ENVISIONING LANGUAGE - AN EXPLORATION OF PERCEPTUAL PROCESSES IN LANGUAGE COMPREHENSION

ISBN: 978-90-76203-95-9

The cover artwork was made by Sanneke Duijf

Printed and bound by Ipskamp Drukkers b.v.

Envisioning language - An exploration of perceptual processes in language comprehension

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van de decanen
in het openbaar te verdedigen op maandag 25 juni 2018
om 16.30 uur precies

door
Markus Ostarek
geboren op 15 januari 1990
te Rosenheim (Duitsland)

Promotor:

Prof. dr. Antje Meyer

Copromotor:

Dr. Falk Huettig

Manuscriptcommissie:

Prof. dr. James McQueen
Prof. dr. Floris de Lange
Prof. dr. Pienie Zwitserlood (Universiteit Münster, Duitsland)

This research was supported by the Max Planck Society for the Advancement of Science, Munich, Germany.

TABLE OF CONTENTS

Chapter 1

GENERAL INTRODUCTION

How do we understand language? What is the nature of the conceptual system that constitutes the foundation for this remarkable capacity? Philosophers - and later linguists, psycholinguists, psychologists, and cognitive neuroscientists - have struggled with these questions for millennia and have yet to reach consensus on several elementary fronts. The dominant view in the second half of the 20th century has been that conceptual processing relies on an amodal symbol system that is strictly divorced from the systems for perception, action, and introspection (Fodor, 1975; Newell, 1980; Pylyshyn, 1980). However, inspired by theoretical considerations (Barsalou, 1999; Damasio, 1989; Harnad, 1990) and empirical developments (Barsalou, 2008; Pulvermüller, 2005), a different perspective has emerged in the past two decades. This view, often referred to as *embodied cognition*, ascribes a central role to the modal systems in high-level cognitive functions including conceptual processing and language comprehension. A central proposal in this framework is that conceptual content is retrieved from those systems that are active during the perception of the corresponding exemplars. As such, a key prediction is that the type of conceptual content activated by a word should systematically map onto the brain regions sensitive to the perceptual features of the corresponding referents. The process of reactivation is often referred to as *simulation*, reflecting the hypothesized similarity to bottom-up perception even though no bottom-up input is present.

As this research program differs so strongly in spirit from the symbolic approach to cognition, to many it may seem like a fresh and new approach. But this way of thinking about the mind rather reflects the renaissance of a view that had been quite dominant for the most part of documented human philosophy. In *de Anima*, the perhaps first comprehensive theory of cognition, Aristotle proposed that mental images (*phantasmata*) have central functions in memory and thought and claimed that "It is impossible to think without an image (*phantasma*)" (Hicks, 2015). Interestingly, he also anticipated the main thesis of embodied accounts of language comprehension by stipulating that the comprehension of spoken words relies on mental images. Similar ideas resurfaced in the work of British empiricists Hume (1739/2003), Berkeley (1710/1999), and Locke (1700), and in that of Kant (1787/1998), among many other philosophers.

One reason why these ideas have been so pervasive might be that they make intuitive sense. Many creative thinkers, such as Albert Einstein, are quoted as having relied heavily on imagery to solve complex problems, and most of us have experienced the colorful mental worlds to which good stories can transport us. Modern cognitive (neuro) science has developed tools that allow an increasingly sharp glimpse into the processes that occur when we engage in conceptual processing. Consequently, in the past two decades a substantial amount of behavioral and neuroimaging work has considerably advanced our understanding of the processes involved in semantic processing (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017).

This thesis combines binocular rivalry, visual interference, eye-tracking, and functional magnetic resonance imaging (fMRI) to further examine the cognitive and neural mechanisms underlying language comprehension. It focuses on a number of questions that remain unanswered and whose answers would be critical for theoretical progress. 1) It still remains unclear to what extent language comprehension involves the retrieval of information from modality-specific systems. This is so because many behavioral paradigms that are typically cited as evidence for that view are subject to alternative explanations and the neuroimaging evidence is inconclusive. 2) To the extent that semantic processing does involve sensory processes, a characterisation of the nature of the activated processes is missing. 3) Even a bullet-proof demonstration that modality-specific processes are *activated* during semantic processing would not be sufficient to conclude that language comprehension relies on simulation. Instead, it would be crucial to determine the extent to which modality-specific processes contribute *functionally* to language comprehension. 4) Recent evidence suggests that language comprehension is highly context-dependent and that a simple one-to-one mapping between words and concepts is therefore not realistic. The mechanisms allowing the context-sensitive retrieval of information remain largely unclear.

Below, I elaborate on each of these points and briefly describe how the experiments presented in this thesis attempted to tackle them. For the most part, in this thesis I examined the role of visual processes for the comprehension of concrete words and sentences. Therefore, this Introduction similarly focuses on the visual domain even though a substantial amount of research has been devoted to the role of other modalities, most notably the motor domain (see for instance Pulvermüller, 2005).

Behavioral evidence

Congruency paradigms

Experimental efforts testing the hypothesis that modality-specific processes are activated for language comprehension come in two flavors: behavioral studies examining how language and perception interact, and neuroimaging studies. Numerous behavioral studies have provided evidence for interactions between semantic and visual processing. Typically, congruency between semantic aspects of words and visual aspects of images leads to facilitated processing in cross-modal priming paradigms, suggesting that language and vision are intertwined (Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Kaschak et al., 2005; Meteyard, Bahrami, & Vigliocco, 2007; Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009; Stanfield & Zwaan, 2001; Zwaan, Madden, Yaxley, & Aveyard, 2004; Zwaan & Pecher, 2012; Zwaan, Stanfield, & Yaxley, 2002).

For instance, in the sentence-picture verification paradigm (Zwaan et al., 2002), participants read sentences implying a certain shape of an object (e.g., *The ranger saw an eagle flying in the sky*, implying outstretched wings). After the sentence, they are presented with a picture either of an eagle or of another object and they have to indicate as quickly and accurately as possible whether the picture shows an object that was mentioned in the sentence. Crucially, half of the time, the target picture shows the object matching the shape implied in the sentence (an eagle with outstretched wings). The well-replicated finding (Engelen et al., 2011; Pecher et al., 2009; Rommers, Meyer, & Huettig, 2013; Zwaan & Pecher, 2012) is that reaction times to shape-matching pictures are faster than to shape-mismatching pictures, even though both trigger clear "yes" responses and participants are usually unaware of the shape manipulation. This strongly suggests that participants retrieve implicit shape information during sentence comprehension and that the pre-activation of shape information facilitates processing of shape-matching target pictures. Further studies using sentence-picture verification provided similar evidence for the activation of orientation (Pecher et al., 2009; Stanfield & Zwaan, 2001), size (de Koning, Wassenburg, Bos, & Van der Schoot, 2017), color (de Koning, Wassenburg, Bos, & van der Schoot, 2017; Mannaert, Dijkstra, & Zwaan, 2017), and distance information (Vukovic & Williams, 2014). These findings are typically interpreted as evidence for perceptual simulation during language comprehension.

However, two recent studies suggest that the orientation effects are unreliable (de Koning, Wassenburg, Bos, & van der Schoot, 2017; Rommers et al., 2013). Moreover, evidence from picture naming pointing to reduced effect sizes for the shape match effect (Rommers et al., 2013) raises the question to what extent previous results generalize to other paradigms and to everyday communication. But an even more fundamental issue is that even if we accept that the robust shape-match effect in the sentence-picture verification paradigm is a good approximation of routine language processing, it does not constitute strong evidence for the recruitment of visual processes as part of comprehension. This is because the paradigm is informative about the conceptual *content* that is activated, not about the underlying *processes* that provide it. Given that the target pictures are not only processed at the visual level, it is inherently difficult to rule out that the observed reaction time effects arise from congruency at a higher semantic level that may be shared across modalities. An alternative account that is entirely compatible with the data thus is that visual processing is not affected by language. Instead, both domains interact at the conceptual level. Therefore, previous congruency paradigms only allow limited conclusions about the kind of processes that are activated during language processing. This is also the case in related paradigms that demonstrated effects of language on visual search (Lupyan & Spivey, 2010b) and categorization (Lupyan & Thompson-Schill, 2012), and cannot easily be avoided as even simple visual tasks such as near-threshold categorization of motion patterns as up or down (Meteyard et al., 2007) and the detection of backward-masked stimuli (Lupyan & Spivey, 2010a) may be subject to semantic analysis (Kouider & Dehaene, 2007). Thus, congruency effects in these paradigms may rely on high-level integration processes (Francken, Kok, Hagoort, & De Lange, 2014; Francken, Meijs, Hagoort, Van Gaal, & De Lange, 2015).

This means that existing congruency paradigms mostly advance our understanding of the conceptual content that is accessed during language comprehension, but have left question marks regarding the nature of the processes that provide this content. Chapter 2 attempts to fill this gap by testing the hypothesis that spoken words activate concept-specific shape information in early visual processing stages. This was achieved by employing a binocular rivalry technique called continuous flash suppression (CFS; Tsuchiya & Koch, 2005) that makes it possible to study effects of language on basic visual perception (Lupyan & Ward, 2013). CFS can be used to study the most basic form of visual detection that occurs without access to semantic information (Stein, Thoma, & Sterzer, 2015). As such, it can solve the interpretational challenge of previous cross-modal priming experiments and can thus be considered a strong test of whether language activates low-level visual processes: If words can enhance the detection of corresponding referents in CFS, this would indicate that words activate category-specific visual processes involved in the earliest stages of conscious vision.

Functional role of visual processes for language comprehension

The studies that implicated visual processes with semantic processing typically relied on correlational methods, as exemplified in the congruency paradigms reviewed above. As such, it remains unclear to what extent they play a functional role for language comprehension. One fundamental assumption of studies showing activation of visual processes during language comprehension is that this activation reflects semantic processing. However, in principle it would be possible that activation of sensory processing systems during language comprehension is epiphenomenal. Some models assume that there are bidirectional links between semantic and the perceptual representations, even though both are separate systems. Thus, sensory-motor processes could be activated as a result of semantic processing, rather than contributing to it (Mahon, 2015; Mahon & Caramazza, 2008).

This issue can be addressed to some extent with methods capable of revealing how quickly sensory processes are activated (Hauk & Tschentscher, 2013; Shtyrov et al., 2014): If they are activated in the first hundreds of milliseconds after word onset (the time window typically associated with semantic processing), it is likely that this reflects access to word meanings. In order to directly demonstrate a functional role however, one would need to show a causal link between sensory processes and comprehension. This has been achieved in the motor domain, where transcranial magnetic stimulation of the motor and premotor cortex has been shown to impact the speed of action language comprehension (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005; Vukovic, Feurra, Shpektor, Myachykov, & Shtyrov, 2017; Willems, Labruna, D'Esposito, Ivry, & Casasanto, 2011). In Chapter 3 I investigated in what task situations low-level visual processes contribute causally to semantic processing of concrete nouns by employing a visual interference technique while participants performed different linguistic tasks on the same set of words.

As described above, previous research has shown that implicit shape information is retrieved during sentence comprehension. To the extent that the retrieval of shape information involves visual processes, no study has tested whether they play a causal role. This would be crucial to determine whether visual processes constitute a functional part of the semantic processing machinery. In chapter 4, I used visual interference techniques aimed at different processing stages to characterize the cognitive processes that drive the shape-match effect in the sentence-picture verification paradigm.

