are indistinguishable from the replies a real Chinese speaker would make (i.e., pass the Turing test in Chinese; see Turing, 1964) - then the computer would understand the meaning of Chinese symbols in the same sense that English people understand the meaning of English words. In response to this argument, Searle (1980) pointed out that imagining himself, who knows no Chinese, doing what the computer does (i.e., receiving the Chinese input symbols, manipulating them purely on the basis of their shape, and finally returning the Chinese output symbols) would not be understanding Chinese. Hence, neither the computer could actually understand Chinese. He argues that unlike words in our head which have intrinsic meaning, symbols in a symbol system have extrinsic meaning. That is, if we compare these symbols to the words in a book we can easily see how their meanings derive from the meanings in our head. Therefore, if the meanings of symbols in a symbol system are extrinsic, rather than intrinsic like the meanings in our heads, then they are not a viable model for the meanings in our heads. In other words, cognition cannot be just symbol manipulation.

Harnad's (1990) version of the symbol grounding problem is known as the "Chinese/Chinese dictionary-go-round". He supposes to learn Chinese as a second language with a Chinese/Chinese dictionary as the only source of information. This would entail an endlessly transition from one meaningless symbol or symbol-string (i.e., the definiens) to another (i.e., the definiendum), without understanding what anything means. Figure 3 depicts an example of a Chinese dictionary entry.

Fig. 3 Chinese dictionary entry for "zebra", which is "striped horse".

Reference: Harnad [1990]

Harnad (1990) pointed out that unlike cryptologists of ancient languages and secret codes whose successful efforts are grounded in a first language and in real world experience and knowledge, the task faced by a purely symbolic model of the mind can never be accomplished. A symbolic model of mind can never get off the symbol/symbol merry-go-round because symbol meaning is not grounded in something other than just more meaningless symbols.

The converse of the symbol grounding problem is the transduction problem, namely how amodal symbols arise in the cognitive system or, in other words, how perceptual states are mapped into amodal symbols.

As we shall see in the next chapter, these shortcomings can be overcome assuming a different format for symbols, that is, a sensory-motor or perceptual format.

II. Grounded Accounts of Knowledge

The grounded accounts of knowledge (also referred to as grounded cognition, situated cognition and embodied cognition hypothesis) are a class of psychological theories aimed at tackling the grounding problem raised initially by Searle (1980) and Harnad (1990). As we saw earlier, the grounding problem concerns the way in which amodal symbols, specifically, and cognition, more generally, are linked to the modalities, the body, and the environment. Thus, rather than studying cognitive mechanisms in isolation, grounded theories foster the study of cognitive mechanisms' relations with the contexts in which they are embedded and on which they depend.

According to Barsalou (2016, p. 1123), the label "grounded" better describes "the central focus of the general perspective by including other forms of grounding beside embodiment, such as multimodal simulation, physical situations, and social situations" (see also Barsalou 2008, 2010; Kiefer & Barsalou, 2013). Indeed, the cornerstone of the grounded approach is to understand not only how the body contribute to cognition but also how the modalities, the physical environment, and the social environment contribute to it (Barsalou, Breazeal, & Smith, 2007).

The theories of grounded cognition range from perceptual theories of concepts (Barsalou, 1999, 2003, 2008, 2016; Prinz, 2002) to cognitive linguistics theories (Lakoff & Johnson, 1980, 1999; Gibbs, 1994) to theories of situated action (Breazeal, 2002; Clark 1997; Gibson, 1979; Prinz, 1997; Steels & Brooks, 1995; Thelen & L. Smith 1994), memory (Conway, 1990, 2002; Glenberg, 1997, 2015a, 2015b; Rubin, 2006) and social simulation (Arbib, 2005; Decety & Grèzes 2006; Gallese, Keysers, & Rizzolatti, 2004; Goldman, 2006; Rizzolatti & Arbib, 1998; Rizzolatti & Craighero 2004; for other proposals within the grounded or embodied framework, see Gallese & Lakoff, 2005; Glenberg & Kaschak, 2002; Kiefer & Pulvermüller, 2011; Martin, Ungerleider, & Haxby, 2000; Pulvermüller, 2005, 2013; Zwaan, 2004, 2016). For the purposes of the present manuscript, I focus on perceptual theories of concepts, all other theories being beyond the scope of this research.

1. An antecedent: The sensory/functional theory. Neuropsychological research has shown selective impairments at the expense of specific categories of information. That is, following a stroke, a viral infection or a neurodegenerative disease such as the Alzheimer disease (AD) or Semantic Dementia (SD), people may lose knowledge of some categories while retaining knowledge of others. For example, people may selectively lose knowledge of living animate entities (i.e., animals), living inanimate entities (i.e., fruit/vegetables), conspecifics (i.e., other people), or nonliving things (i.e., vechicles). There are various different patterns of category-specific deficits[8]. Patients lose knowledge of living things, in

---

[8] Category-specific deficits are aslo known as “agnosia” or “semantic deficits”. Whereas the term “agnosia” implies that the deficit reflects damage

particular animals more often than nonliving things, such as manipulable artefacts. Sometimes patients lose knowledge of a single category, sometimes of multiple categories.

Warrington and her collaborators (Warrington & McCarthy, 1983, 1987; Warrington & Shallice, 1984) put forward a proposal to explain category-specific deficits that has had a broad impact on theoretical accounts of the organization and representation of concepts in the brain. The *sensory/functional theory* (Warrington & McCarthy, 1983, 1987; Warrington & Shallice, 1984; see also Damasio, 1989; Farah & McClelland 1991; Humphreys and Forde, 2001; McRae & Cree's, 2002) assumes that knowledge of a specific category is located near the sensory-motor areas of the brain dedicated to the perception of its instances' perceptual qualities and kind of movements. As a consequence, when a sensory-motor area is damaged, the processing of instances of the specif category (or categories) that rely on that area is impaired. Therefore, a damage to modality-specific brain systems explains category-specific deficits. According to these researchers there are high correlations between certain categories and certain modality-specific systems. Specifically, they suggested that living things such as animals and fruits/vegetables mainly depend on visual perceptual properties for their identification, whereas nonliving things such as vechicles or tools mainly depend on

---

to a particular sensory-motor modality, the label "semantic deficit" implies damage to a higher-order conceptual representation. Particular theories tend to favour one term over the other, based on their particular assumptions about the conceptual system.

functional/associative properties for their identification. Further studies along these lines (Borgo & Shallice, 2001; 2003; Cree & McRae, 2003; Crutch & Warrington, 2003; Vinson, Vigliocco, Cappa, & Siri, 2003) have emphasized the importance of different visual properties for different categories. For example, while the recognition of living things such as *animals* mainly depends on the visual property of motion, the recognition of living things such as *fruits* mainly depends on the visual property of color.

In addition, much recent neuroimaging research has largely shown different neural activations for different categories. For instance, Chao, Haxby, and Martin (1999) and Chao, Weisberg, and Martin (2002) found differential activation for animals and tools. Kanwisher, McDermott, & Chun (1997) showed neural specificity for faces. Further investigations have demonstrated the activation of specific neural areas when specific stimuli such as places (Epstein & Kanwisher, 1998; see also Bar & Aminoff, 2003), bodyparts (e.g., Downing, Jiang, Shuman, & Anwisher, 2001), and written words (e.g., Cohen, Dehaene, Naccache, Lehéricy, Dehaene-Lambertz, Hénaff, & Michel 2000; Glezer, Jiang, & Riesenhuber, 2009) were presented. Chao & Martin (2000) described regions in the dorsal visual pathway, such as posterior parietal cortex, that were differentially recruited when participants viewed manipulable objects such as tools and utensils. Also, semantic knowledge of actions involves different loci of representation in the brain than semantic knowledge of entities, specifically the frontal lobe motor-related areas (see, for example, Hickok, 2014; Kemmerer, 2015). This substantial amount of data corroborated the idea that semantic category or domain is an organizing principle in the brain.

The *sensory/functional theory* assumes the existence of a sensory store that contains conceptual content recoded from the original sensory systems. Similarly, a functional store is hypothesized that contains conceptual content recoded from the motor system. The assumption of multiple systems (sensory, functional) for the representation of knowledge lends the theory a distributed character, whereas the idea of a recoding or transduction of sensory-motor properties into new representations (which later function as stand-alone representations) is typical of the amodal accounts of knowledge. Despite being an amodal account of knowledge, the *sensory/functional theory* has proved crucial for the flourishing of grounded theories of knowledge, the latter sharing with the former the idea that categories, and the conceptual system more broadly, are organized in a modality-specific fashion.

2. The convergence zone theory. A rather different formulation of the *sensory/functional theory* is the *convergence zone theory* (CZ, Damasio, 1989; Damasio & Damasio, 1994). The theory consists of two core components: (1) systems of feature detectors in sensory-motor areas, and (2) conjunctions of modality-specific and cross-modal information in convergence zones. On this view, when an entity is perceived, it activates feature detectors in the relevant sensory-motor areas (systems of these detectors are also known as “feature maps”, see Simmons

& Barsalou, 2003). During visual processing of a cat, for example, some neurons respond to line orientations, vertices, and planar surfaces. Others respond to colour, and direction of movement. The overall pattern of activation across this hierarchically organised distributed system constitutes the visual representation of the concept *cat* in the visual system (Palmer, 1999; Zeki, 1993). Similar patterns of activation arise in other modalities (auditory, motor, etc.).

The states (i.e., patterns of activation) that arise in different sensory-motor areas are then stored in association areas. Damasio refers to these association areas as “convergence zones” and assumes that they exist at multiple hierarchical levels, that is, sensory-motor (i.e., posterior in the brain) as well as higher-level (i.e., anterior in the brain). At the sensory-motor level, CZs store patterns of activation within a particular modality. For example, CZs near visual processing areas store patterns of activation within the visual system, whereas CZs near motor processing areas store patterns of activation within the motor system. Conversely, higher-level association areas link together patterns of activation across modalities. For example, if CZs near visual processing areas store the visual form of a chair and CZs near motor processing areas store the action taken on a chair then a subset of neurons in higher level cross-modal CZs correlates the visual form of the chair with the action taken on it. Thus, subsets of neurons in higher level cross-modal CZs link together earlier conjunctions of neurons present in sensory-motor CZs.

It is assumed that convergence zones become differentially important for representing different semantic categories. For example, because humans frequently interact with tools and other man-made objects, the zone that links object shape and action might be more important for knowledge of artefacts than for knowledge of living things. Similarly, because animals move in characteristic ways, the zone that links shape to movement might acquire special salience for knowledge of animals.

Unlike other *sensory/functional theories*, which assume that conceptual content only exists in other systems that recode patterns of feature maps activation, the *convergence zone theory* posits that conceptual content does only exist in feature maps. According to Damasio (1989) neurons in CZs play no representational roles. That is, they only constitute a means of reactivating previously active patterns in feature maps. Consider the representation of chairs, for example. Damasio assumes that neurons in CZs that link together the visual features of chairs cannot function as a stand-alone representation of this category. Rather these neurons serve to reactivate *chair* features in visual feature maps, which then constitute a conceptual representation of chairs. This is a radical different claim from other *sensory/functional theories* and from amodal accounts in general, which will shape subsequent theories of concepts and conceptual processing. Therefore, neurons in CZs play the important role of reactivating patterns in feature maps during imagery, conceptual processing, and other cognitive tasks (see also Barsalou, 1999). Neurons in a sensory-motor CZ, for example, can reactivate the previously captured sensory-motor state in the absence of bottom-up sensory stimulation. For example, in a recall task or during conceptual processing, neurons in the sensory-motor CZs may re-enact the sensory-motor states that were active while encoding a certain object. The basic idea of re-enactment is essentially the same as that of neural accounts of mental imagery (e.g., Farah, 2000; Grezes & Decety, 2001; Jeannerod, 1995; Kosslyn, 1994). However, while imagery typically restores more complete and vivid images, a re-enactment in memory, conceptualization, and comprehension is only a sketchy record of experience, as we shall see in the next section.

3. The perceptual symbol systems theory. The theory of *perceptual symbol systems* (PSS, Barsalou, 1999; see also Barsalou 2008, 2016) is one of the most prominent theory of concepts within the grounded framework. PSS recovers the basic CZ architecture and shows how a fully functional conceptual system can be built upon it. On this view it is assumed that a concept is a perceptual symbol, namely “a record of the neural activation that arises during perception” (Barsalou, 1999, p. 583). Therefore, concepts or perceptual symbols are conceived as the constituents of a symbol system that are grounded in perception[9]. Barsalou (1999) describes perceptual

---

[9] Rather than referring only to the sensory modalities, “perception” here refers to any aspect of perceived experience including proprioception and introspection.

symbols as unconscious, componential, schematic, and flexible representations. I address each of these characteristics in turn.

In PSS view, concepts, or perceptual symbols, result from something different than conscious subjective experience. They function unconsciously as patterns of neural activation, thus being an alternative to the mental images of classical empiricist theories. Defining a perceptual symbol as an unconscious neural representation has important consequences for the theory of PSS. Specifically, it implies that while the neurons for a particular shape of an object may be active in processing a certain concept such as *chair*, other neurons coding a particular orientation of that same object may be not. As a consequence, perceptual symbols are componential rather than holistic representations, that is, they can be built up from simple parts arranged hierarchically (see also Marr, 1982). This aspect is what mostly differentiates perceptual symbols from mental images.

As exemplified in Figure 4, a perceptual symbol is a schematic record of a perceptual experience, namely it can abstract away from details of position, scale, metric, proportion and viewpoint (see also Prinz, 2002). In fact, attentional mechanisms shape perceptual symbols. As Barsalou (1999, p. 584) explains, “If a configuration of active neurons underlies a perceptual state, selective attention operates on this neural representation, isolating a subset of active neurons. If selective attention focuses on an object’s shape, the neurons representing this shape are selected, and a record of their activation is stored”. Therefore, “the symbol formation process selects and stores a subset of the active neurons in a perceptual state”. For example, during a perceptual experience such as viewing a chair, selective attention may focus on a particular feature of that experience (e.g., the shape of the chair in order to recognise it as a *chair*). As a consequence, other features of the same perceptual experience such as the color, texture, and position of the chair, as well as the surrounding objects, would be filtered out, at least to a significant extent.

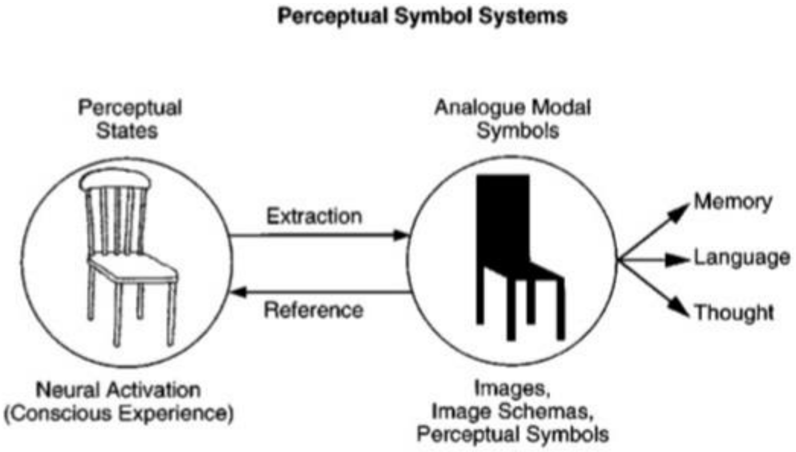

Fig. 4. Subsets of activated neurons in sensory-motor systems are stored in long-term memory to function as symbols. As a consequence, their format is perceptual, and they are grounded in the sensory-motor states that produced them.

Reference: Barsalou [1999]

Moreover, because a perceptual symbol is assumed to be a pattern of neurons, its activation is flexibly adapted to the context at hand. Therefore, in PSS theory perceptual symbols and thus concepts are conceived as flexible representations[10].

In sum, selective attention focuses on particular features of perceived experience on the basis of the context in which the experiencer is immersed, according to the immediate goals of the perceptual experience, etc., and stores records of aspects of that experience in long-term memory, which later function as symbols. Stored perceptual symbols allow recognition of objects on subsequent occasions, and can be modified and updated over time. Indeed, as we experience more objects of the same kind (i.e., more chairs), we refine those symbols that we have already stored. To appropriately represent intervening changes within a category, collections of perceptual symbols must be grouped together. We group these symbols together on the basis of different principles (see Prinz, 2002). For instance, hierarchical symbols consist of different representations of simple parts coinstantiated in a single object. Also, perceptual symbols formed in different modalities may be grouped together on the basis of coinstantiation. If I hear a chirp as my canary flies back into his cage, I may store a record of that chirp along with the visual representation of the canary getting into the cage because the two are coinstantiated, that is, they co-occur. Moreover, objects might change as we are observing them. Think of an unpeeled and a peeled apple, for example. It is the same object, thus the perceptual symbols corresponding to the two states (i.e.,

---

[10] See chapter I, section 3 for more insight on the stable versus flexible representations distinction.

unpeeled and peeled) get grouped together. In addition, we often experience things that co-occur even though they are not physically bounded such as a dog and his bone or a toothbrush with toothpaste. Finally, we could store representations together because they match, namely they are quite similar. Think, for example, of two different dogs such as a collie and a wolf dog.

A different type of principles on the basis of which we group representations together are the causal/explanatory principles, which have been emphasized by theory theorists (e.g., Murphy & Medin, 1985). Causal/explanatory links between representations are more challenging to accomodate in the context of perceptual representations. For example, how does one represent the fact that happiness is causally related to tail wagging? A complete answer to such questions is not yet available in the context of the grounded approach although Prinz (2002, p. 148) has argued that "the failure to see how certain properties can be perceptually represented is almost always a failure of imagination". More compellingly, hybrid models of knowledge offer a concrete answer to this question, as it will be seen in chapter III.