Item-independent concepts

An often neglected type of conceptual dimension that forms a crucial component of the human conceptual system is that of *item-independent* concepts. In contrast to the widely studied class of concepts that are straightforwardly described in terms of their sensory-motor features, these refer to concepts that do not denote tangible referents (Borghi et al., 2017). In particular, I focused on the vertical spatial dimension. This dimension is item-independent because an item's typical location in space does not predict what kind of item it is (big vs. small, animate vs. inanimate, moving vs. stationary, etc.). Therefore, conceptual spatial representations have to generalize across items to derive abstract spatial information. Behavioral experiments have suggested that nouns denoting referents that are typically perceived up or down in space (e.g., *bird*, *mouse*) induce spatial shifts of attention (Gozli, Chasteen, & Pratt, 2013). Eye-tracking studies recently showed that they speed up launch times of subsequent saccades (Dudschig, Souman, Lachmair, de la Vega, & Kaup, 2013; Dunn, 2016). In Chapter 5, I investigated the dynamics of interactions between semantic and spatial information *during* the comprehension of implicitly spatial words. More specifically, I used growth curve analysis to test the hypothesis that processing of implicit up/down words influences the trajectories of concurrently performed saccades.

Neuroimaging evidence

Besides behavioral efforts, an obvious way of probing the mechanics of conceptual processing is to investigate their neural correlates. Embodied accounts predict that the type of semantic content is systematically related to the systems that are recruited as meaning is computed in the brain. In line with this, neuroimaging studies observed that action concepts activate aspects of the fronto-parietal motor system and motion-sensitive temporal areas (Hauk, Johnsrude, & Pulvermüller, 2004; Martin, Wiggs, Ungerleider, & Haxby, 1996; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014; Watson, Cardillo, Ianni, & Chatterjee, 2013), whereas object concepts relying strongly on visual information activate inferior temporal and visual areas (Chao, Haxby, & Martin, 1999; Desai, Binder, Conant, & Seidenberg, 2009; Hwang, Palmer, Basho, Zadra, & Müller, 2009; Lewis & Poeppel, 2014; Martin, 2007; Martin et al., 1996).

Some previous fMRI studies indicate that listening to and producing language about concrete objects activates occipital regions involved in visual perception (Desai, Binder, Conant, & Seidenberg, 2009; Hwang, Palmer, Basho, Zadra, & Müller, 2009) and one MEG study (Lewis & Poeppel, 2014) observed that spoken words with high imageability activate extrastriate cortex ca. 200 ms after word onset, simultaneous with (or even slightly before) lexical effects in middle temporal areas. This suggests that the comprehension of concrete words involves the activation of visual processes. However, these neuroimaging approaches based on univariate activation analysis cannot reveal what information is retrieved and thus the nature of the activated representations remains unclear. Recent developments in the analysis of neuroimaging data exploiting machine learning techniques have made it possible to study how individual words are reflected in fine-grained activity patterns (Haxby, Connolly, & Guntupalli, 2014; Norman, Polyn, Detre, & Haxby, 2006) and allow researchers to establish direct links between hypothesized semantic content and activity patterns (Kriegeskorte & Kievit, 2013).

The neural correlates of spatial word meaning

In Chapter 6 (following up on Chapter 5), I used multi-voxel pattern analysis to examine the hypothesis that the abstract up-down dimension implicit in these words is reflected in direction-specific activation patterns in the cortical network for eye movements. Previous behavioral and eye-tracking evidence demonstrated that spatial aspects of word meanings can interact with spatial attention (Dudschig et al., 2013; Gozli et al., 2013). An open question was whether semantic processing of implicitly spatial words in the absence of a visuo-spatial task activates direction-specific processes in the cortical network for spatial shifts of attention (Corbetta et al., 1998). To directly test this hypothesis, multi-voxel patterns analysis was used to identify brain areas whose patterns distinguish between words with up vs. down associations as well as brain areas whose patterns are shared between up vs. down words and up vs. down saccades.

Task-dependence/flexibility of conceptual processing

Whether or not an aspect of our conceptual knowledge is functionally relevant is likely to depend on task and other contextual demands. Traditionally, cognitive theories have envisioned concepts to be represented in invariant symbolic codes that are immune to momentary contextual factors in order to guarantee conceptual stability (Fodor, 1975; Tyler & Moss, 2001). Based on the highly flexible way in which language is used, it has recently been suggested that conceptual processing is inherently flexible, such that the retrieval of semantic information in response to the same words can differ drastically due to contextual factors (Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008; Lebois, Wilson-

Mendenhall, & Barsalou, 2015; Yee & Thompson-Schill, 2016). Previous studies have demonstrated that access to conceptual content is highly flexible (Hoenig et al., 2008; Lebois et al., 2015; Nieuwland & Van Berkum, 2006; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Van Berkum, Brown, & Hagoort, 1999; Yee, Ahmed, & Thompson-Schill, 2012; Yee & Thompson-Schill, 2016) such that task contexts have a strong impact on the type of information that is accessed. Less is known about the underlying neuro-cognitive mechanisms. One fMRI study showed that brain activity patterns evoked by a word at different points in time can vary substantially (Musz & Thompson-Schill, 2015). Univariate fMRI studies have reported increased activation of modality-specific processing areas when the corresponding features were task-relevant (Hoenig et al., 2008; van Dam, van Dijk, Bekkering, & Rueschemeyer, 2012).

Chapter 7 relied on multi-voxel pattern analysis to investigate how task demands influence the processing of the same set of words in different contexts. Representational similarity analysis allowed us to specifically test how size and animacy information is reflected in brain activity patterns when these dimensions were task relevant vs. irrelevant. Moreover, I tested how consistent fine-grained activity patterns in response to individual words were within-task and between-tasks to map out which brain areas flexibly adapt their response profiles to task demands and which are invariant to them.

References

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Berkeley, G. (1999). *A treatise concerning the principles of human knowledge. 1710*. RS Bear.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., … Shulman, G. L. (1998). A Common Network of Functional Areas for Attention and Eye Movements. *Neuron*, *21*(4), 761–773. https://doi.org/10.1016/S0896-6273(00)80593-0

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1), 25–62.

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017). Mental simulation of four visual object properties: similarities and differences as assessed by the sentence–picture verification task. *Journal of Cognitive Psychology*, 1–13.

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, *54*(7), 493–503.

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M. S. (2009). Activation of sensory–motor areas in sentence comprehension. *Cerebral Cortex*, *20*(2), 468–478.

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading “sun” and looking up: The influence of language on saccadic eye movements in the vertical dimension. *PloS One*, *8*(2), e56872.

Dunn, B. M. (2016). *Which way is up? Grounded mental representations of space*. University of Glasgow.

Engelen, J. A., Bouwmeester, S., de Bruin, A. B., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*(4), 659–675.

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

Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit spatial cues for visual attention. *Journal of Experimental Psychology: General*, *142*(4), 1028.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346.

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

Hauk, O., & Tschentscher, N. (2013). The body of evidence: what can neuroscience tell us about embodied semantics? *Frontiers in Psychology*, *4*.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456.

Hicks, R. D. (2015). *Aristotle De Anima*. Cambridge University Press.

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

Hume, D. (2003). *A treatise of human nature*. Courier Corporation.

Hwang, K., Palmer, E. D., Basho, S., Zadra, J. R., & Müller, R.-A. (2009). Category-specific activations during word generation reflect experiential sensorimotor modalities. *Neuroimage*, *48*(4), 717–725.

Kant, I. (1998). Critique of Pure Reason, trans. P. Guyer and A. Wood. *The Cambridge Edition of the Works of Immanuel Kant*.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*(3), B79–B89.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 857–875.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Lambon Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*(8), 1764–1801.

Lewis, G., & Poeppel, D. (2014). The role of visual representations during the lexical access of spoken words. *Brain and Language*, *134*, 1–10.

Locke, J. (1700). *An essay concerning human understanding*. Awnsham and John Churchil, at the Black-Swan in Pater-Noster-Row, and Samuel Manship, at the Ship in Cornhill, near the Royal-Exchange.

Lupyan, G., & Spivey, M. J. (2010a). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS One*, *5*(7), e11452.

Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236–2253.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201.

Mannaert, L. N. H., Dijkstra, K., & Zwaan, R. A. (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition*, 1–9.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*(6566), 649–652.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013.

Musz, E., & Thompson-Schill, S. L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia*, *76*, 41–51.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*(2), 135–183.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1108–1114.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, *3*(1), 111–132.

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, *24*(11), 2218–2225.

Shtyrov, Y., Butorina, A., Nikolaeva, A., & Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences*, *111*(18), E1918–E1923.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Stein, T., Thoma, V., & Sterzer, P. (2015). Priming of object detection under continuous flash suppression depends on attention but not on part-whole configuration. *Journal of Vision*, *15*(3), 15–15.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 1632–1634.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096.

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

Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147–182.

Vukovic, N., & Williams, J. N. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychologica*, *145*, 98–103.

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action concepts in the brain: an activation likelihood estimation meta-analysis. *Journal of Cognitive Neuroscience*, *25*(8), 1191–1205.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (can) prime furiously. *Psychological Science*, *23*(4), 364–369.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, *28*(4), 611–619.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PloS One*, *7*(12), e51382.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Chapter 2

SPOKEN WORDS CAN MAKE THE INVISIBLE VISIBLE – TESTING THE INVOLVEMENT OF LOW-LEVEL VISUAL PROCESSES IN SPOKEN WORD PROCESSING

Based on: Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible - Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 499.

Abstract

The notion that processing spoken (object) words involves activation of category-specific representations in visual cortex is a key prediction of modality-specific theories of representation that contrasts with theories assuming dedicated conceptual representational systems abstracted away from sensorimotor systems. In the present study, we investigated whether participants can detect otherwise invisible pictures of objects when they are presented with the corresponding spoken word shortly before the picture appears. Our results showed facilitated detection for congruent ("bottle" -> picture of a bottle) vs. incongruent ("bottle" -> picture of a banana) trials. A second experiment investigated the time-course of the effect by manipulating the timing of picture presentation relative to word onset and revealed that it arises as soon as 200-400ms after word onset and decays at 600ms after word onset. Together, these data strongly suggest that spoken words can rapidly activate low-level category-specific visual representations that affect the mere detection of a stimulus, i.e. what we see. More generally our findings fit best with the notion that spoken words activate modality-specific visual representations that are low-level enough to provide information related to a given token and at the same time abstract enough to be relevant not only for previously seen tokens but also for generalizing to novel exemplars one has never seen before.

Introduction

In the last decade, much of the research on the mechanisms underlying language comprehension has focused on the role of modality-specific representations that have traditionally been implicated with vision, hearing, planning and executing movements, etc. A substantial amount of evidence is consistent with the idea that sensorimotor systems are a component in the process chain related to conceptual access in language comprehension (Barsalou, 2008). Numerous studies have found modality-specific activity (Binder & Desai, 2011a; Martin, Wiggs, Ungerleider, & Haxby, 1996; (Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008a; Kiefer et al., 2012), such that conceptual processing involves activation of the visual cortex (Lewis & Poeppel, 2014), auditory regions in superior posterior and middle temporal gyri (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008), gustatory regions (Simmons, Martin, & Barsalou, 2005), primary olfactory cortex (González et al., 2006), or motor regions (Pulvermüller, Shtyrov, & Ilmoniemi, 2005). Based on these findings it has been proposed that perceptual simulation is a prime mechanism for language comprehension (e.g., Barsalou, 2008; Pulvermüller, 2005), or at least an important part of it. Several recent review papers however have pointed out (rightly in our view) that theoretical progress has been limited and some of the fundamental questions remain unclear (Binder & Desai, 2011; Hauk & Tschentscher, 2013; Kiefer & Pulvermüller, 2012; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Willems & Casasanto, 2011).