Once a group of linked perceptual symbols is stored in memory, they constitute a long-term memory network. The schematic symbol formation process can operate in any modality of perceived experience: from sight to audition, from touch to smell, and taste, as well as on proprioception and introspection[11]. Thus, for example, visual symbols might originate in visual areas, auditory symbols in auditory areas, proprioceptive symbols in motor areas, and so forth.

Evidence supporting the claim that perceptual symbols originate (and then become established) in all modes of perceived experience comes from neuroimaging studies. A growing number of neuroimaging studies show that modality-specific brain areas are active during conceptual processing (for reviews, see Binder & Desai, 2011; Martin, 2001, 2007; Martin & Chao, 2001). For instance, when people process colour names (e.g., *yellow*), specific areas in the visual cortex become active (Simmons, Ramjee, Beauchamp, McRae, Martin, & Barsalou, 2007). Conversely, when people process concepts for which the auditory modality is important (e.g., *telephone*) specific auditory areas become activated (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008). These results are consistent with the claim that perceptual symbols are multimodal, that is, they become established in all modalities of experience and they are distributed widely throughout the modality-specific areas of the brain.

---

[11] The modality of proprioception is the one that allows oneself to perceive his own body in the space that surrounds it, and to perceive the strength employed in his own movements. From proprioceptive experience, for example, people derive concepts for hand movements and body positions. The modality of introspection is the one relative to the representation of an entity or event in its absence. It is also the one relative to cognitive operations such as rehearsal, elaboration, search, retrieval, comparison, and transformation, and to emotional states. From introspective experience, for example, people derive abstract concepts such as *happiness*, *sadness*, etc.

Further support in favor of the assumption that perceptual symbols become established in all modes of perceived experience comes from behavioral studies showing the involvement of sensorimotor systems during conceptual processing. As this literature is central to the second part of the present manuscript I will extensively illustrate it in Part 2. I will now focus on further theoretical support for the grounded accounts of knowledge.

4. The principle of the neural reuse. It is widely known that evolution recycles existing mechanisms to perform new functions (Gould 1991; Gould & Vrba, 2008). As we saw in the previous section, Barsalou (1999, 2008, 2016) has argued that the same neural regions that are involved in perception and action are involved in conceptual processing. This mechanism, known as perceptual simulation, is of crucial importance in the context of grounded cognition. A perceptual simulation involves the reenactment of configurations of neurons previously established during our interaction with objects in the world. For instance, while processing the concept *violin*, the auditory system might re-enact states (i.e., patterns of neuronal activation) associated with hearing its sound.

According to Barsalou (2016, p. 1130), the principle of "Neural reuse offers a natural account of what is meant by simulation". Within the grounded framework, the reuse of neural circuitry for various cognitive purposes is assumed to be a central organizational principle that contribute to explain the functional structure of the brain. Specifically, grounded theories argue that neural circuits established for one purpose (i.e., perception) are exapted or recycled during evolution or normal development, and are put to different uses (i.e., conception), often without losing their original functions.

The principle of *neural reuse* can be summarized in three main points: a) neural circuits can continue to acquire new uses after an initial or original function is established; b) the acquisition of new uses need not involve unusual circumstances such as injury or loss of established function; 3) the acquisition of a new use need not involve (much) local change to circuit structure (e.g., it might involve only the establishment of functional connections to new neural partners; Anderson, 2010).

Through the principle of *neural reuse*, and the mechanism of simulation, Barsalou explains how core cognitive functions such as the productivity of human thought and language could arise in the context of grounded cognition. As we saw earlier (chapter I, section 1), productivity is the ability to build and understand a potentially infinite number of linguistic expressions starting from a finite number of linguistic elements using combinatorial and recursive rules. In PSS it is shown how productivity is achieved through schematicity: "[…] if a perceptual symbol for *ball* only represents its shape schematically after color and texture have been filtered out, then information about color and texture can later be added productively. For example, the simulation of a *ball* could evolve into a *blue ball* or a *smooth yellow ball*. Because the symbol formation process similarly establishes schematic representations for colors and textures, these representations can be combined productively with perceptual representations for shapes to produce complex simulations” (Barsalou, 1999, p. 593). In a nutshell, the principle of *neural reuse* further supports the mechanism of simulation through which stored schemas of aspects of experience are re-enacted and combined productively.

According to Barsalou (2008, p. 632), although “Amodal formalisms for symbolic operations may provide a theoretical shorthand for expressing what the brain computes, […] simulation, or something else, may be the mechanism that actually implements these operations”. The fact that Barsalou suggests a way for PSS to be productive indicates that the transition from amodal symbolic accounts of knowledge to grounded theories of knowledge has not implied a replacement of more traditional views on concepts with more recent proposals. Rather, grounded theories have complemented more traditional ones developing their relations with the modalities, the body and the environment.

In sum, according to grounded theories the existence of core cognitive functions such as the productivity of human language and thought is not in question. What is in question is how the brain may actually implement such core cognitive functions, that is, in an amodal format rather than in a sensory-motor format. Within the PSS framework, concepts have a sensory-motor format because conceptual content is, at least in part, reaccessed sensory-motor information. If the *neural reuse* hypothesis is correct, and conceptual processing exploits modality-specific resources, then conceptual representations are more likely to have a modality-specific character rather than an amodal one. However, it is worth noting that although conceptual processing might often reuse systems that underly perception (and action and internal states), Barsalou (2016, p. 1129) acknowledges that conceptual processing might also “draw on integrative and abstractive mechanisms in association areas (Binder, 2016; Simmons & Barsalou, 2003)”. That is, conceptual processing might exploit other systems beyond the sensorimotor ones. It is to that we turn in chapter III.

5. Is there a reductionist claim in grounded accounts? Critics of the grounded and embodied approach (Leshinskaya & Caramazza, 2016; Machery, 2016, Mahon, 2015; Mahon & Caramazza, 2008) argue that there is a reductionist claim within the grounded framework. The reductionist claim would consist in the assumption that sensory-motor mechanisms explain concepts and conceptual processing. Specifically, these authors disagree with the grounded claim that perceptual and motor information is constitutive of knowledge representation and language comprehension.

For instance, Mahon & Caramazza (2008) argue against the interpretation of the activation of motor information when observing manipulable objects as evidence that motor information is constitutive of conceptual content (see Barsalou, Simmons, Barbey, & Wilson, 2003; Boulenger, Roy, Paulignan, Deprez, Jeannerod, & Nazir, 2006; Gallese & Lakoff, 2005; Martin, 2007; Pulvermüller, 2005; Rizzolatti & Craighero, 2004). Indeed, these researchers pointed out that the activation of orthographic information during a phonological task has not led to draw the parallel inference that orthography of words is constitutive of their phonology. To illustrate, in a phonological task (e.g. rhyme judgment) orthographic information is activated (as demonstrated by priming effects) such that rhyme judgments are affected by orthographic similarity. For example, subjects are faster to decide that two words rhyme when they are ortographically similar (e.g., pie-tie) than when they are orthographically dissimilar (e.g., rye-tie; see Donnenwerth-Nolan, Tanenhaus, & Seidenberg, 1981; Seidenberg & Tanenhaus, 1979; for a review see Muneaux & Ziegler, 2004). This result has not been interpreted as evidence that phonology is represented in terms of orthography. Rather, orthography is considered a separate type of information that is promiscuously available to the decision mechanisms. As pointed out by Desai (personal communication), in the context of a phonological task, the rapid activation of task-irrelevant information (i.e., orthography) could be due to a close correspondence between orthography, phonology, and semantics in language use. In reading, for example, orthography activates semantics and phonology, while in writing, semantics and phonology activate orthography. During reading, speaking, and listening, several words are processed per second, which makes rapid activation of this information essential. Thus, the triangle of orthography-phonology-semantics is tightly connected such that activation in phonology activates orthography, which in turn feeds back to phonology. This results from the fact that people have a vast amount of experience in reading and writing, which repeatedly reinforces the tight correspondence between orthography and phonology (and semantics). After all, as emphasized by Desai, learning to read and write is learning the orthography-phonology correspondence. Hence, according to Desai, it should be not surprising that performing a task that involves processing phonology but does not explicitly involve processing orthography leads to the activation of the latter as well without entailing that phonology is represented in terms of ortography.

In addition, Mahon & Caramazza (2008) pointed out that neither the activation of phonological information during a naming task has led to draw the inference that the phonology of words is constitutive of their meaning. To illustrate, in a picture naming task, naming latencies were found to be faster in the phonologically related condition (*hammock – hammer*) than in the phonologically unrelated condition (*hammock – button*) (Morsella & Miozzo, 2002; Navarrete & Costa, 2005). That is, not just the name of the picture being named such as *hammock* was activated, but similar-sounding words such as *hammer* were also activated. According to Desai (personal communication), in this case, the rapid activation of task-irrelevant information (i.e., phonological) might be due to the fact that in a system where words (i.e., phonological, orthographic, or sematic forms of the word) have a distributed representation, a word activates one set of phonemes, which in turn activate other words with an overlapping set of phonemes (along with orthographic and semantic codes). It is worth noting that activation simply does not spread everywhere. Rather, a specific word partially activates other specific, overlapping words. *Hammock* does not activate *button* even though *button* is part of the same overall system, and has the same distributed phonemic representation[12]. Thus, in this case, irrelevant activation is brought about by direct feature overlap, and, as in the previous example, is not arbitrary.

In sum, these two examples demonstrate two ways in which task-irrelevant information might be activated. In one case, this is due to the actual real-life correspondence between two types of codes (orthography and phonology). In the other case, the activation of task-irrelevant information is due to the physical overlap of features in a distributed representation system. In neither case the activation of task-irrelevant information is arbitrary.

Mahon & Caramazza (2008) further pointed out that in order to interpret the activation of motor information as conceptually relevant information one must first reject the alternative interpretation that the activation of the motor system is a merely by-product of the way in which activation spreads throughout the system. In particular, these researchers pointed out that a cascade processing model (McClelland & Rumelhart, 1981) within the amodal symbolic accounts of knowledge is able to explain the phenomenon of motor activation when observing

---

[12] Similarly, in semantics once a concept such as *lion* is activated, it causes some activation in concepts with overlapping features such as *tiger*.

manipulable objects. However, assuming that the sensory-motor brain activations are a merely by-product or ‘Pavlovian’ reflex of conceptual/semantic processing implies that there must be an association between reading/hearing a word and doing an action, and vice versa. Specifically, when we perform a throwing action we use the word *throw* to describe it, thus an association is formed. On subsequently reading/hearing the word *throw*, the motor cortex is activated because of this Pavlovian association. However, it is worth noting that there is little to no association between doing or observing actions (or attending to objects) and using words denoting them. That is, language is not used as a running commentary of the immediate environment[13]. Rather, language is used to convey ideas that are not obvious to the listener/reader, and only much more rarely it is used to convey the details of the current physical environment that is available to and attended by the listener. For example, while sitting at the table eating lunch with a colleague we do not say “I’m sitting, I’m eating” and so on. However, this is what would be required for a Pavlovian association, that is, the co-occurrence of actions and relevant language expressing those actions.

Using language as a commentary would not be possible mainly because of simultaneous occurrence of events and multiple possible levels of description. To illustrate, saying “I’m eating” while actually eating destroys the Pavlovian conditioning for “chewing”, and both of those hurt the

---

[13] The counter-arguments that follows were all suggested by Rutvik Desai (personal communication).

conditioning of “talking” and “speaking” and “sitting”. We especially talk about past events, opinions, future plans or possibilities, implications, wishes, feelings, and so on, but only rarely we do talk about the immediate physical environment and actions that are obvious to our interlocutors. Even in the rare cases when we do describe our current actions, we do it if we think that it is not obvious or clear to the listener and hence needs to be emphasized. Otherwise, we do so for some rhetorical purpose such as sarcasm. For example, it is very common for authors of popular funny sitcoms to make people laugh this way. Even when the explicit task is to describe events, as required for a TV commentator of a soccer match for example, the obvious is left out. That is why soccer commentary on TV does not sound like “he kicks, he kicks, he runs, he kicks, he runs…”. On the other hand, one can and does read about all kinds of events without performing or observing of them. This does not imply that one cannot find examples where there is an actual correspondence between actions and words. For example, if we are in a park playing with a ball and our friends ask us to “throw the ball” we might actually throw the ball to them. However, such instances are very rare compared to the thousands of times that we may read or hear the verb *throw* without performing or observing the action in the immediate temporal vicinity. Secondly, if we throw the ball many times without using the word, then no learning occurs due to the few instances in which that did happen. Third, even if we hear our friends asking us to throw the ball and we do it, performing the throwing action is not an automatic reflex. That is, it is not similar at all to orthography of a word activating the phonology of the word within 200 ms or so, without any decision-making. In the *throwing* situation, we consciously decide to follow through with a request, precisely because actually performing an action upon hearing an action word is a rare event.

Now consider a different example. Imagine following instructions while assembling furniture. The instructions might say ‘put the screw into the hole’ and we might perform the action described. Again, this is a conscious decision involving temporal delay and decision-making, not a reflex action. There is no one-to-one correspondence between instructions and actions. Even the instructions do not describe in detail our actual actions, as the obvious is left out. The instructions might say ‘tighten screws #5 and #6” but do not say ‘pick up the screwdriver, insert the head into the slot of the screw, hold the screwdriver firmly and turn your wrist clockwise, etc.”. Thus, not only actually following instructions constitutes a very small percentage of life and our total language use, it nevertheless is not nearly enough to establish Pavlovian conditioning.

One may attempt to change the argument a bit and say that the conditioning is established during early childhood, when processing child-directed speech (CDS). The intuition is that CDS contains the type of correspondence required for Pavlovian conditioning. However, two problems might be raised. First, even if the conditioning were established in early childhood, it would be quickly eliminated once the child turns say 5-6 years old and learns to read and parents and others stop communicating in a manner they do with babies or toddlers. An even bigger problem is that CDS is not a running commentary of the immediate environment either. This is the fundamental problem of language acquisition. Caregivers simply do not continuously describe the immediate actions and objects to the child. When a mother gets home, she might say “I am home!” or “Look who is home!” But not “I am grasping the doorknob. I am turning the knob anticlockwise. I am pushing the door open. I am taking the first step in the house…”. That is why additional mechanism such as joint attention and statistical analysis need to be assumed in order to explain language acquisition. Therefore, the interpretation of task-irrelevant motor activation as a consequence of Pavlovian conditioning does not work.

However, task-irrelevant activation could still be due to feature overlap as in the *hammock – hammer* example above. Activation of irrelevant phonological information during a naming task is due to the two words sharing a format (distributed set of phonemes), a network, and features. However, the fundamental claim of the amodal account is that concepts are represented as arbitrary amodal symbols that bear no resemblance to sensory-motor processes. An amodal symbol has an arbitrary form that has no similarity to, and contains no information in, the visual, auditory, or motor systems. Just as the word *cup* has no similarity to what a cup looks like or how it is used or what it feels like to use it. Thus, by definition, there cannot possibly be any overlap between an amodal concept and the sensory-motor cortex. Unlike *hammock* that triggers partial activation of *hammer*, the abstract symbol for *throw* cannot spread anything to the motor cortex because they do not share anything. Hence, neither the Pavlovian conditioning nor overlapping features can explain rapid activation of sensory-motor areas during conceptual tasks. There is only one possibility left. The activation in sensory-motor systems represents semantics itself as predicted by grounded accounts of knowledge.

Finally, a more general point regarding the nature of Mahon & Caramazza's (2008) argument. Most recent knowledge of human brain functions comes from 'recording' methods, such as fMRI, PET, EEG, MEG, ECoG, and NIRS, which are correlational by nature and do not show causality. If some researcher entertain principled belief regarding correlational methods showing unnecessary or irrelevant signals, then those researchers must necessarily dismiss all results from all these methods. If they are not willing to do that, then they should not bring up this possibility only when results go against their own theories.

That being said, it is difficult to understand why grounded theories are accused of reductionism especially if, in addition to the above-mentioned arguments that fairly dismiss the skeptical hypothesis (Leshinskaya & Caramazza, 2016; Machery, 2016, Mahon, 2015; Mahon & Caramazza, 2008) against grounded accounts, one also consider that several theorists within the grounded and embodied framework agree that sensory-motor mechanisms are insufficient in and of themselves for explaining concepts and conceptual processing. For example, Barsalou (2008) illustrates how both language and internal states contribute to the representation of concepts above and beyond the sensory-motor modalities. Also, Barsalou (1999) emphasizes that internal states play central roles in conceptual processing, especially for abstract concepts (see also Barsalou & Wiemer-Hastings, 2005). As we saw earlier, Barsalou (2016) also suggests that conceptual processing might draw on integrative and abstractive mechanisms in association areas (see also Simmons & Barsalou, 2003). Similarly, Pulvermüller (2013) argues that disembodied mechanisms in the brain's hub regions contribute to semantic meaning (see also Pulvermüller, 2012; Pulvermüller & Garagnani, 2014). In a seminal paper, Pulvermüller (2013) suggested the *correlation learning* principle, that is, a neural key to understand brain topographies of linguistic and semantic processes. In a nutshell, the core idea behind the principle is that "neurons that fire together wire together and neurons out of synk delink" (Pulvermüller, 2013 p. 462). In order to explain why semantic brain processes have been observed within both sensory and motor areas and multimodal association cortices located far away from sensory and motor fields, Pulvermüller (2013, p. 464) suggested a key role for intermediary areas. "To link the spoken word form *grasp* to the concordant motor movement, or the articulation pattern for pronouncing the word *grass* to specific visual knowledge about color and shape, nerve cells in motor and sensory areas are necessary; in addition, intermediary area neurons are equally required to build circuits that bind sensory and motor information. Neuroanatomical structure determines that the emerging circuits include neurons in modality-nonspecific areas of cortex." Therefore, if on the one hand modality-preferential areas are required to link symbols to information in the sensorimotor system, on the other hand, higher multimodal relay areas are recruited to bridge information coming from different modality-preferential systems (i.e., sensory and motor systems). In sum, the *correlation learning* principle together with the neuroanatomical structure (i.e., cortical connectivity) can explain why semantic processing is distributed over both modality-preferential and multimodal areas. I will further tackle these issues in the next chapter.