One such question is whether processing of (object) words involves activating category-specific low-level visual representations in visual cortex. This is a key prediction of modality-specific theories of representation, which contrasts with theories assuming dedicated conceptual representational systems abstracted away from sensorimotor systems. Some experimental evidence is consistent with such a prediction. A fMRI study found increased activation in Brodmann area (BA) 19 (a visual association area) for visual compared to abstract sentences (Desai, Binder, Conant, & Seidenberg, 2009). Similarly, overtly and covertly producing words related to visually salient concepts activated BA 18 (Hwang, Palmer, Basho, Zadra, & Müller, 2009). Moreover, a recent MEG study benefiting from excellent temporal resolution found a correlation between the imageability of object words and activation in BA 19 as rapid as ca. 200ms after word onset (Lewis &

Poeppel, 2014). While intriguing, these findings do not tell us much about the nature of the representations that were accessed. To be useful for conceptual processing, visual representations need to be specific enough to provide information related to a given concept and at the same time abstract enough to be relevant not only for previously seen tokens (a signature of episodic memory) but also for recognizing novel exemplars and grasping the concept's meaning more generally. Because of these requirements, it is not obvious that low-level sensory areas (as opposed to high-level areas that are based on more holistic representations) are useful for conceptual processing and can provide this sort of information.

In the present study, we directly tested whether low-level visual representations are involved in spoken word processing. Using continuous flash suppression we show that spoken words activate behaviorally relevant low-level visual representations and pin down the time-course of this effect to the first hundreds of milliseconds after word onset.

Thus far, research into object knowledge representation has been largely restricted to high-level visual cortex. While it is uncontroversial that higher level visual areas are strongly implicated in conceptual processing related to objects (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009; Martin, 2007), it remains open what types of representations are involved. On the one hand several recent studies found that activation patterns in these regions are dominated by strikingly low-level visual features (Andrews, Watson, Rice, & Hartley, 2015; Nasr, Echavarria, & Tootell, 2014; Watson, Hymers, Hartley, & Andrews, 2016). A recent study with macaque monkeys suggests that low-level information may be represented most explicitly in anterior regions such as inferior temporal lobe (IT) rather than visual cortex V4 and V1 (Hong, Yamins, Majaj, & DiCarlo, 2016), such that category-orthogonal features of pictures, such as position, size, pose, etc. can be decoded much more reliably from IT than early visual cortex. On the other hand there is evidence from studies with blind individuals suggesting that the properties of high level visual areas are better described as multimodal (e.g., Mahon et al., 2009; for review and discussion, see: Bi, Wang, & Caramazza, 2016), pointing to the possibility that visual information is transduced into conceptual information relatively early in the ventral stream. Though fascinating, the currently available data from research on high-level visual brain regions therefore cannot elucidate to what extent conceptual processing relies on modality-specific visual representations.

A similar issue arises in the context of the available behavioral data. Many studies point to a tight link between conceptual processing of object concepts and visual perception. Words can enhance attention to the corresponding categories in a visual search task (Lupyan & Spivey, 2010b), have a stronger effect on visual categorization than equally informative environmental sounds (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012), can facilitate the perception of shape-matching vs. mismatching objects (Zwaan, Stanfield, & Yaxley, 2002; but see Rommers, Meyer, & Huettig, 2013), etc. However, what all of these studies have in common is that the picture targets are not only processed on the visual level, but also the conceptual level. Therefore it cannot be ruled out that the reported effects arose from congruency on a high-level system instead of the visual system. This problem cannot easily be circumvented, as even simple visual tasks, such as categorizing motion patterns as moving up or down (Meteyard, Bahrami, & Vigliocco, 2007), or detecting backward-masked stimuli (Lupyan & Spivey, 2010a) are still subject to semantic processing (Kouider & Dehaene, 2007) and may therefore involve high-level integration processes (Francken, Meijs, Ridderinkhof, et al., 2015; Francken, Meijs, Hagoort, van Gaal, & de Lange, 2015).

In order to find out whether object words activate low-level visual representations, it is necessary to isolate processes that are unequivocally part of *basic* visual processing and that do not tap into high-level systems. This can be achieved by means of a detection paradigm implemented in the binocular rivalry technique called "continuous flash suppression" (CFS). CFS is typically used in research on subconscious visual processing because of its capacity to render pictures invisible for relatively long periods of time (Tsuchiya & Koch, 2005). CFS disrupts processing of visual stimuli (Gayet, van der Stigchel, & Paffen, 2014; Kang, Blake, & Woodman, 2011; Stein, Hebart, & Sterzer, 2011), such that they elicit only weak activity in visual areas. At the same time it does not hinder other concurrent cognitive tasks, such that, for instance, auditory stimuli can be fully processed. Importantly, CSF can be used to study basic visual detection, which is generally accepted to be a basic visual process. A recent study (Stein, Thoma, & Sterzer, 2015) has scrutinized the factors underlying the detection of pictures in CFS and found that holistic representations played no integral role for this task, further supporting the low-level nature of the process. This makes CFS an ideal tool to study the potential involvement of low-level visual processes in spoken word processing, namely by testing the effect of words on the mere detection of the presence of a stimulus.

Previous research indicates that it is possible for words to boost otherwise invisible pictures into awareness (Forder, Taylor, Mankin, Scott, & Franklin, 2016; Lupyan & Ward, 2013; Pinto et al., 2015). However, these studies cannot be taken as evidence that spoken word comprehension involves rapid activation of low-level visual representations because the paradigms they used did not tap into the time-course typically associated with semantic processing (see Hauk & Tschentscher, 2013 for discussion of the crucial importance of timing). For instance, Lupyan and Ward (2013) had at least 1s and Forder et al. (2016) several seconds between word and picture presentation. These authors used such long delays to address a different research question, namely the influence of expectations on perception, and thus used a set-up that encouraged the use of top-down expectations. Indeed, Pinto and colleagues' (2015) experiment 3 demonstrated that with such long presentation delays the effect disappears when participants are told that the cue word is not predictive of the target picture. Thus, it is unclear from the above research whether spoken word processing involves the rapid activation of low-level visual representations that is predicted by theories that ascribe an important role to modality-specific representation during processing.

The present study

In the present study, we implemented the CFS technique in a priming experiment to test whether spoken (object) words activate low-level visual representations. In particular, we tested whether participants can detect otherwise invisible objects when they are near-simultaneously presented with a spoken word (e.g. the word "bottle") and a suppressed picture (e.g. the picture of a bottle). In contrast to the studies mentioned above, we tested for rapid activation of visual representations by minimizing the delay between word and target picture presentation. Our hypothesis was that spoken object words can activate category-specific low-level visual representations rapidly and in a short time window. Access to category-specific visual representations was operationalized as facilitated detection of nearly invisible pictures that were masked with CFS and only briefly presented (for a duration of 400ms) shortly after word onset.

Experiment 1

Method

Participants

24 native Dutch speakers from the local Max Planck Institute database took part in the study. All participants had normal hearing and normal or corrected-to-normal vision. Five participants had to be excluded (two because of a detection rate <5%, two because of false alarm rates >50%, one because of technical failure) and were replaced by new participants to reach a total of 24 participants. This number should yield adequate statistical power based on similar previous research (Lupyan & Ward, 2013; Pinto et al., 2015).

Stimuli, apparatus and procedure

Participants were asked to wear custom-made prism glasses (prism dioptre: 10 Δ) as well as headphones and to put their head on a chinrest. A separator was placed centrally from the nose of the participant to the computer screen, such that each eye could only see the ipsilateral half of the screen. The stimuli were presented on a computer screen (resolution: 1900x720, refresh rate: 60Hz) using Presentation Software (Version 16.2, www.neurobs.com) in 80cm viewing distance which ensured optimal interocular fusion in this particular setup. The CFS masks were randomly selected every trial from a set of 50 Mondrian-type images composed of 1000 randomly superimposed rectangles of different colors and sizes, which were created in Matlab (similar to Stein, Seymour, Hebart, & Sterzer, 2016). The 16 target pictures were selected from the normed de Groot, Koelewijn, Huettig, and Olivers (2015) database, converted into greyscale and their edges blurred using a Gaussian filter with a radius of 3 pixels in Adobe Photoshop to facilitate suppression. The 16 spoken cue words (listed in the Appendix, mean length: 500ms, range: 287ms-685ms) were recorded by a female native Dutch speaker and cut into single files with Praat (Boersma, 2002).

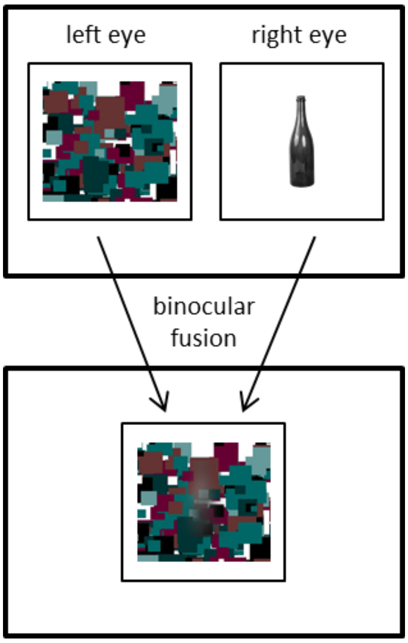


Figure 1: Binocular fusion. A picture was presented to the right eye, while the left eye saw CFS masks changing at ca. 10Hz.

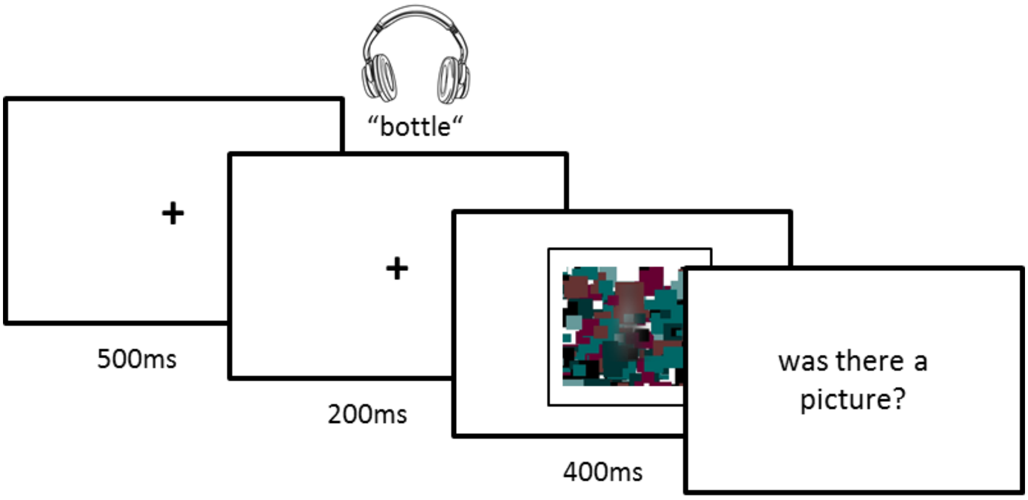


Figure 2: Trial structure. Note that the target picture was presented 200ms after word onset, but the words extended towards target picture offset (mean word length: 500ms).