III. Hybrid accounts of knowledge

What do a record album and a necklace have in common? Of course, not the sound, shape, colour, or actions their structures afford, let alone their names and verbal descriptions. Rather, as many of the readers will have already inferred, these objects share the property of being *possible gifts*. On this type of higher-order generalizations that disregard modality-specific information about objects depends much of our conceptual processing. It is possible that different aspects of the same concept (e.g., what an object looks like versus whether or not it is of a certain type) are stored in different representational formats. To account for this fine-grained differences among conceptual representations, hybrid or pluralistic views have been put forward (e.g., Dove, 2009; Malt, 2010; Patterson et al., 2007; Simmons & Barsalou, 2003). According to such views, some concepts (e.g., concrete nouns) are grounded in perceptual representations while others (e.g., abstract concepts) are amodal or “disembodied”. Malt (2010), for example, conceives relational or thematic representations as being amodal while other concepts as being perceptually grounded. Given the fact that lesions in some perceptual regions result in uni-modal, category-specific semantic deficits (Kan et al., 2003) and lesions in the anterior temporal lobes (ATLs) result in multi-modal, category-general semantic deficits (Patterson et al., 2007), it is at least possible that different neural substrates underly different concepts. In the remainder of this section I introduce and discuss two different hybrid models of knowledge: the *distributed-plus-hub* and *conceptual topography* proposals.

1. The distributed-plus-hub view. According to the *distributed-plus-hub* view, sensory, motor and linguistic information is necessary but not sufficient to explain conceptual processing (Patterson et al., 2007; Lambon Ralph, Sage, Jones, Mayberry, 2010). On this view, in addition to direct neuroanatomical pathways between different sensory, motor and linguistic regions, the neural network for semantic memory requires a single convergence zone or hub that supports the interactive activation of representations in all modalities, for all semantic categories.

Despite the assumption of the existence of a convergence zone, Patterson et al.'s proposal varies in at least two respects if compared to Damasio's (1989) *convergence zone theory*. As we saw earlier, Damasio (1989) proposed the existence of convergence zones, namely association areas in the brain intended to associate different aspects of knowledge. According to Damasio, however, these association areas are assumed to be multiple, hierarchically organized convergence regions. For example, there is a cross-modal CZ that encode associations between visual representations of shape and corresponding actions, another that encodes associations between shape and object name, and so on. By contrast, the *distributed-plus-hub view* posits the existence of a single convergence zone or hub. Patterson et al.'s (2007) claim is that in addition to modality-specific regions and connections, the various different surface representations (i.e., shape) connect to and communicate through, a shared, amodal hub (shown as a red area in Fig. 5, Panel b) that would be located in the anterior temporal lobes.

Furthermore, Damasio suggested that associations between different pairs of attributes are encoded along different neuroanatomical pathways. Thus, convergence zones become differentially important for representing different semantic categories. For example, because humans frequently interact with tools and other man-made objects, the zone that links object shape and action might be more important for knowledge of man-made artefacts than for knowledge of living things. Similarly, because animals move in characteristic ways, the zone that links shape to movement might acquire special salience for knowledge of animals. On the contrary, the *distributed-plus-hub view* does not predict any specific relation between semantic categories and certain modality-specific systems. At the hub stage, associations between different pairs of attributes (such as shape and name, shape and action, or shape and colour) are all processed by a common set of neurons and synapses, regardless of the task. Representations in the hub are assumed to be amodal in that they can be generated from any individual receptive modality and can be used to generate behaviour in any individual expressive modality. As a consequence, it is hypothesized that a damage to the hub should produce a semantic impairment that is independent of the modality of input (objects, pictures, words, sounds, tastes, and so on) and of the modality of output (for example, naming an object, drawing it or using it correctly) given that information in the hub is assumed to be amodal. Figure 5 illustrates the differences between the *convergence zone theory* (referred to as the *distributed-only view*) and the *distributed-plus-hub view*.

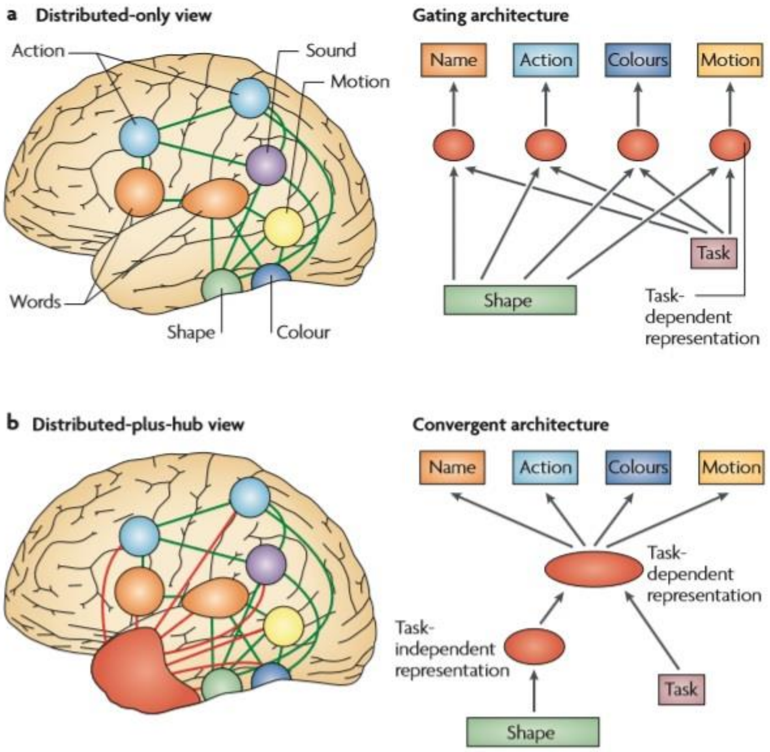

Fig. 5. Neuroanatomical distribution of the cortical semantic network according to *convergence zone theory* (referred to as the *distributed-only view*, Panel a) and *distributed-plus-hub* view (Panel b).

Reference: Patterson et al. [2007]

Evidence supporting the existence of a hub part in the semantic network comes from neuropsychological research.

Lesion studies point to the anterior temporal lobe (ATL) as being crucial for semantic processing. In particular, a lesion-overlap study (Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996) that tested picture naming with anomic[14] patients showed different degrees of correlation between the locus of the lesion and the symptoms. To illustrate, a lesion centred on the left temporal pole resulted in the tightest overlap with impaired naming of *famous faces*, whereas lesions in the anterior, inferior left temporal lobe correlated with impaired naming of *animals*. Scarce performance on *tool*-naming was associated with damage in the posterior, lateral left temporal lobe as well as in the temporo–occipito–parietal junction but with the lowest degree of lesion-symptom overlap. Thus, lesion to the left ATL strongly correlated with impaired naming for two of the three categories tested (i.e., *famous faces*, *animals*, *tools*). When healthy participants performed the same task with the same categories in a PET activation paradigm, all three stimulus types (i.e., *famous faces*, *animals*, *tools*) yielded significant blood-flow increases (relative to a control condition) in the left temporal pole.

In addition, functional (or metabolic) imaging studies have shown dysfunctions in the bilateral anterior temporal lobe for Semantic Dementia (SD) patients. Diehl, Grimmer, Drzezga, Riemenschneider, Förstl, and Kurz, (2004) reported more extensive hypometabolism along the length of the inferior left temporal lobe. Hypometabolism in SD was also detected in the

---

[14] Patients who have difficulties to name objects.

left insula and orbito-frontal areas (Desgranges, Matuszewski, Piolino, Chételat, Mézenge, Landeau, De La Sayette, Belliard, & Eustache, 2007) and in the rostral temporal lobes (Nestor, Fryer, & Hodges, 2006). Nestor et al. (2006) also showed that semantic impairment is much milder in Alzheimer Disease (AD) in which hypometabolism is much more widespread, than in SD in which hypometabolism mainly concerns the ATL.

Consistent with functional imaging studies, structural imaging research indicates relative preservation of the posterior temporal lobe in SD (Desgranges, et al., 2007; Nestor, et al., 2006; Davies, Graham, Xuereb, Williams, & Hodges, 2004; Mummery, Patterson, Price, Ashburner, Frackowiak, & Hodges, 2000). In sum, both metabolic and structural imaging studies of patients with SD suggest that lesions are most pronounced in the anterior and inferior parts of the temporal lobes.

It is worth noting that although other functional imaging studies have also implicated some combination of frontal, posterior temporal, temporo-parietal and parietal regions in the cortical semantic network, Devlin, Russell, Davis, Price, Wilson, Moss, Matthews, and Tyler (2000) demonstrated that the significant anterior temporal lobe activation evident with PET is largely absent with fMRI as a consequence of MRI susceptibility artefact. That is, the anterior temporal lobe is shy to fMRI (see also Patterson et al., 2007).

The strongest ATL activation is usually observed when people are required to recognize or identify unique concepts such as famous building or individuals (i.e., Eiffel Tower or Princess

Diana, see for example, Gorno-Tempini & Price, 2001; Nakamura, Kawashima, Sato, Nakamura, Sugiura, Kato, Hatano, Ito, Fukuda, Schormann, & Zilles, 2000) but also names and even voices (e.g., Gorno-Tempini, Price, Josephs, Vandenberghe, Cappa, Kapur, Frackowiak, & Tempini, 1998; Tsukiura, Mochizuki-Kawai, & Fujii, 2006; Nakamura, Kawashima, Sugiura, Kato, Nakamura, Hatano, Nagumo, Kubota, Fukuda, Ito, & Kojima, 2001). Interestingly, SD patients are profoundly impaired at recognizing famous individuals from photographs, names and verbal descriptions (Snowden, Thompson, & Neary, 2004). More broadly, this impairment seems to reflect a general sensitivity of SD patients to the specificity with which an item is categorized[15]. Patients perform well if a relatively coarse or general categorization of the stimulus is required. For example, SD patients can call a picture *animal*, without being able to name it *chicken* or even *bird* (Hodges, Graham, & Patterson, 1995). This pattern does not arise simply because tasks that require precise classification are more difficult. Indeed, as Rogers and Patterson (2007, see also Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976) showed, healthy adults are faster and more accurate at classifying items at the basic level (for example, *dog*) relative to a more general level (for example, *animal*). Patients with SD show the reverse of the basic level effect: they have greater difficulties at classifying items at the more precise basic or subordinate level. Such findings suggest

---

[15] See also chapter I, section 1 on the sensitivity to specificity of SD.

that semantic tasks that require the distinctive classification of a stimulus place particularly strong demands on the ATL regions that are affected in SD.

Additional evidence supporting the existence of a hub part in the semantic network comes from computational modelling. Computer simulations with neural-network models have shown that networks in which all forms of information about concepts are, at some point, processed through the same population of neurons and synapses exhibit functional properties that explain how the semantic system is able to learn conceptual similarity relationships, that is, higher-order generalizations that disregard modality-specific information about objects (Rogers & McClelland, 2004). Consider, for example, how a convergent zone architecture (referred to as “gating architecture” in Fig. 5, Panel a) might encode information about a pear. The pathway that stores associations between shape and name will learn a representation that encodes both visual and phonological similarity to other known objects. Thus, a pear and a light-bulb will generate similar representations because they have similar shapes; a pear and a bear will generate similar representations because they have similar-sounding names; and a pear and a banana will generate rather different representations because they have different shapes and different-sounding names (Patterson et al., 2007). Therefore, the convergent zone architecture or gating architecture will not encode conceptual similarity relationships, which should capture the fact that pears and bananas are semantically related because they are both fruits whereas pears and light-bulbs or pears and bears are not. One might suggest that attributing greater weight or salience to some sensory or motor features can solve the problem. For instance, if similarity of taste is more salient than similarity of shape or word-sound, then bananas and pears, which are both sweet, might be judged more similar to one another than pears and light-bulbs or pears and bears. As highlighted by Patterson et al. (2007), the problem with this approach is that the salience of a given feature varies from one semantic category to another: colour, for example, is important for categorizing fruits (consider lemons versus limes), but is irrelevant for categorizing toys (see for example, Macario, 1991). Thus, to determine the salience of a given sensory, motor or linguistic feature, one must know to which category the item belongs, but the item is difficult to categorize without knowing the salience of its observed features (Gelman & Williams, 1998; Keil, 1989; Murphy & Medin, 1985). We are at an impasse: in other words, there is no single salience for a given property that will correctly capture semantic similarity for all concepts.

On the contrary, in the *distributed-plus-hub* view (referred to as "convergent architecture" in Fig. 5, Panel b) the same units that code the association between shape and name must also learn to code relationships between shape and colour, shape and action, shape and texture, and so on, as well as complementary mappings (that is, mappings in the other direction) between these surface representations. As a consequence, the internal representations that emerge look very different. They are not dominated by the similarities expressed in any individual modality (or pair of modalities), but instead reflect the similarity relationships that is apparent across all of the modality-specific representations taken together (Patterson et al., 2007). In other words, the intermediate representations that arise in the hub can promote generalization across items that are conceptually related, even if they do not happen to have similar shapes, colours, associated actions, and so on. These representations are amodal in that they can be generated from any individual receptive modality and can be used to generate behaviour in any individual expressive modality. They are semantic in that they express the conceptual similarity relations among concepts that are critical to semantic generalization and induction, even though, in themselves, they have no retrievable content (Rogers & McClelland, 2004). Rogers et al. (2004) implemented a convergent architecture in a fully recurrent connectionist model that was trained to map between simple visual representations of objects, verbal descriptions of the objects, and the objects' names. Interestingly, representations in the model captured aspects of similarity structure that were not apparent when considering the verbal descriptions or the visual representations alone. For instance, considering just visual similarities, fruits and vegetables share many properties with man-made objects whereas, considering just the verbal descriptions, fruits and vegetables are quite distinct from both animals and man-made objects, although they share a few properties with animals. When trained on these patterns, the model acquired representations in which the fruits and vegetables were distinct from animals and man-made objects, but were actually more similar to man-made objects. Thus, the model made the counterintuitive suggestion that fruits and vegetables, although being natural and not man-made, might be represented as being more similar to artefacts than to other natural things like animals in the human semantic system. Consistent with this suggestion, when asked to classify pictures of apples and other fruits and vegetables as being *plant* (correct), *animal*, or *man-made artefact*, SD patients mis-assigned a number of fruits and vegetables to the artefact category, despite making few errors when the choice categories for the apple were *fruit*, *bird* or *land animal* (Rogers et al., 2004).

In sum, the *distributed-plus-hub view* accomodates both a grounded and a more traditional amodal perspective of conceptual organization and processing into a single account. It is worth acknowledging, however, that while Patterson et al. (2007) think that because SD generates highly multi-modal deficits then the ATLs must implement amodal representations, Bonner, Peelle, Cook, and Grossman (2013) think that multi-modal deficits caused by SD suggest that in the ATLs reside heteromodal perceptual representations, namely representations that are encoded in several different perceptual formats (see also McCaffrey, 2015). Thus, additional theories about how association areas such as the ATLs contribute to conceptual knowledge are needed. In this regard, it is worth mentioning what Barsalou recently pointed out: “In recent years, Martin (personal communication) has been asking researchers who use the term "amodal" what they mean by it. Overwhelmingly, he finds that they mean multimodal, not amodal. Sloppy use of "amodal" has resulted in this confusing state-of-affairs" (Barsalou, 2016, p. 1126).

Nonetheless, these hybridizations can prove very fruitful not only for explaining neuropsychological results but also as a plausible and more general account of the organization and representation of concepts in the brain.

2. The conceptual topography theory. The *conceptual topography theory* (CTT, Simmons & Barsalou, 2003) aims at integrating amodal and more recent perceptual theories of category-specific deficits into one single account that shares properties of both classes of theories. In particular, the CTT revises Damasio's (1989) *convergence zone theory*. While Damasio (1989) proposes that conjunctive neurons in CZs play no representational roles and only constitute a means of reactivating previously active patterns in feature maps, CTT assumes that patterns of neurons in CZs can also function as stand-alone representations, in particular during automatised feed-forward processes such as categorisation and word association. For example, during the categorisation of familiar objects (e.g., *chairs*), "active feature detectors feed activation into the conjunctive[16] neurons that integrate chair features. These conjunctive neurons then feed activation to response

---

[16] Simmons & Barsalou (2003) refer to neurons in convergence zones as "conjunctive neurons".

systems, such as the system that vocally produces a category name (e.g., “chair”). In this chain of feed-forward processing, the pattern of active conjunctive neurons functions as a representation sufficient to produce a correct response – reactivating a feature map pattern is not necessary” (Simmons & Barsalou, 2003, p. 456). Conversely, under demanding conditions such as when constructing, manipulating, or evaluating a conceptual representation reactivating a feature map pattern becomes necessary (see also Kan, Barsalou, Solomon, Minor, & Thompson-Schill, 2003).

According to Simmons & Barsalou (2003), CTT contains four sub-systems on each of the six sensory-motor modalities (i.e., visual, auditory, somatosensory, motor, gustatory, olfactory). Specifically, each modality contains feature maps, analytic CZs, holistic CZs, and modality CZs. To illustrate, feature maps detect and represent features such as colour, line orientation, pitch, physical pressure at bodily locations, and so forth. Analytic CZs conjoin modality-specific conjunctions of features, that is, analytic conceptual properties such as shape, color, texture, movement, etc. Holistic CZs conjoin modality-specific conjunctions of features, that is, holistic conceptual properties such as *eyes*, *nose*, *mouth* in the visual modality. Modality CZs conjoin analytic and holistic properties on a single modality. Neurons in these association areas represent properties such as *red*, *round*, *smooth*, etc. In addition to these four subs-ystems, cross-modal CZs integrate the modality-specific CZs.