At the beginning of each trial (see Figure 1 and 2), a central fixation cross was displayed alongside a black rectangular frame (600*600 pixels) in both halves of the screen for 500ms. Next, a spoken word was presented. In picture-present trials, 200ms after word onset the suppressed picture appeared: the CFS masks filled the black frame on one side (random Mondrian-type colorful rectangular shapes changing at a rate of 10Hz) and a grey-scale picture filled 350*350 pixels the other (on half of the trials this side stayed empty as there was no picture). In keeping with Lupyan & Ward (2013), the target pictures were presented to the right eye. The visual stimuli remained on the screen for 400ms and were then replaced by the question "Was there a picture?" to which participants were instructed to respond immediately by pressing the left or right button. On picture-absent trials, the 350*350 pixel area that would otherwise be covered by an image showed only an empty white square (everything else was identical to picture-present trials). Participants were informed that the words neither predicted whether a picture was presented on a given trial nor what picture appeared in picture-present trials. There were a total of 256 trials (corresponding to 8 repetitions of each of the 16 words and targets), including 128 picture-present trials of which 64 had congruent and 64 had incongruent prime words.

One prerequisite for this study was good control over the suppression strength such that, ideally, the detection rate of target pictures would be about 50%. To that end, we first conducted a short staircase procedure (96 trials) that was identical to the main study described above, except that no auditory cues were played and another set of pictures was used. In this staircase procedure, every hit resulted in a slight decrease and every miss in a slight increase of the target picture's contrast, such that the suppression strength usually converges roughly on 50% (as in Lupyan & Ward, 2013, Experiment 3). The final contrast level was then used in the main experiment to counter individual differences in visibility thresholds and avoid ceiling or floor effects.

Analysis

We implemented a design in which all words appeared equally often in picture-present (50%) and picture-absent trials (50%). In picture-present trials, all cue words were followed equally often by congruent (50%) and incongruent targets (50%). Thus, in half of the picture-present trials, the spoken word was congruent with the target picture ("bottle" -> picture of a bottle), while on the other half it was incongruent ("bottle" -> picture of a banana). We chose this design because it rules out any confounds due to differences in low-level features between conditions, as all picture stimuli are evenly distributed over the experimental conditions. Our main prediction that Congruency would lead to facilitated detection was tested in three ways: First, we analyzed detection rates in a binomial mixed effects model, including Congruency (congruent vs. incongruent) as fixed effect and random intercepts and slopes for the effect of Congruency by participant and target. Second, we analyzed d'-scores that take into account the false alarm rate each participant produced. This was achieved by calculating mean detection rates in the congruent and incongruent condition as well as false alarm rates per subject, then calculating d'-scores, and finally comparing the resulting scores in the congruent vs. incongruent condition in paired-samples t-tests. Third, we analyzed reaction times (RTs) on all picture-present trials with hits using a linear mixed effects model with the same fixed and random effects structure as the binomial model above. RTs slower than 2.5s and further than 2.5SDs from the resulting mean per condition were removed from the analysis.

Results

Of primary interest was whether congruent word cues facilitate the detection of almost invisible pictures. Inspection of the hit rate means suggests that this is indeed so (see Figure 3); in congruent trials the mean detection rate was 46.7%, whereas in incongruent trials the mean detection rate was only 41.7% (false alarm rate was 15%). A d'-analysis revealed higher d'-scores (t(23)=3.008, p=0.006) in the congruent (M=1.15, SD=0.63) vs. incongruent condition (M=0.99, SD=0.60). The binomial mixed effects model further backs up this result by revealing a significant difference in hit rates between the congruent vs. incongruent condition (coef=-0.2724, SE=0.1168, z=-2.332, p= 0.0197). Finally, a linear mixed effects model showed that response latencies in correct hit trials were reliably shorter in the congruent vs. incongruent condition (coef=46.30, SE=17.68, t=2.619, p=0.009).

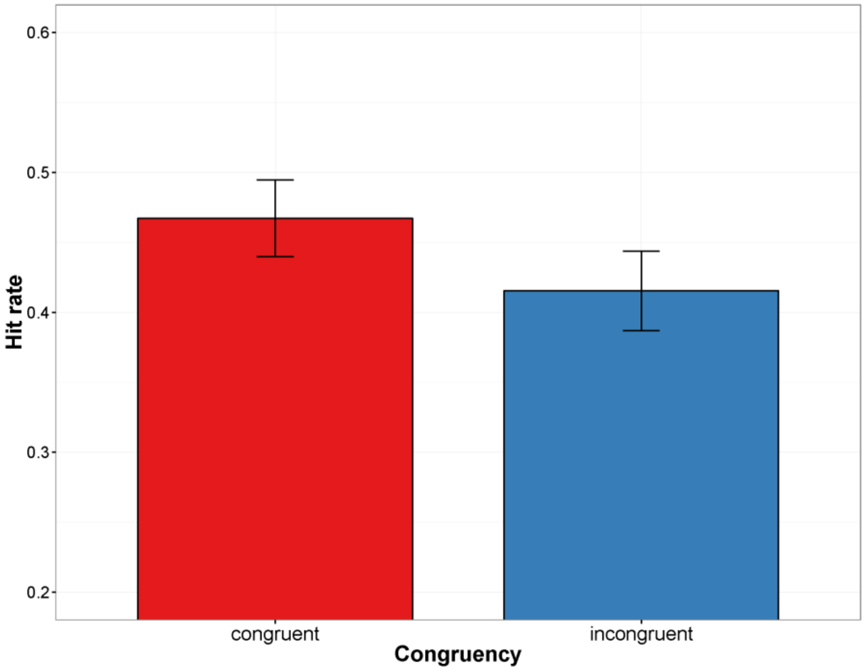


Figure 3: Experiment 1. Mean hit rates in the congruent and incongruent conditions. Error bars indicate 95% confidence intervals.

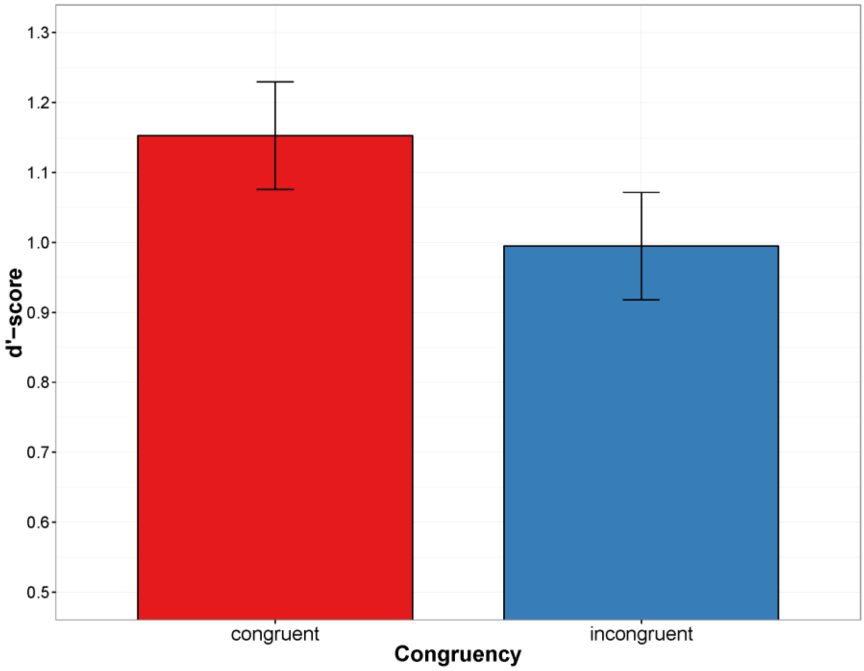


Figure 4: Experiment 1. Mean d'-scores in the congruent and incongruent conditions. Error bars indicate 95% confidence intervals.

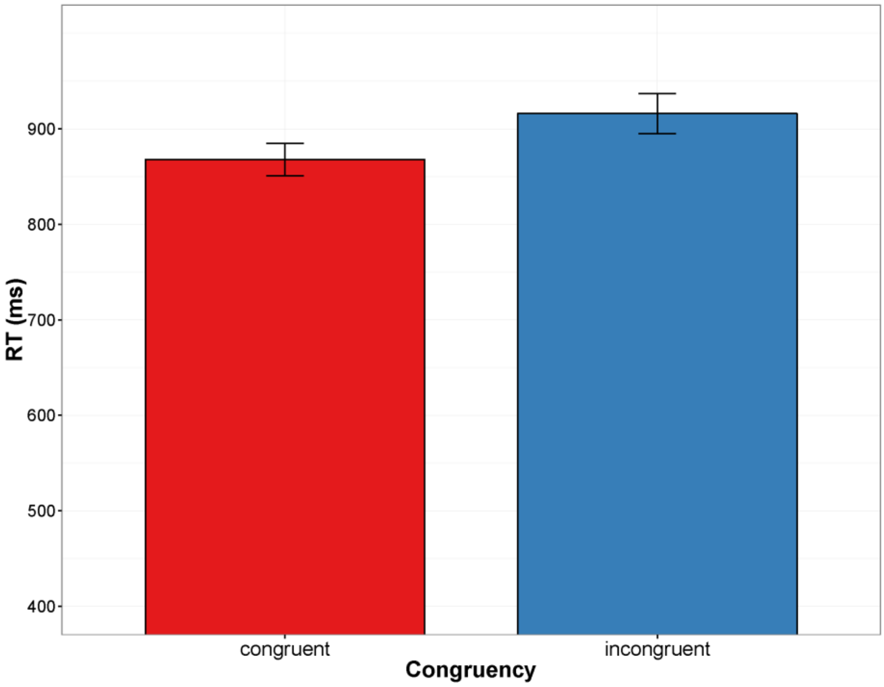


Figure 5: Experiment 1. Mean RTs in the congruent and incongruent conditions on all trials with hits after removing outliers (RTs>2.5s and >2.5SDs from the mean). Error bars indicate 95% confidence intervals.

Discussion

As predicted, processing spoken words made it more likely for participants to detect congruent compared to incongruent suppressed pictures. This result conceptually replicates previous CFS studies that showed that words can be used as cues to affect perception (cf. Lupyan & Ward, 2013; Pinto et al., 2015) and a recent EEG study (Boutonnet & Lupyan, 2015) that demonstrated that words (e.g., *cat*) but not equally informative environmental sounds (a meowing cat) modulate the P1 component in a subsequent picture matching task. Our results go further by demonstrating that spoken word comprehension results in rapid activation of low-level visual representations, as the spoken word and suppressed picture were presented near-simultaneously. We thus conclude that spoken words can activate category-specific visual representations that are low-level enough to influence a process as basic as the mere detection of a stimulus in line with modality-specific theories of conceptual representation. The timing of the spoken word and the following suppressed picture tells us that this effect is present roughly somewhere between 200 and 600ms post word onset. To further pin down its exact time-course, we conducted a second experiment in which the timing of the suppressed picture relative to the prime word was systematically manipulated.

Experiment 2

Experiment 2 was designed to replicate the effect we found in Experiment 1 and to identify its time-course. To that end, we modified the paradigm slightly such that the suppressed pictures were again presented for 400ms, but the timing of the picture relative to word onset was manipulated within-subjects. We were mainly interested in two aspects: First, we wanted to find out how early visual representations are activated during word processing, as this should be informative about the underlying processing dynamics of language comprehension. We predicted early recruitment (within the first 400ms), which would be consistent with the idea that the visual system contributes to conceptual processing of spoken words. Second, and related to that matter, we wanted to see how long-lasting the effect is. In the scenario we envisage, spoken words trigger a rapid and short-lived sweep of activation in visual cortex as part of information retrieval related to word meaning. From that the prediction follows that the effect is only present in the first hundreds of milliseconds post word onset and that it then vanishes. At first sight, this prediction may seem to contradict previous demonstrations of late effects of word cues on detection in CFS (Lupyan & Ward, 2013; Pinto et al., 2015). However, as discussed above, because of the differences in timing our experimental design likely taps mechanisms directly related to initial word processing, whereas these previous studies were designed to capture the effect of expectations on perception that may follow a different time-course.

Method

Participants

38 new native Dutch speakers with normal hearing and vision from the MPI database participated in the experiment. Five of them had to be excluded from the analysis (three because of a mean detection rate > 90%, two because of false alarm rates > 40%), leaving

us with 33 participants. The increased number of participants was chosen to make sure the study is powerful enough, given that we introduce a new factor (Timing) to the design.

Stimuli, apparatus and design

We used the same stimuli, apparatus and design as in Experiment 1 with the following exceptions: 1) We introduced the factor Timing to the design, which had four levels corresponding to SOA 1 (-200-200ms; picture appears 200ms before until 200ms after word onset), SOA 2 (0-400ms; picture appears at word onset until 400ms after word onset), SOA 3 (200-600ms; from 200ms after word onset until 600ms after word onset, exactly as in Experiment 1), and SOA 4 (600-1000ms; from 600ms after word onset until 1000ms after word onset). 2) To ensure sufficient statistical power we increased the number of trials to 384 and changed the ratio of picture-present vs. picture-absent trials to 2:1, as we were primarily interested in detected pictures. This made it possible to repeat each picture twice in each Timing condition per prime word. Again, before the main experiment, participants completed the staircase procedure described above to identify individual detection thresholds that were then used during the main experiment.

Analysis

We predicted that Congruency between the prime word and the suppressed picture would facilitate detection at SOA 2 and 3, but not at SOA 1 and 4. Regarding SOA 1 (-200ms), by the time that the word cue becomes informative the picture has already disappeared and it therefore should not have an effect. With respect to SOA 4 (600ms), we propose that detection performance is not affected by the word primes because it falls outside of the window typically associated with online semantic processing. Category-specific activation in visual cortex triggered by the prime word should therefore have decayed when the picture appears. To test these hypotheses, we used the same set of analyses as before, adapted for the current design as follows. First, the binomial mixed effects model included Congruency and Timing as well as their interaction as fixed effects and random intercepts and slopes for both fixed effects per-subjects and per-items. To analyze in which time-windows there was an effect, we used the function lsmeans from the lsmeans package that computes least-squares means (also called predicted marginal means) for all factor combinations and the contrasts among them. Second, we calculated d'-scores and submitted them to a 2x4 repeated-measures ANOVA for which we predicted a significant interaction. Follow-up paired-samples t-tests were used to determine the effect of Congruency at all four SOAs. Finally, for the RT data we ran a linear mixed effects model on all trials with detected pictures excluding responses slower than 2500ms, as well as responses further than 2.5 SDs from the mean calculated per condition. The model had the same fixed and random effects structure as the binomial model above and all effects were analyzed in the same way.

Results and Discussion

Of primary interest was at what SOA Congruency affected detection. Figures 6 and 8 suggest that detection was facilitated by congruent words at timing 2 and 3, but that there was no difference between congruent and incongruent trials at the other two timings. The 2x4 ANOVA on d'-scores revealed a main effect of Congruency (F(1, 32)=14.066, p=0.001, partial eta squared=0.305) and an interaction between Congruency and Timing (F(3, 96)=3.681, p=0.015, partial eta squared=0.103). Follow-up paired t-tests revealed that, indeed, Congruency had no effect on detection sensitivity at SOA 1 (mean d' difference: 0.048, t(32)<1) and SOA 4 (mean difference: 0.045, t(32)=1.087, p>0.2), while there was a reliable effect at SOA 2 (mean difference: 0.174, t(32)=3.349, p=0.002) and SOA 3 (mean difference: 1.191, t(32)=4.874, p<0.001). In line with this, the binomial mixed effects model on hit rates revealed a significant interaction between Congruency and Timing (Chisquare(3)=7.863, p=0.048), characterized by the pattern that hit rates were higher in the congruent vs. incongruent condition at SOA 2 (coef=0.348, SE=0.12, p=0.003) and 3 (coef=0.415, SE=0.122, p<0.001), but not at SOA 1 (coef=0.081, SE=0.121, p>0.5) and 4 (coef=0.108, SE=1.21, p>0.3)[1]. As a follow-up, we also conducted a logistic growth curve analysis (Mirman, 2016) with a second-order polynomial including fixed-effects of Congruency on both time terms and participant and participant-by-condition random effects on the quadratic term (the full model did not converge, so the linear term was removed because it was not expected to capture key differences). As expected from the results above, it revealed a significant main effect of Congruency (coef=-0.178, SE=0.049, p<0.001) and a significant effect of Congruency on the quadratic term

---

[1] False alarm rates were similar across SOAs (8.2%-10.3%).

(coef=0.218, SE=0.091, p=0.016). Figure 7 shows the model fit. In sum, the hit rate and d'-scores provide converging evidence that our design captured both the onset and offset of the detection facilitation effect.

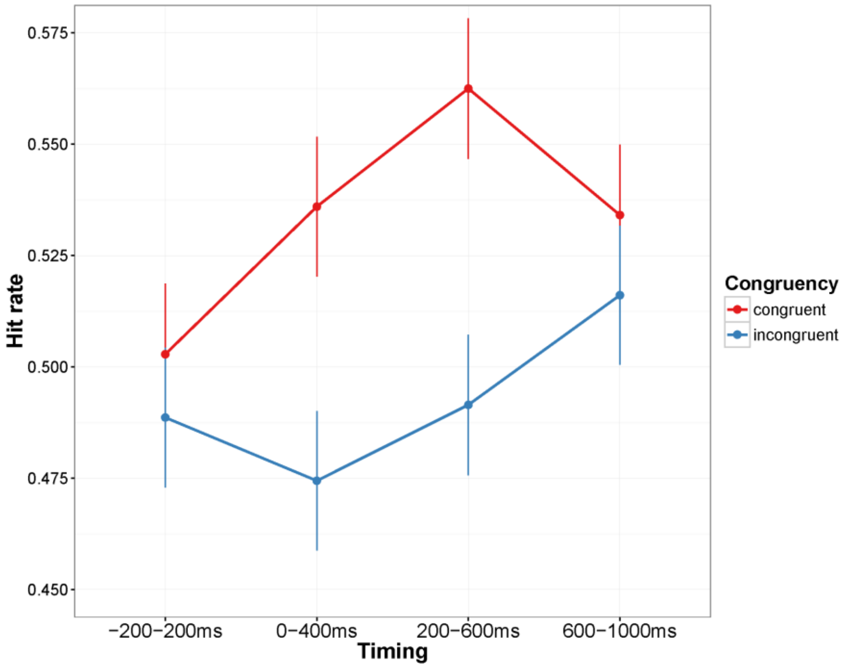


Figure 6: Experiment 2. Mean hit rates in all experimental conditions (congruent vs. incongruent at all four timing conditions). Error bars indicate the standard error.

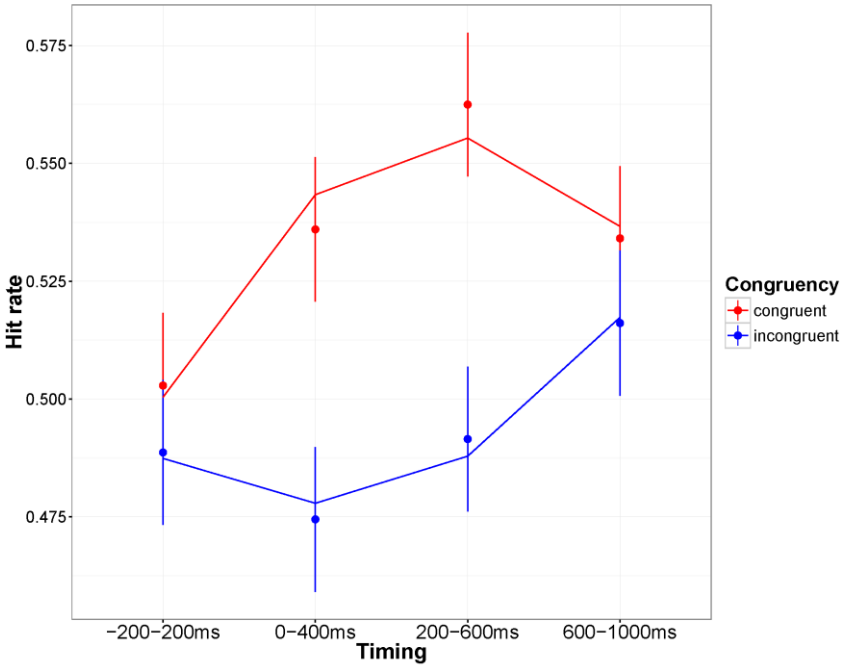


Figure 7: Experiment 2. Model fit of the logistic growth curve analysis depicted by plotting the observed data (dots correspond to the observed means) along with the predicted curve. Error bars indicate the standard error.

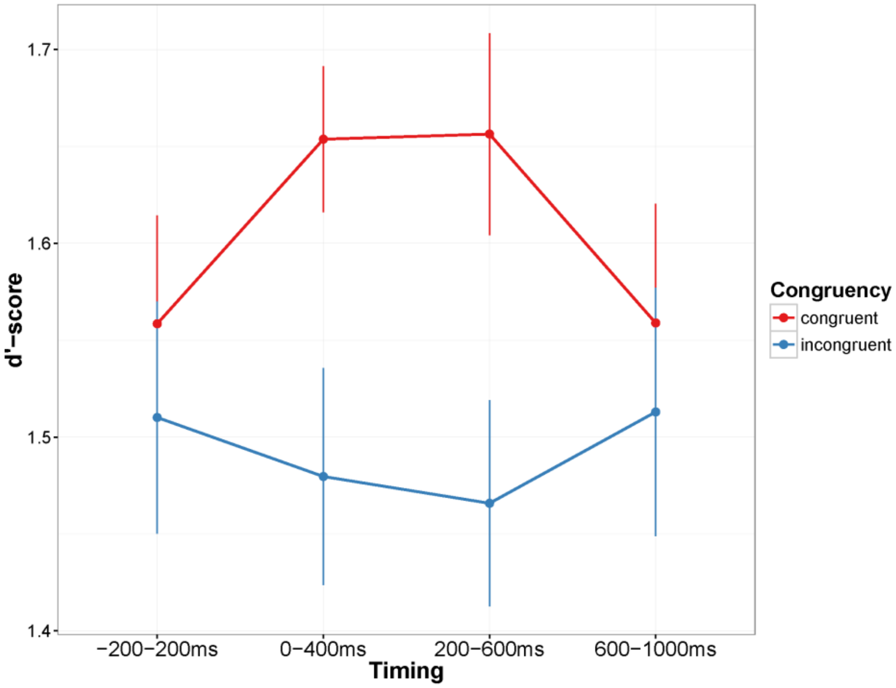


Figure 8: Experiment 2. Mean d'-scores in all experimental conditions (congruent vs. incongruent at all four timing conditions). Error bars indicate the standard error.

RT analyses revealed a main effect of Congruency (coef=0.144, SE=0.063, t=2.287, p=0.022), but no interaction with Timing (p>0.1). This result in combination with visual inspection suggests that RTs were influenced by Congruency at the earliest and latest SOAs. The most likely explanation for this seems to be that at SOA 1 and 4 Congruency affected participants' confidence in their decisions, rather than perceptual processing. Given the clear results from the hit rate and d' data, we consider our results to provide strong evidence for a time-dependent effect, as described below.