The CTT introduces two additional principles to CZ theory: the *similarity in topography* (SIT) and *variable dispersion* principles. Both principles concern the organisation of neurons in CZs. Basically, the SIT principle claims that the physical structure of the world is reflected in the spatial organization or topography of the brain's association areas. More specifically, according to the SIT principle "the spatial proximity of two neurons in a CZ reflects the similarity of the features they conjoin. As two sets of conjoined features become more similar, the conjunctive neurons that link them lie closer together in the CZ's spatial topography" (Simmons & Barsalou, 2003, p. 457). For example, on viewing a human face, large numbers of neurons distributed throughout visual feature maps become active to represent its features. Subsequently, neurons in a visual CZ conjoin these features by associating the respective feature map neurons. According to the SIT principle, the populations of neurons in CZ for a human and a monkey face lie closer together than the populations of conjunctive neurons for a human and an elephant face. Furthermore, the conjunctive neurons that represent all three faces (i.e., human, monkey, and elephant) lie closer together than the conjunctive neurons that represent some completely different type of object, such as a chair. In general, the topographic proximity of neurons in CZ reflects the similarity of the features they link. Figure 6 depicts a schematic illustration of the SIT principle.

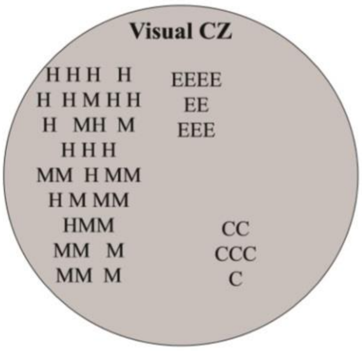

Fig. 6. Illustration of localised conjunctive neurons in a visual CZ for the features of a human (H), a monkey (M), an elephant (E), and a chair(C).

Reference: Simmons & Barsalou [2003]

As with previous amodal accounts of knowledge, statistical structure of concepts is central for the SIT principle. That is, concepts structure mirrors statistical frequency and pairings of features of objects in the world. However, unlike previous amodal accounts, the SIT principle implements this statistical structure topographically in the brain's association areas.

The second additional principle that the *conceptual topography theory* (CTT) develops is the *variable dispersion* principle. Given the distributed character of the CTT, Simmons & Barsalou (2003) assume that neurons that represent a category in a CZ are located in a distributed area that not only contains conjunctive neurons for that category but that also contains conjunctive neurons for other categories. Therefore, "conjunctive neurons for a category are dispersed in clumps, with clumps for other categories falling between" (Simmons &

Barsalou, 2003, p. 459). Furthermore, a given clump may contain conjunctive neurons used by more than one category. The conjunctive neurons for a category are typically not contiguous in a CZ. The *variable dispersion* principle concerns these noncontiguous clusters of conjunctive neurons such that: “In a CZ, the proximity of the noncontiguous clusters for a category reflects the similarity of its instances. As the instances of a category decrease in similarity, its noncontiguous clusters of conjunctive neurons become increasingly dispersed in the CZ’s spatial topography” (Simmons & Barsalou, 2003, p. 459). Figure 7 illustrates the *variable dispersion* principle, showing both low and high dispersion profiles for categories having similar vs. dissimilar instances respectively (e.g., mammals vs. artefacts).

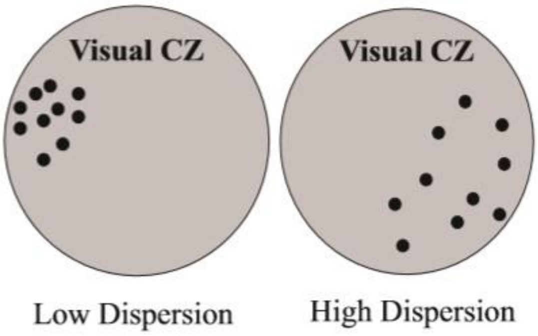

Fig. 7. Illustration of how noncontiguous clusters of conjunctive neurons represent a category, with low dispersion for a category on the left, and high dispersion for a category on the right.

Reference: Simmons & Barsalou [2003]

The *variable dispersion* principle is tightly connected to the SIT principle. If according to the SIT principle, proximity of neurons reflects features similarity, according to the variable dispersion principle, the clusters of conjunctive neurons that represent a category lie closer together as within-category similarity increases.

The *variable dispersion* principle has significant implications for conceptual deficits. For example, within-category similarity is relatively low for artefacts. A fork and a pan do not resemble each other at all. Thus, according to the variable dispersion principle, conjunctive neurons that represent artefacts should be highly dispersed. As a consequence, the *artefacts* category should be less susceptible to damage. In contrast, because *animals* has much higher within-category similarity, its conjunctive neurons should be more tightly localised, and therefore this category should be more susceptible to damage. A related implication is that certain concepts may be more susceptible to lesions in some modality CZs than in others. If the actions afforded by *tools*, for example, are more similar than *tools*' visual properties, tool deficits should be more likely following lesions to motor areas than to visual areas. Conversely, if the visual properties for *animals* are more similar than their other properties, animal deficits should be most likely to follow lesions to visual areas.

The SIT and *variable dispersion* principles also offer an explanation for why conceptual deficits disrupt superordinate categories. Consider the superordinate category of *animals*.

Because its members share many properties (e.g, they all move, eat, have two eyes, have a mouth, have reproductive organs, and so on), the conjunctive neurons that code them should all be mixed together topographically within a modality CZ. Thus a lesion that damages the conjunctive neurons for one basic level category, say *dog*, is likely to damage the conjunctive neurons for other basic level categories (e.g., *cat*, *cow*, *snake*, etc.) that share its prototypical features (Rosch & Mervis, 1975; Hampton, 1993). The result is the loss of a superordinate category, or at least much of it.

In addition, consistent with the *conceptual topography* theory, Fernandino, Binder, Desai, Pendl, Humphries, Gross, Conant, and Seidenberg (2015) found that cortex activation patterns reflected the natural correlations of modalities and attributes in the world (e.g., both visual and somatosensory modalities were associated with shape, manipulation was also associated with shape, the hearing modality was associated with visual motion). In addition, they found that areas previously implicated in multisensory integration were co-activated by the corresponding attributes. Finally, they showed that the only areas activated by all attributes were high-level cortical hubs (angular gyrus, precuneus/posterior cingulate/retrosplenial cortex, parahippocampal gyrus, and medial prefrontal cortex).

In sum, topographical mapping might constitute a fundamental principle of brain organisation at multiple levels.

Summary

Two classes of theories of concepts were introduced and discussed in Part 1: the amodal symbolic accounts of knowledge and the grounded accounts of knowledge. Viable hybrid models of knowledge, that combine aspects of both classes of theories, were then presented. Chapter I illustrated neuropsychological results concerning Semantic Dementia (SD) that are compatible with an amodal format of concepts. Two important distinctions within traditional accounts of knowledge were then discussed: the local vs. distributed distinction and the stable vs. flexible distinction. Behavioral, electrophysiological and fMRI results suggesting a distributed and flexible organization is more likely were illustrated.

Chapter II introduced grounded theories of knowledge as an attempt to tackle and solve the symbol grounding problem or problem of intentionality that affects amodal accounts. Grounded theories purport to show that the reuse of neural circuitry for various cognitive purposes is a central organizational principle that contribute to explain the functional structure of the brain. Behavioral, electrophysiological and fMRI evidence supporting this view was presented. The final section of chapter II addressed the skeptical claims of opponents of grounded accounts with a series of counter-arguments.

Chapter III presented hybrid models of knowledge suggesting fruitful combinations of amodal and grounded aspects of

theories of concepts to overcome the strict dichotomy between the two outlined views.

This general review of amodal, grounded, and hybrid accounts of knowledge was aimed at showing that the transition from amodal to grounded theories of knowledge has not represented a fracture in the cognitive science. On the contrary, grounded cognitition has complemented traditional approaches taking into account the modalities, the body, and the environment’s influence on cognitve mechanisms (Barsalou, 2016).

Importantly, it was pointed out that the distributed and flexible organization of concepts developed within the traditional accounts of knowledge can also work in the context of the grounded accounts. In addition, a continuity between amodal and grounded theories of knowledge was highlighted that rests on the idea that the conceptual system is organized in a modality-specific fashion. Indeed, the modality-specific organization of knowledge is the main organizational principle of some amodal accounts such as the *sensory/functional theory* (Warrington & McCarthy, 1983, 1987; Warrington & Shallice, 1984) as well as of grounded theories such as the *perceptual symbol systems theory* (PSS, Barsalou, 1999, 2008). Hybrid models further confirm that a blending between the two perspectives (i.e., amodal and grounded) can prove very advantageous.

In conclusion, by focusing on the continuity rather than the fracture between the two approaches to the representation of knowledge in the brain, I did not intend to lessen the scope of the debate on the representational format of concepts. Rather, my aim was to show that the two perspectives are not absolutely incompatible with one another and a compromise is desirable and needed much more than a sterile dichotomy.

PART 2

THE DEBATE ON THE FORMAT OF CONCEPTS: EXPERIMENTAL EVIDENCE

OVERVIEW

There is increasing evidence that modality-specific representations underlie concepts (e.g., Pecher, Zeelenberg, & Barsalou, 2003; van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008), memory (e.g., Glenberg, 1997), and language comprehension (e.g., Glenberg & Kaschak, 2002; Stanfield & Zwaan, 2001).

For the purposes of this dissertation, I will be focusing on evidence suggesting that modality-specific representations underlie concepts and conceptual processing. Neuroimaging research shows that modality-specific cortical areas related to sensory and motor processing are involved in semantic processing (see Binder & Desai, 2011; Kiefer & Pulvermüller, 2012 for reviews). The finding that sensorimotor systems are engaged during conceptual processing boosts the idea that conceptual content is distributed in the sensorimotor systems.

Behavioral literature offers further evidence in support of the assumption that perceptual information is engaged in conceptual processing showing a cost for performance in terms of speed and accuracy when two different modalities alternate, compared to when the same modality is presented (Pecher, Zeelenberg & Barsalou, 2003). This effect, known as the Modality-Shifting or Modality-Switch effect (henceforth, MSE), has been claimed to be the result of a perceptual simulation. This research reports four experiments aimed at exploring this claim.

Experiments 1 & 2 (Study 1) use a standard priming paradigm to test whether the MSE is the result of the automatic activation of sensory information. Indeed, as we saw earlier, PSS view assumes that perceptual simulations are automatic, that is, they function unconsciously as re-enactment of patterns of neural activation in sensory-motor areas of the brain rather than as conscious mental images. Crucially, we manipulated the stimuli’s presentation modality across experiments such that Experiment 1 had written sentences as stimuli, whereas Experiment 2 had aurally presented sentences. This manipulation is aimed at verifying whether the MSE is a robust effect that arises in both reading and speech processing.

Experiments 3 & 4 (Study 2) were designed to further assess the scope of the MSE. The main purpose of Study 2 was to test whether the impact of the mode of presentation of stimuli (i.e., visual: through the monitor, aural: through a pair of headphones) on the conceptual MSE is affected by the depth of processing required by the task. Specifically, Experiment 3 was aimed at examining the influence of the mode of presentation of stimuli on the MSE in a property verification priming paradigm, whereas Experiment 4 was aimed at examining the same issue in a lexical decision priming paradigm.

IV. STUDY 1: EXPERIMENTS 1 & 2

The Modality-Switch Effect: Visually and Aurally Presented Prime Sentences Activate Our Senses[17]

Introduction

Object's properties can be perceived through different sensory modalities. Thus, while detecting the color of a traffic light in a cross-road mainly involves the visual modality, perceiving the melody of a violin during a classic concert mainly involves the auditory modality.

According to grounded theories of knowledge (Barsalou, 2008; for a recent discussion see Pecher 2013), sensory information is also active when we process the concepts *TRAFFIC LIGHT* and *VIOLIN*[18]. In other words, processing of concepts would imply a re-enactment of previously recorded and integrated perceptual information concerning the objects or entities they refer to. Hence, a similar pattern of neural activation in sensory systems would be preserved in representation: while processing the concept *VIOLIN*, for instance, the auditory system would re-enact states associated with hearing its sound. This re-enactment is also known as perceptual simulation.

According to embodied and grounded theories (see also Barsalou, 1999, 2003, 2009; Glenberg, 1997), the re-enactment

---

[17] The final publication is available at Frontiers in Psychology via http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01668/full.

[18] Henceforth, I use uppercase italics for concepts (*VIOLIN*) and lowercase italics for properties of the concepts (*melody*).

evoked by linguistic stimuli represents a form of simulated experience. It is worth mentioning that the notion of simulation varies in detail and depth (Borghi & Cimatti, 2010; for a review see Decety & Grèzes, 2006). More specifically two slightly different views are taken into account in the Embodied Cognition theories. According to the first, the notion of simulation is mainly based on the re-enactment of past sensorimotor experience (Barsalou, 1999). The second view stresses the predictive aspect of simulation, suggesting that the automatic simulated re-enactment of the observed actions and objects is at the basis of a direct form of action preparation and comprehension (e.g., Gallese, 2009). Here we mainly focus on simulation as a form of multimodal re-enactment of previously sensory experiences.

A growing number of neuroimaging studies show that modality-specific brain areas are active during conceptual processing (for reviews, see Martin, 2001, 2007; Martin & Chao, 2001). For instance, when people process colour names (e.g., *YELLOW*), colour areas in the visual cortex become active (Simmons et al., 2007). Conversely, when people process concepts for which the auditory modality is important (e.g., *TELEPHONE*), auditory areas become activated (Kiefer et al., 2008). These results are consistent with the claim that people simulate concepts in sensory systems.

The behavioral literature offers further evidence in support of the assumption that perceptual information is engaged in conceptual processing, showing a cost for performance in terms of speed and accuracy when two different modalities alternate, compared to when the same modality is presented (Marques, 2006; Pecher, Zeelenberg & Barsalou, 2003, 2004; van Dantzig, Pecher, Zeelenberg & Barsalou, 2008; Vermeulen, Niedenthal, Luminet, 2007). This effect, known as the Modality-Shifting effect or Modality-Switch effect (henceforth, MSE) was initially found in a pure perceptual study by Spence, Nicholls & Driver (2001). Participants were faced with a visual, tactile, or auditory signal that could appear on the left or on the right. Their task was to detect the location of the signal (i.e., left or right) as rapidly as possible by pressing one of two pedals. Performance was faster and more accurate for trials that were preceded by a same-modality trial (e.g., visual-visual) than for trials that were preceded by a different-modality trial (e.g., auditory-visual).

Crucially, the MSE was replicated using a conceptual task (Pecher et al., 2003). Pecher and colleagues (2003) used a property verification task (see Collins & Quillian, 1969; Conrad, 1972; Glass & Holyoak, 1975; Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974; Solomon & Barsalou, 2001 and 2004): participants were presented with short sentences having a ‘concept can be property’ scheme (e.g., ‘*BANANA* can be *yellow*’) and had to verify whether the property was true of the concept. Related pairs of property verification sentences alternated throughout the task: a context sentence (i.e., the one presented first) was always followed by a target sentence. Properties in both context and target sentences could be in one of six modalities (vision, audition, taste, smell, touch, and action). The key manipulation consisted in the fact that each target sentence could be preceded by a sentence with a property in the same or in a different modality. Results showed that properties were verified faster and more accurately in same-modality trials than in different-modality trials. For instance, participants were faster and more accurate when verifying the property *pastel* for *BABY CLOTHES*, if they previously verified the property *yellow* for *BANANA* (both visual) rather than the property *rustling* for *LEAVES* (auditory context – visual target). This finding suggests that conceptual processing strongly relies on perceptual and motor information.

However, two possible criticisms of the study by Pecher et al. (2003) lay on the fact that (1) their property verification paradigm might have involved less automatic processes compared to those that a simulation would entail (on the automaticity of simulation see Jeannerod, 2006; Pulvermüller, 2005); (2) the MSE with conceptual representations could be explained assuming that concepts are abstract, amodal symbols rather than grounded in perception and action systems (see Mahon & Caramazza, 2008 for a discussion).

As to the first criticism, indeed, it has been argued that simulations are fast, implicit and automatic and only involve exogenous attention. In a recent study, Connell & Lynott (2012) linked perceptual attention to conceptual processing (on the relationship between concepts and attention see also Myachykov, Scheepers and Shtyrov, 2013). These authors claimed that while exogenous attentional mechanisms are involved when incoming stimuli automatically grab attention, endogenous attentional mechanisms are involved when people consciously focus attention on a particular modality (see also Connell & Lynott, 2010). Thus, only exogenous attentional mechanisms would be at work during a perceptual simulation, inducing, for instance, the automatic pre-activation of specific sensory modalities during reading. Automatically pre-activated specific modalities could then interfere with or facilitate the subsequent processing of semantic information yielding the MSE (see also Connell & Lynott, 2014). However, Pecher et al. (2003) had participants performing a double property verification task on each trial, one on the context and one on the target sentence. In addition, no time limit was provided to carry out the task. Therefore, participants were possibly lead to rely on strategic processes involving endogenous attention, such as constructing a mental image of the concept and property described in the sentences. Although mental imagery can be considered as "the best known case of […] simulation mechanisms, [it] typically results from deliberate attempts to construct conscious representations in working memory, [whereas] other forms of simulation often appear to become active automatically and unconsciously outside working memory" (Barsalou 2008, p. 619, see also Kiefer & Barsalou, 2013; Kiefer & Pulvermüller, 2012). For instance, Pulvermüller et al. (2000) showed that semantic activation in the sensorimotor cortex in passive reading tasks was present ~ 200 ms after word onset which would reflect stimulus-triggered early lexico-semantic processes (i.e. simulation) rather than post-lexical processes (i.e. imagery, see also Pulvermüller, 2005; on the generation of mental images see Farah, Weisberg and Monheit, 1989). Since in Pecher et al.'s paradigm participants had to perform a property verification task also on the context sentence and each sentence was presented until a response was given, one could reasonably argue that post-lexical processes involving endogenous attentional mechanisms could explain the MSE.