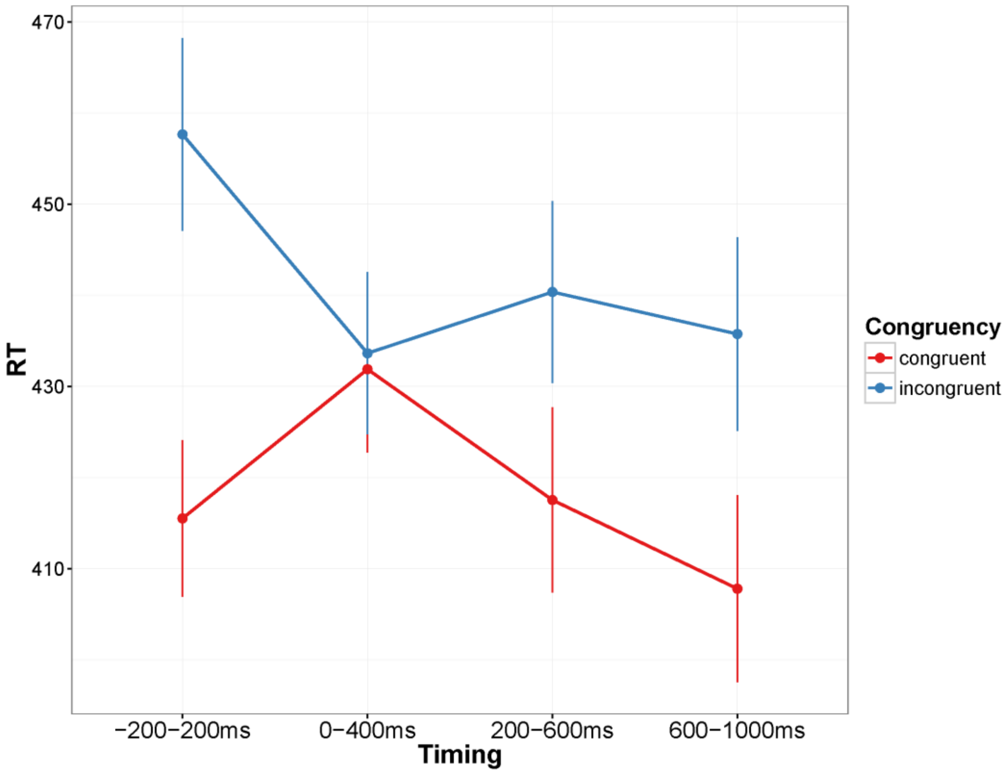


Figure 9: Experiment 2. Mean RTs in all experimental conditions (congruent vs. incongruent at all four timing conditions), calculated on trials with hits after removing outliers (RTs>2.5s and >2.5SDs from the mean). Error bars indicate the standard error.

General Discussion

The results of Experiment 2 replicate the effect we observed in Experiment 1 and also revealed the dynamic activation and deactivation of low-level visual representations over time during spoken word processing. In addition to our replication of the effect at timing 3, where the nearly invisible picture is presented 200-600ms after word onset, the novel effect at timing 2 (0-400ms) in combination with the absence of an effect at 200ms before until 200ms after word onset suggests that category-specific low-level visual representations were first accessed at 200-400ms after word onset. Strikingly, for most of our words this is right around the uniqueness point that disambiguates them from other possible words (and for three words even before). This is reminiscent of Lewis and Poeppel's (2014) MEG study that found a correlation between imageability and occipital cortex activation (in BA 19) at 160-190ms post onset, simultaneously or even before typical lexical effects (such as cohort and frequency effects) in temporal areas. A likely explanation for this even quicker activation compared to our study is that these authors only used monosyllabic words consisting of 3-4 phonemes that were simply shorter than the stimuli used here. Access to visual areas around the uniqueness point is consistent with a role for semantic processing. Furthermore, the absence of an effect at the latest time window (600-1000ms after word onset) suggests that the recruitment of low-level visual representations is limited to the first hundreds of milliseconds of spoken word processing. At around word offset, the effect has completely vanished suggesting a rapid decay after initial rapid activation.

Thus, we observed activation of visual representations in precisely the time-window that is associated with semantic access. Given the behavioral relevance of the activated representations for the detection paradigm we used, we can conclude that spoken words (referring to concrete concepts) activate modality-specific visual representations that are low-level enough to provide information related to the perception of a given token and at the same time abstract enough to be relevant not only for previously seen tokens (a signature of episodic memory) but also for generalizing to novel exemplars one has never seen before (a signature of semantic memory). A mechanism that could accomplish this is the activation of visual features that are diagnostic of the category the word refers to. Attention to a semantic category can shift the tuning of neuron populations in occipital and temporal cortex towards that category (Çukur, Nishimoto, Huth, & Gallant, 2013). Similarly, our results suggest that words may rapidly trigger the activation of low-level visual representations and as a result enhance processing of visual items sharing some of these representations.

Note that our present findings do not contradict the previous CFS studies reporting late cuing effects (e.g., Lupyan & Ward, 2013; Pinto et al., 2015). Their elegant studies add on to the increasing literature on how expectations can affect perception. We would like to point out that both their findings and our present results could be interpreted in a predictive coding framework and may rely on similar neural mechanisms (described in D'Esposito & Postle, 2015; Pratte & Tong, 2014; Scocchia, Valsecchi, & Triesch, 2014). Importantly however regarding our research question, the present study demonstrates the rapid involvement of low-level visual representations during *online* spoken word processing in a situation where the words were no reliable cues for the visual targets.

Future work could usefully investigate the underlying processing dynamics in terms of information exchange between sensorimotor systems on the one hand and high-level integration systems on the other. There is now good evidence that parts of the parietal lobe (especially the angular gyrus) and/or the anterior temporal lobe (ATL) are consistently involved in conceptual processing encompassing information from multiple modalities and it has been proposed that they could function as hubs that integrate information from distributed cortical sites (Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009; Patterson, Nestor, & Rogers, 2007; Xu, Lin, Han, He, & Bi, 2016). Strong evidence comes from studies showing modality-independent semantic impairments induced by ATL damage (e.g., Mummery et al., 2000) or by transcranial magnetic stimulation on ATL (Pobric, Jefferies, & Lambon Ralph, 2010a; 2010b), suggesting a functional role of ATL for conceptual processing. Using electrocorticography, Chen and colleagues (2016) recently managed to decode semantic information related to concepts from a wide array of categories from ATL activity patterns, supporting its category-general role for semantic processing. Similarly, Correia et al. (2014) were able to decode word meanings from ATL using fMRI.

Moreover, there are some clues as to what neural mechanisms could underlie communication between the different systems. ATL has been shown to engage in increased connectivity with modality-specific visual or auditory regions during semantic processing. When information from multiple modalities has to be combined, theta oscillations in ATL and distributed cortical networks become phase-locked as a function of the featural content required (van Ackeren PhD thesis, 2014, chapters 3 and 6; van Ackeren & Rueschemeyer, 2014; van Ackeren, Schneider, Müsch, & Rueschemeyer, 2014). Coutanche and Thompson-Schill (2014) found that the identity of expected objects participants were thinking about could be decoded from ATL patterns and that decoding accuracy was predicted by activation patterns in visual areas processing shape and color information related to the object, further pointing to the functional link between visual and high-level systems in conceptual processing. In the light of these findings, our data fit with the view that words trigger (potentially simultaneously) the activation of modality-specific as well as modality-independent systems that remain coupled for several hundreds of milliseconds to build a conceptual representation in concert.

Acknowledgements: We would like to thank Gary Lupyan and two anonymous reviewers for their helpful comments, as well as Laura Geurts and Jessica Aguilar Diaz for their help during data acquisition.

Appendix

Word stimuli: banana, crane, palm tree, rake, ball, syringe, pen, butterfly, brain, bottle, hat, cushion, rocket, flute, (a kind of) sweet

Corresponding Dutch words: banaan, kraan, palm, hark, bal, spuitje, pen, vlinder, hersens, fles, hoed, kussen, raket, fluit, bonbon

References

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of vision*, *15*(7).

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617-645.

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

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767-2796.

Bi, Y., Wang, X., & Caramazza, A. (2016). Object Domain and Modality in the Ventral Visual Pathway. *Trends in cognitive sciences*, *20*(4), 282-290.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5*(9/10), 341-345.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience*, *35*(25), 9329-9335.

Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., ... & Lambon Ralph, M. L. (2016). The 'when'and 'where'of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex*, *79*, 1-13.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. *Cerebral Cortex*, bhu057.

Cox, D., & Hong, S. (2015). Semantic-Based Crossmodal Processing During Visual Suppression. *Name: Frontiers in Psychology*, *6*, 722.

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience*, *34*(1), 332-338.

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *The Journal of Neuroscience*, *34*(14), 4766-4775.

Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning From Vision. *Trends in cognitive sciences*, *19*(11), 677-687.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature neuroscience*, *16*(6), 763-770.

De Groot, F., Koelewijn, T., Huettig, F., & Olivers, C. N. (2015). A stimulus set of words and pictures matched for visual and semantic similarity. *Journal of Cognitive Psychology*, 1-15.

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M. S. (2009). Activation of sensory–motor areas in sentence comprehension. *Cerebral Cortex*, bhp115.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual review of psychology*, 66, 115.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. Cognition, 143, 93-100. http://doi.org/doi: 10.1016/j.cognition.2015.06.008.

Forder, L., Taylor, O., Mankin, H., Scott, R. B., & Franklin, A. (2016). Color Terms Affect Detection of Color and Color-Associated Objects Suppressed from Visual Awareness. *PloS one*, *11*(3), e0152212.

Francken, J. C., Meijs, E. L., Ridderinkhof, O. M., Hagoort, P., de Lange, F. P., & van Gaal, S. (2015). Manipulating word awareness dissociates feed-forward from feedback models of language-perception interactions. *Neuroscience of Consciousness*, *2015*(1), niv003.

Francken, J. C., Meijs, E. L., Hagoort, P., van Gaal, S., & de Lange, F. P. (2015). Exploring the automaticity of language-perception interactions: Effects of attention and awareness. *Scientific reports*, 5.

Gayet, S., Van der Stigchel, S., & Paffen, C. L. (2014). Breaking continuous flash suppression: competing for consciousness on the pre-semantic battlefield. *Frontiers in psychology*, 5.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage*, *32*(2), 906-912.

Hauk, O., & Tschentscher, N. (2013). The body of evidence: what can neuroscience tell us about embodied semantics?. *Frontiers in Psychology*, *4*, 50.

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

Hong, H., Yamins, D. L., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature neuroscience*, *19*(4), 613-622.

Hwang, K., Palmer, E. D., Basho, S., Zadra, J. R., & Müller, R. A. (2009). Category-specific activations during word generation reflect experiential sensorimotor modalities. *Neuroimage*, *48*(4), 717-725.

Kang, M. S., Blake, R., & Woodman, G. F. (2011). Semantic analysis does not occur in the absence of awareness induced by interocular suppression. *The Journal of Neuroscience*, *31*(38), 13535-13545.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & 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.

Lewis, G., & Poeppel, D. (2014). The role of visual representations during the lexical access of spoken words. *Brain and language*, *134*, 1-10.

Lupyan, G., & Spivey, M. J. (2010a). Making the invisible visible: auditory cues facilitate visual object detection. PLoS ONE, 5(7), e11452. http://doi.org/10.1371/journal.pone.0011452.

Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. Attention, Perception, & Psychophysics, 72(8), 2236-2253. http://doi.org/10.3758/APP.72.8.2236.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. Journal of Experimental Psychology-General, 141(1), 170-186. http://doi.org/10.1037/a0024904.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196-14201.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*(3), 397-405.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25-45.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs language affects low-level visual perception. *Psychological Science*, *18*(11), 1007-1013.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788-804.

Mirman, D. (2016). *Growth curve analysis and visualization using R*. CRC Press.

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.

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *The Journal of Neuroscience*, *34*(20), 6721-6735.

Patterson, K., Nestor, P. J., & 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.

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of vision*, *15*(8), 13-13.

Pratte, M. S., & Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *Journal of vision*, *14*(3), 22-22.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576-582.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of cognitive neuroscience*, *17*(6), 884-892.

Shtyrov, Y., Butorina, A., Nikolaeva, A., & Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences*, *111*(18), E1918-E1923.

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, *15*(10), 1602-1608.

Scocchia, L., Valsecchi, M., & Triesch, J. (2014). Top-down influences on ambiguous perception: the role of stable and transient states of the observer. *Frontiers in human neuroscience*, *8*.

Stein, T., Thoma, V., & Sterzer, P. (2015). Priming of object detection under continuous flash suppression depends on attention but not on part-whole configuration. *Journal of vision*, *15*(3), 15.

Stein, T., Hebart, M. N., & Sterzer, P. (2011). Breaking continuous flash suppression: a new measure of unconscious processing during interocular suppression?. *Frontiers in human neuroscience*, *5*.

Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2014). Rapid fear detection relies on high spatial frequencies. *Psychological science*, *25*(2), 566-574.

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. L. (2010). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current Biology*, *20*(10), 964-968.

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

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, 24, 2218-2225.

Tanaka, Y., & Sagi, D. (1998). A perceptual memory for low-contrast visual signals. *Proceedings of the National Academy of Sciences*, 95(21), 12729-12733.

van Ackeren, M. J. (2014). *Oscillatory neuronal dynamics during lexical-semantic retrieval and integration* (Doctoral dissertation, University of York).

van Ackeren, M. J., & Rueschemeyer, S. A. (2014). Cross-modal integration of lexical-semantic features during word processing: evidence from oscillatory dynamics during EEG. *PloS one*, 9(7), e101042.

van Ackeren, M. J., Schneider, T. R., Müsch, K., & Rueschemeyer, S. A. (2014). Oscillatory neuronal activity reflects lexical-semantic feature integration within and across sensory modalities in distributed cortical networks. *The Journal of Neuroscience*, *34*(43), 14318-14323.

Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *NeuroImage*, *124*, 107-117.

Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, 2, 116.

Xu, Y., Lin, Q., Han, Z., He, Y., & Bi, Y. (2016). Intrinsic functional network architecture of human semantic processing: Modules and hubs. *NeuroImage*, *132*, 542-555.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological science*, *13*(2), 168-171.

Chapter 3

A TASK-DEPENDENT CAUSAL ROLE FOR LOW-LEVEL VISUAL PROCESSES IN SPOKEN WORD COMPREHENSION

Based on: Ostarek, M., & Huettig, F. (2017). A Task-Dependent Causal Role for Low-Level Visual Processes in Spoken Word Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1215.

Abstract

It is well established that the comprehension of spoken words referring to object concepts relies on high-level visual areas in the ventral stream that build increasingly abstract representations. It is much less clear whether basic low-level visual representations are also involved. Here we asked in what task situations low-level visual representations contribute functionally to concrete word comprehension using an interference paradigm. We interfered with basic visual processing while participants performed a concreteness task (Experiment 1), a lexical decision task (Experiment 2), and a word class judgment task (Experiment 3). We found that visual noise interfered more with concrete than with abstract word processing, but only when the task required visual information to be accessed. This suggests that basic visual processes can be causally involved in language comprehension, but that their recruitment is not automatic and rather depends on the type of information that is required in a given task situation.

Introduction

In recent decades, the view that the conceptual system underlying language comprehension is to a non-negligible extent based on sensory-motor (and potentially other kinds of) experience has been very dominant. In recognition of the symbol grounding problem (Harnad, 1990), most contemporary theories of conceptual processing assume that it relies at least partly on modality-specific systems (e.g., Barsalou, 1999; 2008; 2016; Binder & Desai, 2011, Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012). Binder and colleagues (2016), for instance, have proposed a framework of conceptual processing based on a set of primitive experiential features that capture the various kinds of experience associated with a given concept (sensory, motor, affective, social, etc.). Their approach focuses on the kinds of cognitive/neural experience related to a concept meaning that, for instance, all neural states triggered by encounters with dogs are part of the experience related to the concept *dog*. A model based on such experiential attributes was recently used to predict brain activation patterns in response to sentences (Anderson et al., 2016). Relevant for the present study, their model predicts that object concepts tend to rely on visual processes. An important question is what types of visual processes are involved during (e.g. spoken) word processing. Here, we investigated the recruitment of low-level visual processes during object word processing and their dependence on task demands.

The visual system has been implicated in the processing of object concepts in the past decades based on neuroimaging studies that localized their neural correlates to high-level visual cortex that is generally assumed to build increasingly holistic object representations along the posterior-to-anterior axis during perception (Bracci & Op de Beeck, 2016; Di Carlo & Cox, 2007). Initial univariate fMRI studies on object recognition localized selective activation in response to pictures of animate vs. inanimate items to distinct portions of fusiform gyrus and lateral occipital cortex (Chao, Haxby, & Martin, 1999; Chao, Weisberg, & Martin, 2002; for review see Martin, 2007). More recently, several fMRI studies using multivariate pattern analysis (MVPA) reported successful decoding of conceptual categories based on picture stimuli from ventral temporal areas (Haxby et al., 2001; Shinkareva et al., 2008). Similar results have been obtained using written words (Devlin, Rushworth, & Matthews, 2005) as well as spoken words

(Noppeney, Price, Penny, & Friston, 2006), and recent MVPA studies managed to decode the conceptual category across different input modalities from activity patterns in ventral occipitotemporal cortex (Shinkareva et al., 2011; Simanova et al., 2014). Finally, several studies suggest that lesions in visual association cortex lead to specific impairments for concepts that depend on visual knowledge (Capitani et al., 2009; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Tranel, Damasio, & Damasio, 1997). Overall, there seems to be a fair amount of consensus that high-level visual representations are consistently used for conceptual processing of concrete objects because they provide representations that are both abstract and grounded (Barsalou, 2016; Binder & Desai, 2011; Binder, Desai, Graves, & Conant, 2009; Fernandino et al., 2015a; Martin, 2007), although it is certainly still debated to what extent sensory systems and conceptual systems overlap and to what extent they are separate (Binder, 2016; Goldinger et al., 2016; Mahon, 2015).

For the question of what types of visual processes are involved it is important to note that the relative consistency of the neural systems recruited by certain types of concepts does not imply conceptual stability. For one thing, even the typical category effects (such as inanimate vs. animate) are localized to varying cortical areas across studies (Thompson-Schill, 2003; Devlin et al., 2002; Martin & Chao, 2001; for a review see Gerlach, 2007). Moreover, context effects in language processing are ubiquitous (Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008; Kiefer & Martens, 2010; Lebois, Wilson-Mendenhall, & Barsalou, 2015; Nieuwland & van Berkum, 2006; Solomon & Barsalou, 2004; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; van Berkum, Brown, & Hagoort, 1999; Yee, Ahmed, & Thompson-Schill, 2012), such that, for instance, even the blueness of the concept BLUE is not always activated (Besner, Stolz, & Boutilier, 1997; Goldfarb, Aisenberg, & Henik, 2011; Jacoby, Lindsay, & Hessels, 2003; see Kemmerer, 2014b for discussion) and different modality-specific systems can be recruited for the same object words depending on whether the task stresses visual or action properties (Hoenig et al., 2008). These findings suggest that conceptual processing is not adequately described as the retrieval of fixed conceptual representations but rather as a constructive process in which concepts are 'built on the spot' recruiting different kinds of representational systems depending on task demands. In such an account it is conceivable that conceptual processing makes use of basic visual processes to serve high-level (object) word comprehension, at least when detailed visual information is useful. Such an account receives support from recent studies reporting the activation of low-level representations in the motor domain (Hauk, Johnsrude, & Pulvermüller, 2004; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014) and the visual domain (Correia et al., 2014; Desai, Binder, Conant, & Seidenberg, 2009; Hwang, Palmer, Basho, Zadra, & Müller, 2009; Mitchell et al., 2008; Ostarek & Huettig, 2017). We are aware that it is not straightforward to draw a clear-cut low vs. high level distinction in modality-specific systems. However, we find it useful to distinguish between processes related to individual local features that tend to be processed in early visual areas and processes related to holistic perception of entire objects that tends to rely on ventral temporal areas.

The present study

Based on these considerations, it seems plausible that language comprehension can be achieved by different processing channels that vary dynamically with contextual demands. One such channel, on which we focused in this study, is the access to low-level visual processes. There is some neuroimaging evidence that object words can activate occipital cortex (Correia et al., 2014; Hwang, Palmer, Basho, Zadra, & Müller, 2009; Lewis & Poeppel, 2014; Mitchell et al., 2008), and we have recently found that object words rapidly activate basic low-level visual processes related to the mere detection of a stimulus using continuous flash suppression (Ostarek & Huettig, 2017). However, the *functional role* of low-level visual processes for concrete word comprehension across different processing situations is still unknown. One important step was made by an elegant study (Meteyard, Zokaei, Bahrami, & Vigliocco, 2008) showing interference of dot patterns moving up vs. down with the comprehension of verbs that imply motion in the corresponding direction (e.g., with *rise* vs. *fall*). In the present study, we used a paradigm allowing us to test for a functional involvement of basic visual processes during semantic processing of a class of words that rely on visual knowledge more generally. We interfered with basic visual processing while participants were performing different tasks on individual nouns to investigate whether and under what circumstances low-level visual representations are *causally* involved in the comprehension of concrete words. We used dynamic Mondrian-type masks changing at ca. 10Hz consisting of rectangles of different colors and sizes. Thereby interference was targeted at visual processes related to edge detection, the perception of colors and orientations, processes that are typically considered low-level as they are the basic building blocks upon which higher-level representations depend.

While it is generally agreed that knowledge about visual features is crucial for our understanding of concepts referring to physical objects (e.g., Huettig, Mishra, & Olivers, 2012) the long-standing and crucial question of the nature of the representations underlying this knowledge remains controversial. The flexible conceptual processing account described in the Introduction predicts that different kinds of features are recruited in situations that require different kinds of information. In line with this, we tested the hypothesis that low-level visual processes are recruited specifically in situations where in-depth visual information needs to be retrieved. The rationale we followed in this research is that if (and only if) visual information associated with words is retrieved (at least partly) from low-level visual representations and these are functionally involved in the comprehension process, then interfering with basic visual processing should result in a cost for object word processing. We manipulated the task requirements across three experiments to assess the extent to which dynamic low-level visual interference hindered the comprehension of concrete words when retrieval of visual information was task relevant (concreteness judgement task, Experiment 1), when visual information was irrelevant and semantic processing was minimal (lexical decision task (LDT), Experiment 2), and when visual information was irrelevant but semantic processing was necessary (word class judgment task, Experiment 3).

Experiment 1

In Experiment 1, we measured reaction times (RTs henceforth) to highly visually imageable concrete words and low imageability abstract words in a concreteness judgement task with and without concurrent visual interference. We predicted that concrete words should be affected more strongly by visual noise because the affirmative decision that a word refers to a physical object should heavily rely on visual information represented in brain regions processing low-level visual information. While it is conceivable that abstract word meanings are grounded in sensory experience (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Vigliocco et al., 2014), it seems unlikely that visual features are accessed with a degree of specificity comparable to that required for object word processing. RTs to abstract words thus served as a baseline condition allowing us to separate possible general distraction effects of the visual noise manipulation from 'functional interference' effects related specifically to semantic processing of visually dominant concepts. We thus predicted an interaction between word type and visual interference, such that responses to concrete words should be slowed down more than responses to abstract words.