As to the second criticism, van Dantzig and colleagues (2008) sought evidence for the involvement of sensory information in conceptual processing that could not be explained by amodal symbols. According to amodal symbols accounts of concepts (Barsalou & Hale, 1993; Collins & Quillian, 1969; Smith & Medin, 1981), modal representations are turned into abstract, amodal symbols that represent knowledge about experience. Although being amodal, these symbols might still be organized so that to reflect their modality. The MSE with conceptual representations (Pecher et al., 2003) could hence hinge on connections between these symbols. van Dantzig et al. (2008) investigated the effect of a perceptual task such that of Spence and colleagues on a conceptual task such that of Pecher and colleagues. More specifically, the authors asked participants to perform a perceptual left/right spatial discrimination task followed by a conceptual property verification task, with the latter used as the target task. On each trial, participants first detected left/right visual, auditory or tactile signals (i.e., a light flash, a tone or a vibration), as in Spence et al. (2001), and then judged whether a visual, auditory or tactile property was true of a concept, as in Pecher et al. (2003).

Results indicate that participants were faster at verifying whether a property was true of a concept if that property was in the same sensory modality as the immediately preceding perceptual signal. Hence, participants, for example, were faster at verifying that *BABY CLOTHES* are *pastel* if they previously detected a light flash rather than a tone or a vibration. This finding provides evidence that pure perceptual processing (i.e., perceiving stimuli without any semantic meaning) affects the activation of conceptual processing. Since no meaningful relationship existed between the perceptual signals of the first task and the concepts of the second task the authors could conclude that the MSE cannot be explained by amodal symbols.

The present study aims at investigating whether the MSE is the result of the activation of sensory information when exogenous attentional mechanisms are involved. To this end, we introduced two key modifications of Pecher et al. (2003) and van Dantzig et al.'s (2008) studies. First, we implemented a priming paradigm in which context sentences required no task and were presented for a limited amount of time (from now on we will refer to these sentences as "prime sentences"). By using such a priming paradigm we aimed at preventing participants from deliberately drawing upon strategic processing for comprehending prime sentences. Our aim was to rule out the possibility that the involvement of sensory information in language comprehension was the consequence of a late post-lexical strategy to imagine objects and objects properties. Given that recent studies (Trumpp et al., 2013 and 2014) showed that subliminally presented sound and action words can activate auditory and motor systems, we reasonably hypothesized to find the MSE although no instructed task was required on prime sentences presented for a limited amount of time.

The second key difference is that we used prime sentences that made a linguistic description of the pure perceptual stimuli used in Spence et al. (2001) and in van Dantzig et al. (2008) studies so that to exclude connections between amodal symbols as a possible explanation of the effect. Given that language comprehension involves the construction of a perceptual simulation (Barsalou, 1999; Zwaan, 2004), and that perceptual simulations only involve exogenous attentional mechanisms, it is likely that reading or listening to a linguistic description of a pure perceptual stimulus could pre-activate specific sensory modalities, which could then facilitate or interfere with the processing of subsequent semantic information.

Moreover, in order to avoid any possible semantic association between prime and target sentences, concepts were either semantically unrelated or low semantically related. In Appendix A we report a norming study we have conducted to assess semantic relatedness of our stimuli (see also Marques, 2006 on the effects of semantic relatedness). To illustrate, our participants were first presented with a prime sentence describing a *LIGHT* or a *SOUND*'s perceptual property (e.g., "*LIGHT* is *flickering*"; "*SOUND* is *echoing*") and then with a target sentence (e.g., "*BUTTER* is *yellowish*", "*BRUSHWOOD crackle*") upon which a property verification judgment was to be made.

Finally, we also included a further manipulation by introducing neutral prime stimuli, that is, prime stimuli which did not convey any sensory information. Our purpose was to compare performances on target sentences preceded by sensory information (i.e., visual and auditory prime sentences) with performances on target sentences that were not preceded by sensory information (i.e., neutral prime sentences). Since neutral prime items were not expected to trigger a perceptual simulation, that is, they were not expected to involve any attentional mechanisms which could pre-activate a specific sensory modality, we either predict neither facilitation nor interference due to the fact that participants were unable to pre-activate a sensory modality.

We ran an Experiment in which prime and target sentences conveying both visual and auditory contents were presented either visually or aurally[19]. We predicted to find the MSE even with this modified property verification paradigm. In other words, we expected to find a better performance when prime

---

[19] Initially there were two distinct experiments to this study. Experiment 1 in which prime and target sentences were presented visually on the monitor, and Experiment 2 in which prime and target sentences were presented aurally through headphones. However, for purposes of exposition, the two experiments have become one with two conditions (visual, auditory).

and target sentences share the same modality compared to when they do not.

Method

Participants

Sixty-four students from the University of Bologna (43 females; mean age: 20.26, SD: 1.58) participated in this study in return for course credits. All participants were Italian native speakers, had normal or corrected-to-normal vision and hearing by self-report, and were naïve as to the purpose of the experiment. Participants were randomly assigned to one of the two between-subjects conditions (visual vs. auditory). The experiment was approved by the Psychology Department's ethical committee of the University of Bologna. Written informed consent was obtained from all individual participants included in the study. Minors did not take part in this study.

Materials

Prime items

We used 96 prime items. Forty-eight consisted of 24 visual and 24 auditory concept-property pairs. The concepts presented were always "*LIGHT*" and "*SOUND*" and the properties were adjectives associated with them (e.g., "*flickering*/*echoing*", for the visual and auditory concepts, respectively). Twenty-four properties were used, 12 for the visual and 12 for the auditory prime sentences. These properties were repeated once throughout the experiment. Twenty properties out of 24 were taken from the norming study by Lynott and Connell (2009), who classified several object's properties along a unimodality – multimodality continuum. The twenty properties we selected from their pool were all unimodal, being mainly perceived either through the sense of sight or through the sense of hearing. Lynott and Connell (2009) found indeed that using unimodal properties instead of multimodal ones leads to a markedly larger MSE. Since our experimental design needed 24 proprieties, following Lynott and Connell's (2009) combined criterion, we selected four further properties after 50 Italian adjectives had been rated by 22 participants (see Appendix A). For an overview of the visual and auditory prime sentences see Appendix B.

The other 48 prime items consisted of neutral stimuli, that is, for the visual condition a meaningless strings of symbols (e.g. # ° ^ ? *) and for the auditory condition a white noise. Both served to create a neutral modality compared to the same and different ones.

Target sentences

We used 96 target sentences: forty-eight critical target sentences, consisted of half visual and half auditory concept-property pairs taken from the van Dantzig et al.'s study (2008). In these critical pairs the property was always true of the concept (e.g., "*BUTTER* is *yellowish*", "*BRUSHWOOD crackles*"). Each pair was used only once in both the visual and auditory condition of the experiment. Two properties were repeated once across the pairs, although paired with different concepts (i.e., "a *BEE buzzes*", "a *FLY buzzes*"; "*BROCCOLI* is *green*", "*SPINACH* is *green*"). For an overview of the visual and auditory target sentences see Appendix B. The remaining forty-eight stimuli were filler sentences, always taken from van Dantzig et al. (2008). In the filler sentences the property was always false of the concept. Twelve filler sentences had a false visual property (e.g., "the *WATER* is *opaque*"), 12 had a false auditory property (e.g., "the *COMB sings*"), whereas the remaining 24 filler sentences had a false property that did not belong to any modality (e.g., "the *BED* is *sleepy*"). This latter type of fillers was used in order to avoid participants from basing their answers on a superficial word-association strategy, rather than on deeper conceptual-processing (see also Solomon & Barsalou, 2004).

For both the visual and auditory condition, each participant was presented in total with 96 prime sentences (48 modal and 48 neutral) followed by 96 target sentences (48 critical and 48 filler) throughout the experimental session. Prime and target items were randomly combined to form same, different and neutral modality conditions. Each target sentence appeared in the same, different and neutral modality conditions, counterbalanced across lists. This resulted in a comparable distribution of semantic relatedness and stimulus size measures across experimental conditions. To sum up, the critical targets could be combined with: 1) a neutral prime item (neutral modality); 2) a same-modality prime sentence (visual-visual; auditory-auditory, same modality) or 3) a different-modality prime sentence (visual-auditory, auditory-visual, different modality).

Procedure

The experiment took place in a dimly lit room. Participants sat in front of a computer screen, at a distance of about 60 cm. For the visual condition, each trial started with the presentation of a fixation cross (0.5 cm x 0.5 cm) for 500 ms. Immediately after the fixation, the prime sentence appeared in the middle of the screen for 1500 ms. Then, the target sentence was displayed on the center of the screen until a response was given or until 3000 ms had elapsed. Prime and target sentences ranged from 5.9 cm to 17.3 cm (from 9 to 29 characters) which resulted in a visual angle range between 5.6° and 16.5°. All words were bold lowercase Courier new 18. These measures were the same across all conditions. Participants were instructed to read the prime and target sentences and then judge, as quickly and accurately as possible, whether in each target sentence the property was true of the concept. Participants underwent a short practice session of 24 stimuli (different from those used in the experimental blocks), during which a feedback was given about their response. For the auditory condition, the procedure was the same, except that (1) a “bip” sound was presented in alternative to the fixation cross in order to announce the beginning of a new trial; (2) the prime and the target sentences were presented aurally, through headphones, for 2000 ms and 4000 ms respectively. In both the visual and the auditory condition, half of the participants pressed the “s” and “k” keys of a “qwerty” keyboard when the property was respectively true and false of the concept, that is, when the target was a critical or a filler sentence, respectively. The other half of the participants was assigned to the reverse mapping.

In order to control for sequence effects, we avoided to present the same modality for more than two consequent trials. For example, a prime sentence in the visual modality could be followed by another visual prime sentence only once. Then an auditory or neutral prime had to be shown. The same rule held for the target sentences. Two different sequences, composed of the same 192 concept-property pairs, were built. In both visual and auditory conditions, the sequence presentation was balanced across participants, such that half of the participants was presented with one sequence and the remaining ones with the other.

Participants underwent two blocks of 48 prime sentences followed by 48 target sentences each (24 critical and 24 filler) and could take a short break between them. The experiment lasted approximately 20 minutes.

Results

Responses to filler sentences were discarded. Omissions (3.74%), Incorrect responses (17.90%) and RTs faster/slower than the overall participant mean minus/plus 2 standard deviations (3.61%) were excluded from the analyses.

Mean Response Times (RTs) of the correct responses were submitted to a Repeated Analysis of Variance (ANOVA) with *Modality* (same vs. different vs. neutral) as the within-subject factor and *Condition* (visual vs. auditory) as the between-subjects factor (see Table 1 and Figure 8 for the results).

Results indicated that the main effect of *Modality*, F(2, 124) = 58.32, MS_e = 13302.62, p< .001, η_p^2 = .485, was significant. Paired-sample t-tests showed that decision latencies for same modality targets (M = 2000 ms, SD = 502.73 ms) were shorter than for different modality targets (M = 2070 ms, SD = 535.67 ms) t(63) = 5.7, p <.001 and decision latencies for neutral modality targets (M = 2216 ms, SD = 616.95 ms) were longer than for both same and different modality targets t(63) = 8.1, p <.001, t(63) = -6.4, p <.001.

The main effect of *Condition*, F(1, 62) = 320.32, MS_e = 146787.41, p< .001, η_p^2 = .838, resulted as significant, showing that the auditory condition was slower than the visual one (2590 ms vs. 1660 ms, respectively). However, it is worth mentioning that this result is due to a technical specification in the procedure: aurally presented prime and target sentences lasted longer than the visual ones, considering that spoken sentences need to be listened to until the end before participants could be able to release a response, while visually presented sentences were completely available at once.

The interaction between the *Modality* and *Condition* factors was significant, F(2, 87.1) = 7.88, MS_e = 18941.26., p< .001, η_p^2 = .113.

Paired-sample t-tests in the visual condition showed that decision latencies for same modality targets were faster than for different modality targets t(31) = 3.2, p <.01, whereas decision latencies for neutral modality targets were slower than for both same and different modality targets t(31) = 6.5, p <.001, t(31) = -4.1, p <.001. Similarly, paired-sample t-tests in the auditory condition showed that decision latencies for same modality targets were faster than for different modality targets t(31) = 4.9, p <.001, whereas decision latencies for neutral modality targets were slower than for both same and different modality targets t(31) = 6.5, p <.001, t(31) = -5.4, p <.001. In order to investigate the difference between the magnitude of the MSEs found, we run an additional Univariate analysis of Variance with the magnitude of the MSEs as dependent variable and the *Condition* as the only between-subjects factor. The magnitude of MSEs was computed by subtracting the mean RT for the same modality from the mean RT for the different modality. Results showed that the MSE found for the visual condition (50 ms) did not differ from the one found for the auditory one (90 ms), F(1, 62) = 2.8, p= .10, η_p^2 = .043.

In order to exclude a speed accuracy trade-off, mean of the incorrect responses and omissions were submitted to an ANOVA with the same factors as those of the RTs analysis. As to the incorrect responses, neither the main effects, nor the interaction were significant, $F_s < 2$, $p_s > .74$, $\eta_p^2 < .004$. As to the omissions, results indicated that the main effect of *Modality* $F(2, 124) = 11.32$, $MS_e = 36.22$, $p < .001$, $\eta_p^2 = .155$ was significant. In addition, the interaction between *Modality* and *Condition* was significant $F(2, 124) = 4.31$, $MS_e = 51.35$., $p < .05$, $\eta_p^2 = .065$. Paired sample t-tests showed that in the visual condition participants made more omissions in the neutral modality (3.7%) than in the different one (1.4%), $t(31) = 2.5$, $p < .05$. While in the auditory condition all the comparisons resulted significant showing that participants made less omissions in the same modality (1.3%) than in the different one (3.2%), $t(31) = 2.3$, $p < .05$, whereas omissions in the neutral modality (7.6%) outreached omissions in both different and same modalities $t(31) = 2.6$, $p < .05$, $t(31) = 3.3$, $p < .05$.

General Discussion

The goal of the current study was to investigate whether the MSE is the result of the activation of sensory information due to exogenous attentional mechanisms. We used a different paradigm from previous studies in order to exclude strategic processing and amodal symbols accounts of concepts as possible explanations of the effect. In line with the hypotheses, our findings showed a robust MSE, that is, a facilitation for the processing of those target sentences the modality of which was formerly primed by a linguistically described perceptual stimulus. In other words, participants were faster when they responded to a target sentence in the same modality as the previous prime sentence rather than different. These results confirm that when the target's modality correspond to the one pre-activated by the content of the prime sentence, RTs are speeded, while when these modalities do not correspond the time needed to complete the task is slowed down.

It is worth noting that our findings also showed slower RTs and a higher percentage of omissions for the neutral modality compared to the different modality. One might argue that the different modality could be expected to be the slowest modality. Indeed, activating information that does not correspond with what has to be processed later (i.e., different modality) should interfere with the processing of subsequent information and, thus, should require longer response times overall. However, the slowest performances observed with the neutral modality were possibly due to the fact that in this case the prime items (i.e. meaningless strings of symbols or white noise) were perceptually non informative. Unlike the visual and auditory prime sentences, the neutral prime did not pre-activate any specific sensory modality, neither correspondent nor non-correspondent. If the account for the MSE and the hypothesis that a neutral prime do not pre-activate any sensory modalities are correct, we could assume that the neutral prime did not trigger any perceptual simulation. Since no perceptual simulation took place with neutral prime items, participants could not take advantage of a general activation of the sensory system and this consequently resulted in an overall delay and a higher occurrence of omissions in the processing of the target sentences. This result is in line with a recent study by Connell and colleagues (Connell, Lynott & Dreyer, 2012), in which the conceptual processing of non-manipulable objects (e.g., cars or windmills) was not influenced by either a prior tactile or proprioceptive stimulation, showing that perceptually informative stimuli implied no facilitation effect but rather slowed down the response time needed to complete a task on perceptually non informative stimuli.

A potential concern is that participants could rely on a word association strategy to perform the property verification task upon target sentences. However, in the current experiment participants could not carry out a superficial processing of stimuli, using only word-level representations, for at least two reasons. First, the semantic domains across prime and target sentences were very distant to allow for a word association strategy (see also Marques, 2006): while target sentences described perceptual properties of objects, prime sentences described properties of two perceptual categories (i.e., light and sound), hence no main semantic association was available across them. In addition, in order to avoid participants using the word association strategy, we drew upon highly associated concepts and properties on false trials (i.e., fillers). Indeed, previous studies (James, 1975; McCloskey, & Glucksberg, 1979; Smith, Shoben & Rips, 1974) showed that manipulating the difficulty of false trials varies the depth of processing on true trials (see also Solomon & Barsalou, 2004). Therefore, rather than have participants reject unassociated false properties for concepts in the filler trials (e.g., *unripe* for *BED*), we had participants reject associated ones (e.g., *sleepy* for *BED*). For associated false properties, participants could not respond “false” on the basis of the word association strategy because the concept and the property were actually somehow associated (i.e., sleepy people go to bed). Thus, in order to determine whether the property was true of the concept, participants must access conceptual knowledge for *BED* and *sleepy* and realize, for instance, that rather than being sleepy a bed is used by sleepy people.

Overall, the results of the present study boost and broaden previous findings which showed a significant MSE during an on-line perceptual task (Spence et al., 2001), a property verification task (Pecher et al., 2003) and across perceptual and conceptual tasks (van Dantzig et al., 2008). More broadly, our results support the accounts of the role of perceptual attention on conceptual processing (Connell & Lynott, 2010, 2012, 2014) showing that exogenous attentional mechanisms are at work during perceptual simulation and are responsible for the MSE. Although we cannot completely rule out that the MSE we found is due to strategic or imagery processes, the use of a standard priming paradigm represents an important difference compared to previous work. Indeed, while in previous studies the sensory modality was likely to be strategically activated when performing the task on the context sentence, in our experiment we found a MSE even though participants were not required to perform any task on the prime sentences. That is, in our experiment it was completely unnecessary to directly and explicitly pre-activate a specific sensory modality, therefore the MSE we found is likely to be due to an implicit and indirect pre-activation of sensory modalities. Ultimately, we showed that the MSE also occurs when participants are prevented from drawing upon strategic processing, furthering the hypothesis that the MSE arises from a simulation process during which exogenous attention operates. In addition, we showed that not only a perceptual stimulus (van Dantzig et al., 2008) but also a perceptual linguistically described stimulus triggers the pre-activation of a sensory modality: reading or listening to a sentence describing a light or a sound's perceptual property sufficed to spark off a simulation, even though no task was required on that sentence.

In conclusion, the simulation of an object varies considerably across occasions. When reading or listening to a sentence involving a particular object in a certain situation, implicit perceptual attention (i.e., exogenous attention) activates a specific modality. If that modality had been previously activated by either a perceptual stimulus or a perceptual linguistically described stimulus, the processing of semantic information that relates to that modality in the sentence is facilitated. This is far from implying that any given object does only relate to a certain modality. Rather, other relevant modalities might be temporarily inhibited. In facts, modalities represented in simulations vary on the basis of their activation. Future exploration of the MSE could use this modified property verification paradigm with multimodal concepts in order to investigate what happens when multiple modalities compete during a simulation.

Table and Figure

Table 1: Mean Response Times (in Milliseconds) with Standard Deviations in parenthesis, as a Function of *Modality* (same, different, neutral) for both visual and auditory conditions. The MSE is computed by subtracting response times in the same modality from response times in the different modality. Asterisks denote significant differences.

	Visual Condition	**Auditory Condition**
Same	1538 (178.3)	2462 (202.2)
Different	1588 (206.5)	2552 (245.4)
Neutral	1676 (222.2)	(349.1)
MSE	50*	90*

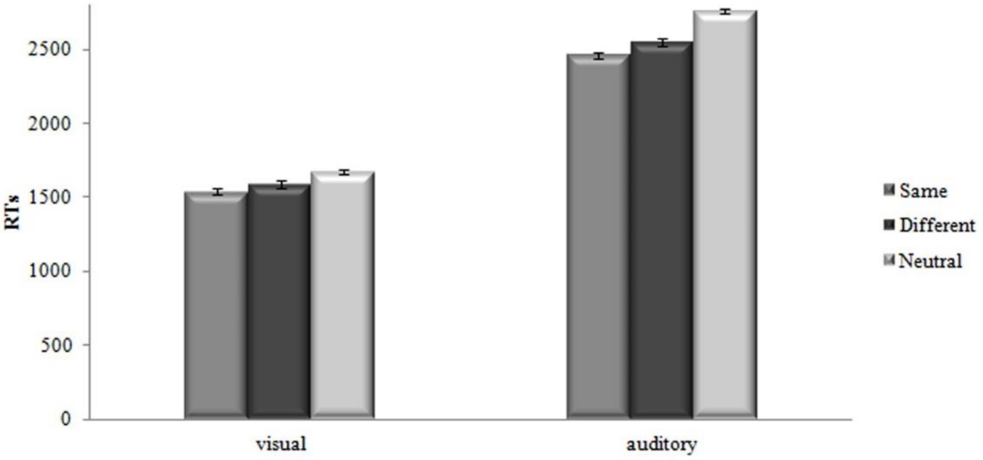

Figure 8: Mean Response Times (in Milliseconds) as a Function of *Modality* (same, different, neutral) for both visual and auditory conditions. Bars are standard Errors.

Appendix A

Rating of the 50 Italian adjectives

A set of 50 words was selected from the Italian dictionary Sabatini-Coletti. Each word (either an adjective or present participle of a verb) could belong to the auditory or to the visual sensory modality (e.g., ritmato, accecante).

The rating was administered with an on line procedure. Twenty-two participants, all Italians native speakers (13 females; mean age: 23 years old; sd. 4 years) were tested. To avoid order presentation effects, participants were split into two equal groups having two different orders of item presentation. Participants' task was to rate, on 5 separate 5-points Likert scales (where 1 = not at all; 5 = greatly) the extent to which each item is experienced through each of the five senses. For instance, participants were faced with the item "bright" and were asked to rate to what extent they experienced this property by the touch, the hearing, the sight, the sense of smell and by the taste.

Participants' average score for each item in each modality was computed. In order to compute the modality exclusivity and modality strength indexes as in Lynott and Connell (2009) and compare our 5-points (1-5) Likert scale to their 6-point (0-5) scale we applied the following conversion formula: (5/4* the score obtained for the item in each modality) – 5/4. This allowed us to keep the threshold of the modality strength and the modality exclusivity to 3.5 and .65 respectively. Four items (2 visual, 2 auditory) scored strong on modality strength (3.5-5) and high on modality exclusivity (65%-100%) showing their unimodality.

Appendix B

Overview of the visual and auditory prime and target sentences

Visual Prime Stimuli	**Visual Prime Stimuli**	**Visual Target Stimuli**	**Visual Target Stimuli**
English version	*Italian translation*	*English version*	*Italian translation*
Light is bright	La luce è intensa	Butter is yellowish	Il burro è giallognolo
Light is colourful	La luce è colorata	Broccoli is green	Il broccolo è verde
Light is dazzling	La luce è abbagliante	Chocolate is dark brown	Il cioccolato è marrone
Light is dim	La luce è soffusa	An eggplant is dark purple	La melanzana è viola
Light is flickering	La luce è intermittente	An inner tube is black	La camera d'aria è scura
Light is gleaming	La luce è splendente	A cassette tape is black	La musicassetta è nera
Light is glowing	La luce è raggiante	A diamond glistens	Il diamante brilla
Light is gold	La luce è dorata	A squirrel is red-brown	Lo scoiattolo è rossiccio
Light is shimmering	La luce è scintillante	An ice cube is transparent	Il cubetto di ghiaccio è trasparente
Light is translucent	La luce è luminosa	A cellar is dark	La cantina è buia
Light is blinding	La luce è accecante	A jellyfish is translucent	La medusa è lucida
Light is flashing	La luce è lampeggiante	A leopard is spotted	Il leopardo è maculato
		An orca is black-and-white	L'orca è bianca e nera
		Peppermint is white	La mentina è bianca
		A chessboard is chequered	La scacchiera è a quadri
		A razorblade is silver	La lametta è argentata
		A tennis ball is yellow	La palla da tennis è gialla
		A walnut is brown	La noce è bruna
		A wasp is striped	La vespa è striata
		A swimming pool is azure blue	La piscina è azzurra
		Ham is pink	Il prosciutto cotto è rosa
		Honey is golden-yellow	Il miele è ambrato
		Mayonnaise is light yellow	La maionese è giallina
		Spinach is dark green	Gli spinaci sono verdi

Auditory Prime Stipuli	**Auditory Prime Stipuli**	**Auditory Target Stimuli**	**Auditory Target Stimuli**
English version	*Italian translation*	*English version*	*Italian translation*
Sound is deafening	Il suono è assordante	A bee buzzes	L’ape ronza
Sound is echoing	Il suono è echeggiante	A flute is high-pitched	Il fauto è di tno acuto
Sound is husbed	Il suono è sommesso	A scooter hums	Il motorino strepita
Sound is loud	Il suono è alto	A bicycle bell rings	Il campanello trilla
Sound is mute	Il suono è muto	A church organ clangs	L’organo vibra
Sound is noisy	Il suono è rumoroso	A cricket chirps	Il grillo canta
Sound is shrill	Il suono è penetrante	A saxophone blares	Il sassofono è squillante
Sound is sonorous	Il suono è altisonante	A ship’s hom is low- pitched	Il fischio della nave è basso
Sound is squealing	Il suono è stridente	A siren wails	La sirena urla
Sound is husky	Il suoni è rauco	A station hall is noisy	La stazione è chiassosa
Sound is croaking	Il suono è gracchiante	A tram grinds	Il tram sferraglia
Sound is audible	Il suono è udibile	A triangle jingles	Il triangolo tintinna
		A trumpet sounds shrill	La tromba è stridula
		A typewriter rattles	La macchina da scrivere ticchetta
		A fly buzzes	La mosca ronza
		An airplan is loud	L’aereo è roboante
		A truck honks	L’autocarro strombazza
		An alarm clock ticks	L’orologio fa tic tac
		Autumn leaves rustle	Le foglie frusciano
		Brushwood crackles	Il sottobosco crepita
		High heels tap	I tacchi alti picchettano
		Thunder rumbles	Il tuono rimbomba
		Pans clang	Le pentole urtano fragorosamente
		A railroad crossing rings	Il passaggio alivello suona

V. STUDY 2: EXPERIMENTS 3 & 4

The Multilevel Modality-Switch Effect: What Happens when We See the Bees Buzzing and Hear the Diamonds Glistening[20]

Introduction

In people's everyday life, the majority of experiences involve multiple sensory modalities. We are thus required to be able to switch across different sensory modalities in different situations. A classic example involves the musicians in an orchestra: they are required to be able to quickly process visually presented auditory contents (i.e., sheet music along with the conductor's gestures) in order to perform. However, this only happens through years of studying. Indeed, recent research has shown that people experience a cognitive cost in shifting attention between different sensory modalities. Interestingly, such cognitive cost occurs both when switching between events presented in different modalities (Spence, Nicholls & Driver, 2001) as well as when switching between sentences having different-modality contents (Pecher, Zeelenberg & Barsalou, 2003). For example, switching from the sentence "*BLENDER* is *loud*" to the sentence "*BANANA* is *yellow*" incurs a processing cost much like switching from an auditory tone to a light flash. This phenomenon is known as Modality-Shifting or Modality-Switch Effect (hereafter MSE).

---

[20] The final publication is available at Springer via http://dx.doi.org/10.3758/s13423-016-1150-2.

The MSE with language has been extensively explored with both behavioral (Marques, 2006; Pecher, Zeelenberg & Barsalou, 2004; Scerrati, Baroni, Borghi, Galatolo, Lugli & Nicoletti, 2015; van Dantzig, Pecher, Zeelenberg & Barsalou, 2008; see also Vermeulen, Niedenthal, Luminet, 2007 for a similar result with emotional concepts) and ERP studies (Collins, Pecher, Zeelenberg, Coulson, 2011; Hald, Marshall, Janssen, Garnham, 2011; Hald, Hocking, Vernon, Marshall & Garnham, 2013). Whether the finding of a purely perceptual phenomenon during conceptual processing is just the epiphenomenal result of spreading activation or evidence supporting the assumption that perceptual information is engaged in conceptual processing is debated. On the one hand, it has been argued that the conceptual system is separated from sensory information (*disembodied cognition* hypothesis, see Mahon & Caramazza, 2008; Mahon & Hickok, 2016 for discussions). On this account, the MSE would reflect the way in which activation spreads throughout the system, therefore it would not reveal anything about semantic processing. On the other hand, it has been assumed that the perceptual and conceptual systems are tightly interwoven and share the same processing mechanisms. Proponents of the grounded accounts of knowledge (Barsalou, 2008; for recent reviews see Borghi and Caruana, 2015; Pecher, 2013) assume that knowledge representation and processing is achieved by reactivating aspects of experience. In particular, conceptual processing would imply constructing a sensorimotor simulation of the objects or events concepts refer to. Such a simulation would involve the partial reactivation of those brain areas that were also active during our interaction with the concepts' referents. For example, on processing the concept DOG, brain areas that represent visual, auditory, tactile, olfactory, gustatory, affective, and motor information about dogs would be liable to partial reactivation. Importantly, simulations are sketchy records of experience that can be flexibly adapted to the context and task at hand (Barsalou, 1999; Gallese, 2009).

Recently, Scerrati et al. (2015) obtained evidence that sensorimotor simulations can also be triggered by a perceptual, linguistically described stimulus presented in a sensory modality different from vision (i.e., the auditory modality). Participants were presented with a prime sentence describing a light or a sound's perceptual property (e.g., "The light is flickering", "The sound is echoing") then they were required to perform a property-verification task on a target sentence with a vision-related or a hearing-related content (e.g., "Butter is yellowish", "Leaves rustle"). The sensory modality activated by the content of the prime sentence could be compatible with the target's content modality (e.g., vision-vision: "The light is flickering" followed by "Butter is yellowish") or not (e.g., vision-audition: "The light is flickering" followed by "Leaves rustle"). Crucially, the stimuli's presentation modality was manipulated such that half of the participants was faced with written prime and target sentences while the other half was faced with spoken prime and target sentences. Results showed that participants were faster at judging whether a certain property was true of a given concept when the target's content modality corresponded to the one pre-activated by the content of the prime sentence with both visual and aural presentation of stimuli.

In the present study we were interested in examining whether switching between different mode of presentation (i.e., visual, aural) across prime and target sentences conveying a sensory content brings about a modality switching cost. Specifically, we aimed at understanding whether and how the conceptual MSE is modulated by the mode of presentation of stimuli. To our knowledge, no previous study has explored this issue in regard to the MSE. Interestingly, however, different studies found that sentence processing can be affected by mode of presentation. Kaschak, Zwaan, Aveyard and Yaxley (2006, Experiment 2) showed that participants were faster in making sensibility judgements on target sentences when the direction of motion implied by the sentence with a hearing-related content (e.g., "The commuter had just arrived on the platform when the subway roared into the station") and the direction of motion depicted by a concurrent auditory stimulus were the same, provided that both the sentence and the stimulus were aurally presented. In a different yet related study, Vermeulen, Corneille and Niedenthal (2008) showed that asking people to store three visual or auditory items (i.e., pictures or sounds) in short-term memory for a subsequent memory task resulted in a worse performance in an intervening property verification task when the latter concerned sentences involving properties in the same modality as that of the stored items (*interference* hypothesis). Vermeulen et al. (2008) suggested that the general attentional load imposed upon participants together with the high complexity of the dual-task paradigm used in their study moderated switching costs. On the basis of this previous evidence, we expect that the mode of presentation of sentences might be relevant in modulating the MSE. Specifically, given that neither we manipulate attentional load, nor we use a dual task paradigm, we expect to observe a facilitation when the prime and the target share the same presentation and content modality as in prior studies where switching costs were found.

Whether and how the conceptual MSE is affected by the mode of presentation of stimuli may hinge on task demands. Connell & Lynott (2014) found that task-specific implicit perceptual attention preactivates modality-specific systems leading to facilitated representation of semantic information related to those modalities. That is, preactivating the visual system through the presentation of strongly visual words (e.g., “cloudy”) facilitated performance in the lexical decision task, whereas preactivating the auditory system through the presentation of strongly auditory words (e.g., “noisy”) facilitated performance when the task was reading aloud. In the present research, we used two different tasks: the property verification and the lexical decision task[21] (LDT, McNamara, 1992). We

---

[21] This study too was initially conceived as a two experiments study. Experiment 1 used a property verification priming paradigm and Experiment 2 used a lexical decision priming paradigm. However, for purposes of

believe that the mode of presentation of stimuli might differently impact the conceptual MSE on the basis of the depth of processing required by the task. With the property verification task we predict to observe a better performance when the presentation and the content modalities of target sentences are congruent (e.g., “Butter is yellowish” presented visually) compared to when they are incongruent (e.g., “Butter is yellowish” presented aurally) due to the depth of processing required by the task. With a less conceptually engaging task such as the LDT, instead, we expect that the mode of presentation might feature prominently compared to the content modality of sentences.

Method

Participants

One hundred twenty-eight students from the University of Bologna (79 females; mean age: 21.45, SD: 2.37) participated in the experiment in exchange for course credit. Sixty-five participants were randomly assigned to the property verification task condition whereas 63 participants were randomly assigned to the LDT condition. All participants were Italian native speakers, had normal or corrected-to-normal vision and hearing by self-report, and were naïve as to the purpose of the experiment. The experiment was approved by the Psychology Department’s ethical committee of the University of Bologna.

---

exposition, the two experiments have become one with two conditions (property verification, lexical decision).

Written informed consent was obtained from all individual participants included in the study. Minors did not take part in the study.

Materials

Twenty-four prime sentences and forty-eight target sentences were used in this experiment. Stimuli were the same as in Scerrati et al. (2015). Half of the prime sentences had a vision-related content (e.g., “the *LIGHT* is *flickering*”) whereas the other half had a hearing-related content (e.g., “the *SOUND* is *echoing*”). Properties in the visual and auditory prime sentences were taken from the norming study by Lynott and Connell (2009) and from a rating of 50 Italian adjectives (see Appendix A in Study 1). Each of the 24 prime sentences was repeated four times throughout the experiment, twice they were aurally presented over closed-ear headphones and twice they were visually presented on the screen.

Target sentences were taken from the van Dantzig et al.’s study (2008). Twenty-four had a vision-related content (e.g., “a *WALNUT* is *brown*”) and twenty-four had a hearing-related content (e.g., “a *BEE buzzes*”). In these critical pairs the property was always true of the concept. Each pair was used only once. Two properties were repeated once across the pairs, although paired with different concepts (i.e., “a *BEE buzzes*”, “a *FLY buzzes*”; “*BROCCOLI* is *green*”, “*SPINACH* is *green*”). For an overview of the visual and auditory prime and target sentences see Appendix B in Study 1. Prime and target sentences were the same across tasks.

As for the property verification task, an additional set of 48 filler sentences, always taken from van Dantzig et al. (2008) was used. In the filler sentences the property was always false of the concept. Twelve filler sentences had a false visual property (e.g., "the *WATER* is *opaque*"), 12 had a false auditory property (e.g., "the *COMB sings*"), whereas the remaining 24 filler sentences had a false property that did not belong to any modality (e.g., "the *BED* is *sleepy*"). This latter type of fillers was used in order to avoid participants from basing their answers on a superficial word-association strategy, rather than on deeper conceptual-processing (see Solomon & Barsalou, 2004).

As for the LDT, an additional set of 48 filler sentences featuring a non-word was used. In half of the filler sentences the non-word was the concept word, whereas in the other half the non-word was the property word. Non-words were generated altering two of the consonants or the double consonant keeping unchanged the vowels so as to preserve the phonotactic rules of Italian.

Each participant was presented with 96 prime sentences followed by 96 target sentences (48 critical and 48 filler) throughout the experimental session. Prime and target sentences were randomly combined to form four modality conditions, that is: different-different (DD, when both the presentation and the content modalities do switch from prime to target sentence), different-same (DS, when the presentation modality does switch but the content modality does not), same-different (SD, when the content modality does switch but the presentation modality does not) and same-same (SS, when the prime and the target sentences share the same presentation and content modalities). For example, a visually presented prime sentences with a vision-related content (e.g., “the *LIGHT* is *flickering*”) could be combined with: 1) an aurally presented target sentence with a hearing-related content (e.g., “a *BEE buzzes*”, DD); 2) an aurally presented target sentence with a vision-related content (e.g., “a *WALNUT* is *brown*”, DS); 3) a visually presented prime sentences with a hearing-related content (e.g., “a *BEE buzzes*”, SD); 4) a visually presented prime sentences with a vision-related content (“a *WALNUT* is *brown*”, SS). Each target sentence appeared in all modality conditions, counterbalanced across participants.

Procedure

The stimuli were presented on a 17 inch monitor (1.6 Ghz refresh rate). The participants sat at a viewing distance of about 60 cm from the monitor in a dimly-lit room. They were invited to wear a pair of headband headphones before starting the experiment. Each trial started with the presentation of a fixation cross (0.5 cm x 0.5 cm) for 500 milliseconds (ms). Immediately after the fixation the prime sentence appeared on the screen or was delivered through headphones for 2000 ms. Then, the target sentence was displayed on the screen or delivered through headphones until a response was given or until 4000 ms had elapsed. Visually presented prime and target sentences ranged from 5.9 cm to 17.3 cm (from 9 to 29 characters) which resulted in a visual angle range between 5.6° and 16.5°. All sentences were bold lowercase Courier new 18 and were presented in black in the center of a white background. Participants were instructed to read or to listen to the prime and target sentences and then judge, as quickly and accurately as possible, whether in each target sentence the property was true of the concept (property verification task condition), or whether in each target sentence there was a non-word or not (LDT condition). In both task conditions, half of the participants pressed the “s” key of a “qwerty” keyboard when either the property was true of the concept or there was a non-word in the target sentence and the “k” key when either the property was false of the concept or the target sentence did not contain a non-word. The other half of the participants was assigned to the reverse mapping.

The order of presentation of each prime-target sentence was completely randomized across participants. Participants underwent a short practice session of 32 stimuli (different from those used in the experimental blocks) before starting the experiment. The experiment consisted of one block of 96 prime-target pairs and lasted approximately 15 minutes.

Results

In the property verification task condition, five participants (all females) were excluded from the analysis: Four of these participants failed to reach an accuracy score of 65% while the other participant responded 35% of the trials in less than 300 ms, indicating that she may have misconceived the task and tried to respond on the prime sentence also. Sixty participants therefore remained for further analysis. Responses to filler sentences were discarded. Omissions (5.93%), Incorrect responses (21.42%) and RTs faster/slower than the overall participant mean minus/plus 2 standard deviations (2.19%) were excluded from the analyses. In the LDT condition, three participants (two females) failed to reach an accuracy score of 65%. Their data were removed, leaving sixty participants for further analysis. Responses to filler sentences were discarded. Omissions (5.03%), Incorrect responses (7.04%) and RTs faster/slower than the overall participant mean minus/plus 2 standard deviations (2.60%) were excluded from the analyses.

Mean Response Times (RTs) of the correct responses were submitted to a Repeated Analysis of Variance (ANOVA) with *Mode of Presentation* (different vs. same), *Content Modality* (different vs. same) and *Target Congruency* (incongruent vs. congruent) as the within-subject factors for the two tasks (property verification vs. lexical decision) separately. Data are shown in Table 2.

In the property verification task condition there was a main effect of *Mode of Presentation*, $F(1, 59) = 4.582$, $MS_e = 75789.90$, $p < .05$, $\eta_p^2 = .072$, that is, decision latencies were faster when the *Mode of Presentation* was the same across prime and target sentences rather than different (M: 2036 ms vs. 2090 ms). The analysis also revealed a main effect of *Target Congruency* ($F(1, 59) = 18.633$, $MS_e = 65906.25$, $p < .001$, $\eta_p^2 = .240$), that is, decision latencies were faster when the *Mode of Presentation* and the *Content Modality* of the target were congruent rather than incongruent (M: 2013 ms vs. 2114 ms). No other main effect or interaction turned out to be significant, $F_s < 2.66$, $p_s > .108$.

In the LDT condition, there was a main effect of *Mode of Presentation*, $F(1, 59) = 6.544$, $MS_e = 59889.45$, $p < .05$, ηp2 = .1, that is, decision latencies were faster when the *Mode of Presentation* was the same across prime and target sentences rather than different (M: 1942 ms vs. 1999 ms). No other main effect or interaction turned out to be significant, $F_s < 1.420$, $p_s > .238$.

Mean of the incorrect responses were submitted to an ANOVA with the same factors as those of the RTs analysis. In the property verification task condition no main effect or interaction turned out to be significant, $F_s < 2.247$, $p_s > .139$. In the LDT condition there was a significant interaction between Mode of Presentation and Target Congruency, $F(1,59) = 4.484$, $MS_e = 140.56$, $p < .05$, $\eta_p^2 = .071$. Paired-sample t-tests showed that percentage of ERs was higher when the *Mode of*

Presentation was the same across prime and target sentences and the target was congruent compared to different and incongruent (9.5% vs. 6.8%), $t(59) = -2.075$, $p <.05$ same and incongruent (9.5% vs. 5.8%), $t(59) = -2.225$, $p < .05$ and different and congruent (9.5% vs. 5.9%), $t(59) = -2.327$, $p < .05$. No other main effect or interaction turned out to be significant, $F_s < 2.683$, $p_s > .107$.

General Discussion

The present research investigated whether and to what extent switching between different modes of presentation (i.e., visual, aural) across prime and target sentences affects the conceptual MSE. Although previous studies investigated how sentence processing can be affected by mode of presentation of linguistic stimuli, such relationship had not previously been studied in the context of the MSE. Given that the impact of the mode of presentation of stimuli on language processing may be modulated by task demands (see Connell and Lynott, 2014 for a similar result in a different context), we compared performance on a property verification priming paradigm with performance on a lexical decision priming paradigm, each involving different levels of conceptual processing.

In keeping with our hypothesis, we found evidence for the involvement of the mode of presentation of stimuli in both the property verification and the lexical decision task. Crucially, results from both tasks showed that the presentation-driven effect weakens the conceptual MSE. Indeed, a conceptual MSE was observed in the property verification task, but not in LDT, as expected; however, it did not reach significance.

Interestingly, the property verification task highlighted an effect of the target congruency. That is, we found that participants were slower in deciding whether a certain property was true of the concept when the presentation and the content modality were incongruent for the target (e.g., “a BEE buzzes” presented visually) compared to when they were congruent. Such a within-target MSE is in line with the results of van Dantzig et al. (2008) showing that when a perceptual stimulus (i.e., a light flash, a tone or a vibration) and a subsequent target sentence were in a different sensory modality decision latencies were slower compared to when they were in the same modality. Our results broaden their finding showing such an effect within the same stimulus, that is, when the processing of perceptual and conceptual information overlaps in time. It is worth noting that such interference only occurred with the property verification task. Therefore, it seems likely that since the lexical decision task did not emphasize conceptual processing, it only recruited the semantic system to a certain extent insufficient to generate interference between the two systems.

In sum, our findings show that conceptual processing is not only affected by switching between sensory modalities on a semantic level (i.e., content modality of stimuli) but also by switching between sensory modalities on a purely perceptual level (i.e., mode of presentation of stimuli). Interestingly, our results also demonstrate a task-dependent, complex interplay of perceptual and semantic information taking place within the target. These findings question the view according to which the MSE does not reveal anything about semantic processing as claimed by the opponents of the grounded accounts of knowledge (Mahon & Caramazza, 2008).

We conclude that the MSE is a task-related, multilevel effect which can occur on two different levels of information processing, i.e., perceptual and semantic. We interpret these results as further evidence supporting the likelihood that the perceptual and conceptual systems are tightly interwoven and share the same processing mechanisms as claimed by the simulation account of conceptual processing (Barsalou, 1999, 2003, 2008).

Table

Table 2: Mean Response Times (in Milliseconds) and Percentage of Errors with Standard Deviations in parenthesis as a Function of *Mode of Presentation* (MoP: Different, Same), *Content Modality* (CM: Different, Same) and *Target Congruency* (TC: incongruent, congruent) for each tasks separately.

	Property verification				**Lexical decision**			
	RT		**ERs**		**RT**		**ERs**	
	D	**S**	**D**	**S**	**D**	**S**	**D**	**S**
MoP	2090 (338.6)	2036 (340.2)	21 (17)	21 (15.9)	1999 (337.9)	1942 (268.8)	6 (10.4)	7 (10.9)
CM	2080 (357.2)	2046 (321.6)	20 (16.4)	22 (16.5)	1972 (309.7)	1968 (297)	7 (10.9)	6 (10.4)
TC	**I**	**C**	**I**	**C**	**I**	**C**	**I**	**C**
	2114 (340.5)	2013 (338.3)	22 (16.2)	20 (16.7)	1972 (304.7)	1968 (302)	6 (10)	7 (11.3)

CONCLUDING REMARKS

This dissertation has motivated and defended an analysis of the format of concepts. The analysis gives a treatment of the ongoing theoretic debate on the format of concepts, contrasting important classes of theories weighing in on the debate: the amodal, grounded and hybrid accounts of knowledge. Importantly, the analysis also explores predictions coming from such accounts on experimental grounds, testing the scope and robustness of the Modality-Switch Effect (MSE), a cost for performance in terms of speed and accuracy occurring when two different sensory modality properties for concepts alternate compared to when the same sensory modality property is presented.

The theoretic and experimental investigations developed in this dissertation provide new insights in order to construct a unified account of the nature of concepts that emphasizes important mechanisms of brain organization and functioning ensuing from all of the examined theories. In particular, the modality-specific organization of knowledge common to amodal sensory/functional and grounded theories of knowledge together with the reuse of sensory-motor circuitry for the representation of concepts typical of the grounded framework, and the topographical mapping between the physical structure of the world and the spatial organization or topography of the brain's association areas developed in the context of the hybrid accounts of knowledge (the *conceptual topography* theory, CTT) all seem very important principles of neural organization and functioning.

Perhaps the most surprising results comes from the experimental investigation of the MSE carried out in this dissertation. Four experiments were devised to investigate different predictions coming from both amodal and grounded perspectives. First, if conceptual processing exploits sensorimotor systems, then a cost similar to that found in the perception literature should occur when verifying different-modality properties for concepts. Study 1 confirms this prediction, replicating Pecher et al.'s (2003) results, and further demonstrates that the MSE is due to exogenous attentional mechanisms that automatically activate sensory information during a perceptual simulation.

Second, if the MSE was due to symbols in an amodal symbols system organized so that to reflect their modality, then results were not supposed to show a MSE when participants were not required to perform any task on the prime sentences. That is, in these experiments (Study 1) it was completely unnecessary to directly and explicitly pre-activate a specific sensory modality. Nevertheless, participants could not avoid it. Therefore, the MSE found by these studies is likely to reflect an automatic pre-activation of sensory modalities as it is expected to occur during a perceptual simulation.

Third, the discovery that the MSE also emerges with spoken sentences that automatically triggers the pre-activation of a sensory modality constitutes an innovative demonstration that the effect arises during both reading and speech processing. Thus, not only reading but also listening to a sentence describing perceptual properties suffice to spark off a simulation as shown by Experiment 2 (Study 1).

Fourth, if the MSE does not reveal anything about semantic processing as claimed by the opponents of the grounded accounts, then one could not explain results from Study 2 (Experiments 3 & 4). Specifically, the finding that the mode of presentation of stimuli affects the MSE in both the property verification (Experiment 3) and the lexical decision task (Experiment 4) is intriguing and difficult to accomodate in the context of amodal accounts. Crucially, results from both tasks showed that the presentation-driven effect weakens the conceptual MSE leading to the conclusion that the MSE is a multilevel effect. Interestingly, these results demonstrate a task-dependent, complex interplay of perceptual and semantic information not easy to account for in the amodal framework.

Overall, although it is possible that an amodal conceptual hub exists in the brain, this does not necessarily entail that all conceptual knowledge has an amodal format. The experiments reported in this dissertation show that perceptual mechanisms are involved in conceptual processing. Hybrid solutions seems the most promising given extant data from behavioral, neuroimaging and neuropsychological research.

Of course, a number of problems concerning the format of concepts remain unsolved. For example, we cannot determine the representational format of specific brain regions (Martin, 2016). Nevertheless, we can hypothesize that conceptual processing use modality-specific resources whatever format concepts have (Barsalou, 2016).

In sum, I hope to have shown that the research on the representation and processing of concepts in the brain can play an important role for discovering new and important cognitive mechanisms underlying the organization of knowledge in the human brain. Understanding how we organize our knowledge is of crucial importance given that without an organization it is easy to imagine how our mental life could turn in a chaotic muddle that would prevent any kind of thought, activity and not last consciousness.

Bibliography

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

Arbib, M.A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105–67.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, 38, 347–358.

Barclay, J.R., Bransford, J.D., Franks, J.J., McCarrell, N.S., and Nitsch, K.E. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 13(4), 471-481.

Barlow, H. (1972). Single units and cognition: A neuron-doctrine for perceptual psychology. *Perception*, 1(4), 371-394.

Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, 10(1), 82-93.

Barsalou, L.W. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. In U. Neisser (Ed.), *Concepts and Conceptual Development:*

Ecological and Intellectual Factors in Categorization. Cambridge: Cambridge University Press.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.

Barsalou, L.W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18:5-6, 513-562.

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

Barsalou, L.W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society*, 364, 1281–1289.

Barsalou, L.W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science*, 2, 716–724.

Barsalou, L.W. (2016). On staying grounded and avoiding Quixotic dead ends. *Psychonomic Bulletin & Review*, 23, 1122-1142.

Barsalou, L.W., Breazeal, C., & Smith, L.B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing*, 8, 79–91.

Barsalou, L.W., and Hale, C.R. (1993). Components of conceptual representations: From feature lists to recursive frames. In: *Categories and concepts: Theoretical views and inductive data analysis*. E. van

Mechelen, J. Hampton, R. Michalski & P. Theuns (eds). Academic 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, 84–91.

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

Binder, J.R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, 23(4), 1096-1108.

Binder, J.R., & Desai, R.H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, 15(11), 527-536.

Bonner, M.F., Peelle, J.E., Cook, P.A., & Grossman, M. (2013). Heteromodal conceptual processing in the angular gyrus. *Neuroimage*, 71, 175-186.

Borghi, A.M., & Caruana, F. (2015). “Embodiment theory”. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, Oxford: Elsevier, Vol 7, pp. 420-426.

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

Borgo, F., & Shallice, T. (2003). Category specificity and feature knowledge: Evidence from new sensory-quality categories. *Cognitive Neuropsychology*, 20, 327–353.

Boulenger, V., Roy, A.C., Paulignan, Y., Deprez, V., Jeannerod, M. et al. (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.

Bowers, J.S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, 116(1), 220-251.

Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley: University of California Press.

Buckner, R.L., Petersen, S.E., Ojemann, J.G., Miezin, F.M., Squire, L.R. & Raichle, M.E. (1995). Functional anatomical studies of explicit and implicit memory retrieval tasks. *Journal of Neuroscience*, 15, 12–29.

Chao, L.L., Haxby, J.V., & Martin, A. (1999). Attribute-based neural substrates in posterior temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2, 913-919.

Chao, L.L., & Martin, A. (2000). Representation of manipulable manmade objects in the dorsal stream. *NeuroImage*, 12, 478-484.

Chao, L.L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category related cortical activity. *Cerebral Cortex*, 12, 545–551.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123 (Pt. 2), 291–307.

Collins, A.M., and Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428.

Collins, A.M., and Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 240-247.

Collins, J., Pecher, D., Zeelenberg, R., Coulson, S. (2011). Modality switching in a property verification task: an ERP study of what happens when candles flicker after high heels click. *Frontiers in Psychology* 2:10. doi: 10.3389/fpsyg.2011.00010

Connell, L. and Lynott, D. (2010). Look but don’t touch: Tactile disadvantage in processing modality specific words. *Cognition* 115, 1–9.

Connell, L. and Lynott, D. (2012). When does perception facilitate or interfere with conceptual processing? The effect of attentional modulation. *Frontiers in Psychology*, 3, 1-4. doi: 10.3389/fpsyg.2012.00474

Connell, L. and Lynott, D. (2014). I See/Hear What You Mean: Semantic Activation in Visual Word Recognition Depends on Perceptual Attention. *Journal of Experimental Psychology: General*, 1-7. doi: 10.1037/a0034626.

Connell, L., Lynott, D., & Dreyer, F. (2012). A functional role for modality-specific perceptual systems in conceptual representations. *PLoS ONE* 7, e33321. doi:10.1371/journal.pone.0033321.

Conrad, C. (1972). Cognitive economy in semantic memory. *Journal of Experimental Psychology*, 92, 149-154.

Conway, M.A. (1990). *Autobiographical Memory: An Introduction*. Buckingham, UK: Buckingham University Press.

Conway, M.A. (2002). Sensory-perceptual episodic memory and its context: autobiographical memory. In A. Baddeley, J.P. Aggleton, & M.A. Conway (Eds.), *Episodic Memory: New Directions in Research*, (pp. 53-70). Oxford: Oxford University Press.

Cree, G.S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132, 163–201.

Creutzfeldt, O., Ojemann, G., and Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Experimental Brain Research*, 77(3), 451-475.

Crutch, S.J., & Warrington, E.K. (2003). The selective impairment of fruit and vegetable knowledge: A multiple processing channels account of fine-grain category specificity. *Cognitive Neuropsychology*, 20, 355–372.

Damasio, A.R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25–62.

Damasio, A.R. & Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain*, MIT Press.

Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R.D., & Damasio, A.R. (1996). A neural basis for lexical retrieval. *Nature*, 380(6574), 499-505.

Davies, R.R., Graham, K.S., Xuereb, J.H., Williams, G.B., & Hodges, J.R. (2004). The human perirhinal cortex and semantic memory. *European Journal of Neuroscience*, *20*(9), 2441-2446.

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

Desgranges, B., Matuszewski, V., Piolino, P., Chételat, G., Mézenge, F., Landeau, B., de la Sayette, V., Belliard, S., & Eustache, F. (2007). Anatomical and functional alterations in semantic dementia: a voxel-based MRI and PET study. *Neurobiology of aging*, *28*(12), 1904-1913.

Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E., Matthews, P.M., & Tyler, L.K. (2000). Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *Neuroimage*, *11*(6), 589-600.

Diehl, J., Grimmer, T., Drzezga, A., Riemenschneider, M., Förstl, H., & Kurz, A. (2004). Cerebral metabolic patterns at early stages of frontotemporal dementia and semantic dementia. A PET study. *Neurobiology of aging*, *25*(8), 1051-1056.

Donnenwerth-Nolan, S., Tanenhaus, M.K., & Seidenberg, M.S. (1981). Multiple code activation in word recognition: evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(3), 170-180.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, 110(3), 412-431.

Downing, P.E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470–2473.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598-601.

Farah, M.J. (2000). The neural bases of mental imagery. In M.S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp.965–974). Cambridge, MA: MIT Press.

Farah, M.J., and McClelland, J.L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120(4), 339-357.

Farah, M.J., Weisberg, L.L., Monheit, M. (1989). Brain activity underlying mental imagery: Event-related potentials during mental image generation. *Journal of Cognitive Neuroscience*, 1(4), 302-316.

Fernandino, L., Binder, J.R., Desai, R.H., Pendl, S.L., Humphries, C.J., Gross, W.L., Conant, L.L, & Seidenberg, M.S. (2015). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*, bhv020.

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

Fodor, J.A. (1987). *Psychosemantics*. Cambridge MA: MIT/Bradford.

Fodor, J.A. and Pylyshyn Z.W. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. In S. Pinker and J. Mehler, (Eds.), *Connections and Symbols*, Cambridge, Massachusetts: MIT Press.

Gallese, V. (2009). Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research*, 73, 486-498.

Gallese, V., Keysers, C., Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Science*, 8, 396-403.

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

Garagnani, M., Wennekers, T., and Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, 27(2), 492-513.

Gelman, R., Williams, E.M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. William (Ed.) *Handbook of child psychology: Cognition, perception, and language*, (pp. 575-630). Hoboken, NJ, US: John Wiley & Sons.

Gibbs, R.W. Jr. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. New York: Cambridge University Press.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. New York: Houghton Mifflin.

Glass, A.L., & Holyoak, K.J. (1975). Alternative conceptions of semantic memory. *Cognition*, 3, 313-339.

Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–55.

Glenberg, A.M. (2015a). Few believe the world is flat: How embodiment is changing the scientific understanding of cognition. *Canadian Journal of Experimental Psychology*, 69, 165–171.

Glenberg, A.M. (2015b). Response to Mahon: Unburdening cognition from abstract symbols. *Canadian Journal of Experimental Psychology*, 69, 181–182.

Glenberg, A.M., & Kaschak, M.P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558–565.

Glezer, L.S., Jiang, X., & Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the “visual word form area”. *Neuron*, 62, 199–204.

Gorno-Tempini, M.L., & Price, C.J. (2001). Identification of famous faces and buildings. *Brain*, *124*(10), 2087-2097.

Gorno-Tempini, M.L., Price, C.J., Josephs, O., Vandenberghe, R., Cappa, S.F., Kapur, N., Frackowiak, R.S., & Tempini, M.L. (1998). The neural systems sustaining face and proper-name processing. *Brain*, *121*(11), 2103-2118.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Gould, S.J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *Journal of Social Issues* 47, 43–65.

Gould, S.J., & Vrba, E. (2008). *Exaptation. Il bricolage dell'evoluzione*. Torino: Bollati Boringhieri.

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

Hald L.A., Marshall J-A, Janssen D.P., Garnham, A. (2011). Switching modalities in a sentence verification task: ERP evidence for embodied language processing. *Frontiers in Psychology* 2:45. doi: 10.3389/fpsyg.2011.00045

Hald L.A., Hocking I., Vernon D., Marshall J-A, and Garnham, A. (2013). Exploring modality switching effects in negated sentences: further evidence for grounded representations. *Frontiers in Psychology*. 4:93. doi: 10.3389/fpsyg.2013.00093

Hampton, J. (1993). Prototype models of concept representation. In I. Van Mechelen & R. S. Michalski (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67-95). Londra: Academic Press.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica* D 42, 335-346.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press.

Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition*. New York, NY: Norton.

Hodges, J.R., Graham, N., and Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory*, 3(3-4), 463-495.

Hoenig, K., Sim, E.J, Bochev, V., Herrnberger, B., and Kiefer, M. (2008). Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motion and motor related areas. *Journal of Cognitive Neuroscience*, 20(10), 1799-1814.

Humphreys, G.W. and Forde, E.M.E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" Neuropsychological deficits. *Behavioral and Brain Sciences*, 24(3), 453-509.

James, C. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception & Performance*, 104, 130-136.

Jeannerod, M. (1995). Mental-imagery in the motor context. *Neuropsychologia*, 33, 1419–1432.

Jeannerod, M. (2006). *Motor cognition: What actions tell the Self*. Oxford University Press, Oxford.

Kan, I.P., Barsalou, L.W., Solomon, K.O., Minor, J.K., & Thompson-Schill, S.L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *CognitiveNeuropsychology*, 20, 525–540.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.

Kaschak, M.P., Zwaan, R.A., Aveyard, M., Yaxley, R.H. (2006). Perception of auditory motion affects language processing. *Cognitive Science,* 30, 733-744.

Keil, F.C. (1992). *Concepts, kinds, and cognitive development*. MIT Press.

Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, 22, 1068–1075.

Kiefer, M., & Barsalou, L.W. (2013). Grounding the human conceptual system in perception, action, and internal states. In W. Prinz, M. Beisert, & A. Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 381-407). Cambridge: MIT Press.

Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48, 805-825.

Kiefer, M., Sim, E.J, Herrnberger, B., Grothe, J., and Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience*, 28(47), 12224-12230.

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

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

Lakoff, G, & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Lambon Ralph, M.A., Sage, K., Jones, R.W., & Mayberry, E.J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, 107(6), 2717-2722.

Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. *Psychonomic Bulletin & Review*, 23(4), 991-1001.

Lynott, D., Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41 (2), 558-564.

Macario, J.F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, *6*(1), 17-46.

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review*, 23(4), 1090-1095.

Magnuson, J.S. & Nusbaum, H.C. (1993) Talker differences and perceptual normalization. *Journal of the Acoustical Society of America*, 93, 2371.

Mahon, B.Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, 30, 420–429.

Mahon, B.Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-*Paris, 102, 59–70.

Mahon B.Z., & Hickok, G. (2016). Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychonomic Bullettin & Review*, 1-18.

Malt, B. C. (2010). Why we should do without concepts. *Mind & Language*, 25 (5), 622-633.

Marques, J.F. (2006). Specialization and semantic organization: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition*, 34 (1), 60-67.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Marschark, M., & Cornoldi, C. (1991). Imagery and verbal memory. In C. Cornoldi & M. A. McDaniel (Eds.), *Imagery and cognition*, (pp. 41-56). New York: Springer.

Martin, A. (2001). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.) *Handbook of functional neuroimaging of cognition* (pp. 153–186). Cambridge, MA: MIT Press).

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

Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic bulletin & review*, 1-12.

Martin, A., & Chao, L. (2001). Semantic memory and the brain: Structure and process. *Current Opinion in Neurobiology*, 11, 194–201.

Martin, A., Ungerleider, L.G., & Haxby, J.V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), *Higher cognitive functions: The new cognitive neurosciences* (pp. 1023–1036). Cambridge, MA: MIT Press.

McCaffrey, J. (2015). Reconceiving conceptual vehicles: Lessons from semantic dementia. *Philosophical Psychology*, 28 (3), 337-354.

McClelland, J.L., and Rogers, T.T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, 4(4), 310-322.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception. *Psychological Review*, 88(5), 375-407.

McCloskey, M., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for model of semantic memory. *Cognitive Psychology*, 11, 1-37.

McNamara, T.P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1173-1190.

McRae, K., & Cree, G.S. (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category-specificity in mind and brain* (pp.211–249). Hove, UK: Psychology Press.

Medin, D.L., and Shoben, E. (1988). Context and Structure in Conceptual Combination. *Cognitive Psychology*, 20, 158–190.

Morsella, E. & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28 555-563.

Mummery, C.J., Patterson, K., Price, C.J., Ashburner, J., Frackowiak, R.S.J., & Hodges, J.R. (2000). A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Annals of neurology*, *47*(1), 36-45.

Muneaux, M. & Ziegler, J.C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, 19, 641-660.

Murphy, G.L. (2002). *The big book of concepts*. Cambridge: The MIT Press.

Murphy, G.L., and Medin D.L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review*, 92, 289–316.

Myachykov, A., Scheepers, C., and Shtyrov Y. Y. (2013). Interfaces between language and cognition. *Frontiers in Psychology*, 4, 1-2. doi: 10.3389/fpsyg.2013.00258

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K. Ito, K., Fukuda, H., Schormann, T., & Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. *Brain*, *123*(9), 1903-1912.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., & Kojima, S. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia*, *39*(10), 1047-1054.

Navarrete, E. & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, 53, 359-377.

Nestor, P.J., Fryer, T.D., & Hodges, J.R. (2006). Declarative memory impairments in Alzheimer's disease and semantic dementia. *Neuroimage*, *30*(3), 1010-1020.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4, 135-183.

Newell, A., & Simon, H.A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the Association for Computing Machinery*, 19, 113–126.

Paivio, A. (1966). Latency of verbal associations and imagery to noun stimuli as a function of abstractness and generality. *Canadian Journal of Psychology*, 20(4), 378-387.

Palmer, S.E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Patterson, K., Nestor, P.J., and Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976-987.

Pecher, D. (2013). The perceptual representation of mental categories. In D. Reisberg (Ed.). *The Oxford Handbook of Cognitive Psychology*. New York: Oxford University Press.

Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2003). Verifying the properties of object concepts across different modalities produces switching costs. Psychological Science, 14, 119-124.

Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. Psychonomic Bulletin and Review, 11, 164-167.

Plaut, D.C. (2002). Graded modality specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, 19(7), 603-639.

Pobric, G., Jefferies, E., and Lambon Ralph, M.A. (2010). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia*, 48(5), 1336-1342.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129-154.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22(2), 253-336.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. Nature Reviews *Neuroscience*, 6, 576–582.

Pulvermüller, F. (2012). Meaning and the brain: The neurosomatics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics*, 25, 423–459.

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17, 458–470.

Pulvermüller, F., & Garagnani, M. (2014). From sensorimotor learning to memory cells in prefrontal and temporal association cortex: A neurocomputational study of disembodiment. *Cortex*, 57, 1–21.

Pulvermüller, F., Härle, M. & Hummerl, F. (2000). Neurophysiological distinction of verb categories. *Neuroreport*, 11, 2789-2793.

Pulvermüller, F., Lutzenberger, W., and 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.

Pylyshyn, Z.W. (1980). Computation and cognition: Issues in the foundation of cognitive science. *Behavioral and Brain Sciences*, 3, 111-169.

Pylyshyn, Z.W. (1984). *Computation and cognition*. Cambridge MA: MIT/Bradford.

Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102-1207.

Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neuroscience*, 21, 188-94.

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

Roediger, H.L., & McDermott, K.B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.) *Handbook of Neuropsychology*, (vol. 8(2), pp.63-121), Elsevier.

Rogers, T.T., Lambon Ralph, M.A., Garrard, P., Bozeat, S., McClelland, J.L., Hodges, J.R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111(1), 205-235.

Rogers, T.T., & McClelland, J.L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Rogers, T.T., & Patterson, K. (2007). Object categorization: reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451.

Rosch, E. (1978). Principles of Categorization. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Rosch, E., & Mervis, C.B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573–605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, 8, 382–439.

Rubin D.C. (2006). The basic-systems model of episodic memory. Perspective Psychological Science, 1, 277-311.

Rumelhart, D.E., and McClelland, J.L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Cambridge, MA: MIT Press.

Samuel, A.G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474-494.

Samuel, A.G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language*, 26, 36-56.

Schacter, D.L. (1995). Implicit memory: A new frontier for neuroscience. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*, MIT Press.

Searle, J. (1980). Minds, brains and programs, Behavioral and Brain Science, 417-457.

Seidenberg, M.S. & Tanenhaus, M.K. (1979). Orthographic Effects on Rhyme Monitoring. *Journal of Experimental Psychology: Human Learning and Memory,* 5, 546-554.

Simmons, W.K., and Barsalou, L.W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology,* 20(3), 451-486.

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, 2802–2810.

Smith, E.E., and Medin, D.L. (1981). *Categories and concepts*. Harvard University Press.

Smith, E.E., Shoben E.J., and Rips L.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81(3), 214-241.

Snowden, J.S., Thompson, J.C., & Neary, D. (2004). Knowledge of famous faces and names in semantic dementia. *Brain, 127*(4), 860-872.

Solomon, K.O., Barsalou, L.W. (2001). Representing properties locally. *Cognitive Psychology*, 43, 129-169.

Solomon, K.O., Barsalou, L.W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, 32 (2), 244-259.

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, 63, 330-336.

Steels, L., & Brooks, R. (1995). *The Artificial Life Route to Artificial Intelligence: Building Embodied SituatedAgents*. Mahwah, NJ: Erlbaum.

Thelen, E., & Smith L.B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Trumpp, N.M., Traub, F., & Kiefer, M. (2013). Masked priming of conceptual features reveals differential brain activation during unconscious access to conceptual action and sound information. *PLoS ONE* 8(5): e65910. doi:10.1371/journal.pone.0065910.

Trumpp, N.M., Traub, F., Pulvermüller, F., Kiefer, M. (2014). Unconscious automatic brain activation of acoustic and action-related conceptual features during masked repetition priming. *Journal of Cognitive Neuroscience*, 26(2), 352-364.

Tsukiura, T., Mochizuki-Kawai, H., & Fujii, T. (2006). Dissociable roles of the bilateral anterior temporal lobe in face-name associations: An event-related fMRI study. *Neuroimage, 30*(2), 617-626.

Turing, A.M. (1964). Computing machinery and intelligence. In A. Anderson (Ed.), *Minds and Machines*, Prentice Hall, Engelwood Cliffs, NJ.

Tyler, L.K., and Moss, H.E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, 5(6), 244-252.

Van Dantzig, S., Pecher, D., Zeelenberg, R., Barsalou, L.W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, 32, 579-590.

Vermeulen, N., Niedenthal, P.M., &Luminet, O. (2007). Switching between sensory and affective systems incurs processing costs. *Cognitive Science*, 31, 183-192.

Vermeulen, N., Corneille, O., Niedenthal, P.M., (2008). Sensory load incurs conceptual processing costs. *Cognition*, 109, 287-294.

Vigliocco, G., Vinson, D.P., Lewis, W., and Garrett, M.F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48(4), 422-488.

Vinson, D.P., Vigliocco, G., Cappa, S., & Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language*, 86, 347-365.

Warren, R.M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392-393.

Warrington, E.K., & McCarthy, R. A. (1983). Category specific access dysphasia. *Brain*, 106, 859-878.

Warrington, E.K., & McCarthy, R.A. (1987). Categories of knowledge: Further fractionations and an attempted integration. *Brain*, 110, 1273-1296.

Warrington, E.K. & Shallice, T. (1984). Category-specific semantic impairments. *Brain*, 107, 829-854.

Wennekers, T., Garagnani, M., and Pulvermüller, F. (2006). Language models based on Hebbian cell assemblies. *Journal of Physiology* (Paris), 100(1-3), 16-30.

Wittgenstein, L. (1953). *Philosophical investigations*. Macmillan.

Yee, E., & Thompson-Schill, S.L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, 23. doi: 10.3758/s13423-015- 0948-7.

Zeki, S. (1993). *A vision of the brain*. Cambridge, MA: Blackwell Scientific Publications.

Zwaan, R.A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp.35–62). NewYork, NY: Academic Press.

Zwaan, R.A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, 23, 1028-1034.

Overview of Publication Status of Chapters in Thesis

Chapter number and title	Original text (not published before)	Submitted: no feedback received	Submitted: revision requested or revision submitted	Accepted/published (specify journal or book)
1. Chapter one – Amodal symbolic accounts of knowledge	x			
2. Chapter Two - Grounded accounts of knowledge	x			
3. Chapter Three - Hybrid accounts of knowledge	x			
4. Chapter Four - Study 1 “The Modality-Switch Effect: Visually and Aurally Presented Prime Sentences Activate our Senses”				x Frontiers in Psychology (2015)
5. Chapter Five - Study 2 “The Multilevel Modality-Switch Effect: What Happens when We See the Bees Buzzing and Hear the Diamonds Glistening”				x Psychonomic Bulletin & Review (2016)