Method

Participants

50 native Dutch speakers participated in the experiment. They all had normal or corrected-to-normal vision and hearing, were paid and gave written consent. Participant numbers were determined in advance on the basis that approximately 50 participants should yield ample statistical power to detect the effects of interest in this kind of semantic task, given previous work (e.g., Neely, Keefe, & Ross, 1989). Data collection was stopped when this number was reached. The study was approved by the ethics board of the faculty of social sciences of the Radboud University.

Stimuli and apparatus

The visual stimuli were presented on a computer screen (resolution: 1900x720, refresh rate: 60Hz) in 55cm viewing distance using Presentation Software (Version 16.2, www.neurobs.com). They included a central fixation cross to initiate each trial and the visual interference patterns (random Mondrian-type colorful rectangular shapes changing at 12Hz) which are typically used for continuous flash suppression studies because they interfere strongly with low-level visual processing (see Tsuchiya & Koch, 2005, for discussion). 52 highly visually imageable concrete words (henceforth "concrete words") and 52 low imageability abstract words (henceforth "abstract words") that were selected on the basis of their high vs. low visual imageability were recorded by a female Dutch speaker using Audacity Software. A norming study with the 50 participants of Experiment 1 (conducted after the main experiment) showed that the two word types indeed differed strongly in visual imageability[2] (likert scale 1-7; concrete words: mean=6.4, SD=0.2; abstract words: mean=2.9, SD=0.8; $p<0.001$), which was used as a proxy to gauge how visually dominant the concepts are. The word types also differed significantly ($p<.001$) in terms of concreteness ratings (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014); on a likert scale from 1-5 abstract words had a mean of 2.2 (0.7), concrete words 4.7 (0.3). Words from the two word types were matched for length (concrete: mean=518ms, SD=72ms vs. abstract: mean=530ms, SD=79ms, $t<1$), number of syllables (abstract: mean=1.8, SD=0.5, concrete: mean=1.6, SD=0.5, $t<1$), number of phonemes (abstract: mean=5.0, SD=0.9, concrete: mean=4.8, SD=1.1, $t<1$), and frequency (abstract: mean=63.6, SD=70.6, concrete: mean=64.9, SD=94.9, $t<1$) using the SUBTLEX-NL database (Keuleers, Brysbaert, & New, 2010). A list of all stimuli is in the Appendix.

Procedure

At the beginning of each trial a fixation cross appeared at the center of the screen for 800ms. Immediately afterwards participants heard a spoken word accompanied either by a blank screen or by visual noise. Participants were instructed to decide as quickly and as accurately as possible whether the spoken word was concrete or abstract by pressing one of two buttons with their left or right index finger (RTs were measured from word onset). When a response was made or 2s had elapsed after word onset a blank screen replaced the visual noise patterns (when present) and was shown for 1000ms after which the next trial started. Every word was presented twice; once with and once without visual noise to obtain a 2 (Word Type; concrete object vs. abstract) x 2 (Visual Condition; flashes vs. blank screen) within-subjects design and a total of 208 experimental trials. The experiment was organized in two blocks with all 104 spoken words presented in random order. Half of the words were presented alongside visual noise in the first block and half of them in the second block (the order was counterbalanced between participants). This

---

[2] In particular, participants were asked how strong a mental image they experienced when hearing the word (1 = no specific image whatsoever, 7 = a very strong and clear mental image).

means that both Word Type and Visual Condition were mixed pseudo-randomly in both blocks.

Analysis

Six items were removed because they produced mean error rates greater than 25% and one because of a technical problem[3] (6.7% of all trials). Trials with incorrect responses (3.8%) and RTs slower than 2000ms (0.9%) were excluded from the analysis. The remaining data set was trimmed for outliers by removing trials with RTs that were more than 2.5SDs from the mean (2.8%). This amounts to 1508 trials (14.5%) removed in total. RTs were analyzed with linear mixed effects models (Baayen, Davidson, & Bates, 2008) implemented in RStudio (Version 0.98.945, 2009-2013 RStudio, Inc.) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). The full model contained Word Type and Visual Condition as well as the interaction between the two as fixed effects, and by-participants random slopes and intercepts for both fixed effects and by-items slopes and intercepts for the effect of Visual Condition[4]. *β*-values (representing estimates based on scaled RTs), t-values and p-values were obtained by using the function get_pvalues from the psy811 package. Additionally, for the critical interaction the full model was compared to a second model from which only the interaction term was removed using a likelihood ratio test. Finally, to estimate the effect size of the predicted difference in the impact of visual noise on concrete vs. abstract words we calculated cohen's d based on least-square means using the lsmeans package.

Results and Discussion

There was a main effect of Word Type (β=-.33; SE=.07; t=-4.65; p<.001), reflecting faster responses to concrete words. This commonly observed "concreteness effect" is typically attributed to richer representations of concrete words arising from differences in imageability or context availability between concrete and abstract words (Paivio, 1986; 2013; 2014; Schwanenflugel, 1991; Schwanenflugel & Shoben, 1983; Schwanenflugel, Harnishfeger, & Stowe, 1988). There was also a main effect of Visual Condition (β=-.13; SE=.03; t=4.76; p<.001) reflecting slower responses on trials with visual noise. Critically,

---

[3] The audio file corresponding to the word *verlies* (loss) was corrupted such that it was very difficult to recognize it.

[4] A model that additionally included random effects for the interaction term did not converge.

there was a robust interaction between Word Type and Visual Condition (β=-.1; SE=.04; t=3.00; likelihood ratio test: $\chi^2(1)=9.01$; p=.002). As predicted, visual noise affected responses to concrete words more strongly. The mean difference was 21ms with Cohen's d=0.61 which is commonly interpreted as a medium effect size.

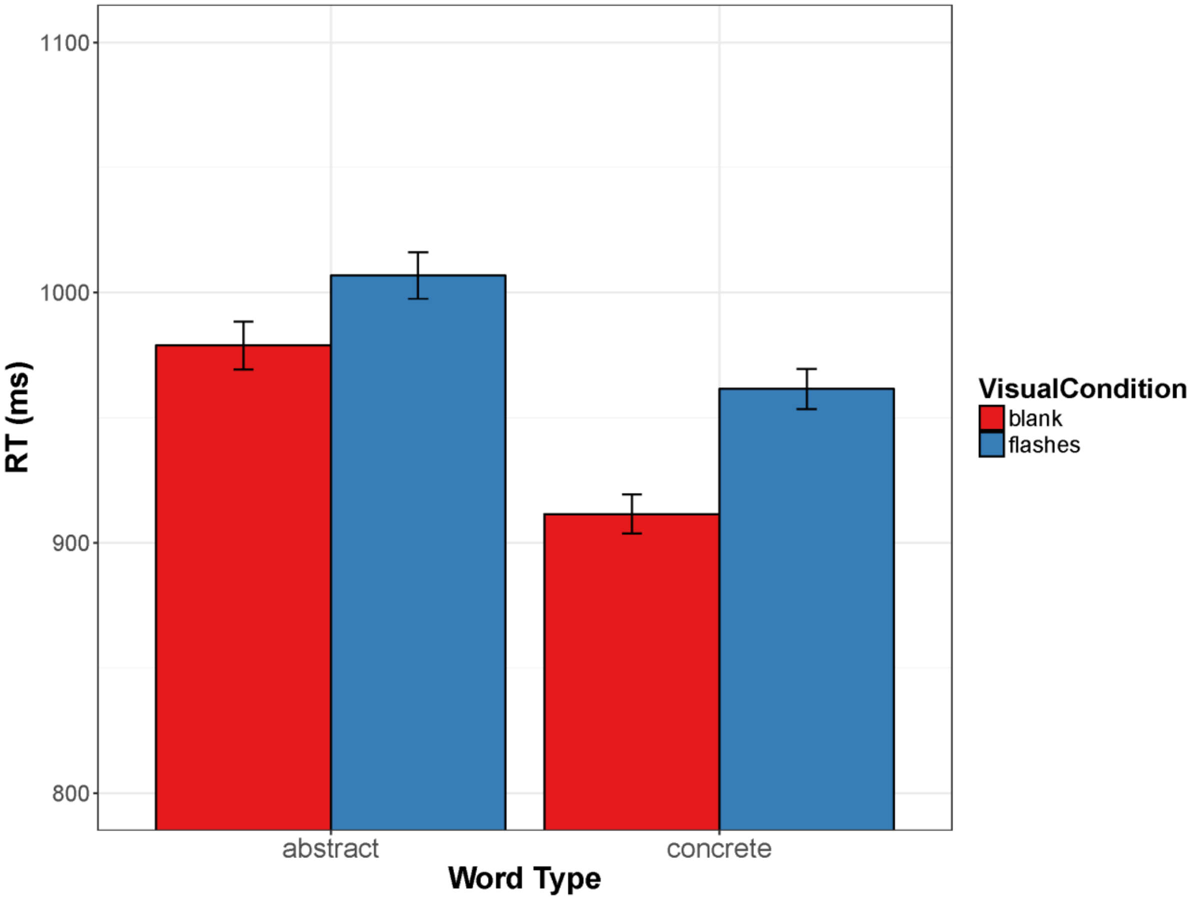


Figure 1: Experiment 1. Mean RTs of all four conditions, error bars indicate 95% confidence intervals.

Our results therefore suggest that concrete words rely more strongly on visual features and, more importantly, that the visual system was recruited for the comprehension of concrete words. The present finding that visual interference disrupts semantic processing of concrete words constitutes direct evidence that basic low-level visual processes can be causally involved in word comprehension. One study along these lines reported that dot patterns moving up or down at near-threshold interfered with the comprehension of verbs with up/down association (rise vs. fall) when their direction was incongruent with the verb meaning (Meteyard, Zokaei, Bahrami, & Vigliocco, 2008). While that study focused on one particular perceptual feature of a class of verbs, our data speak to the comprehension of concrete words more generally. What is striking about our findings is that the visual interference technique we used specifically interferes with basic low-level processes, as it only consists of horizontal and vertical components and colors, all of which recruit early visual cortical regions (see Yuval-Greenberg & Heeger, 2013). This indicates that participants did not solely rely on high-level representations but built low-level visual representations on the spot. The recruitment of visual processes typically involved in local feature processing is consistent with the online construction of an idiosyncratic visual representation typical of bottom-up processing and episodic memory retrieval (Danker & Anderson, 2010; Ueno et al., 2007). We propose that, in the current task situation, spoken words activate processes akin to conscious visual perception in order to solve the sensory judgement task.

Experiment 2

Though the results of Experiment 1 are striking, it is of crucial importance for our research question to determine to what extent the visual interference effect reported here is task-dependent. Does spoken object word processing always rely on low-level visual processes? Or, alternatively, are low-level visual representations only involved in spoken word processing when the task requires retrieval of visual information? We tested this issue by repeating the Visual Noise study in an auditory lexical decision task (LDT).

Note that although access of semantic information may not be strictly necessary to perform this task, a vast number of lexical decision studies have shown that semantic information is typically accessed and speeds up RTs (e.g., Bentin, McCarthy, & Wood, 1985; Fischler, 1977; Holcomb & Neville, 1990; Hutchison, 2003; Perea & Rosa, 2002; Thompson-Schill, Kurtz, & Gabrieli, 1998). This makes it an interesting testing ground for the hypothesis that task demands determine what kinds of information are retrieved during language processing: As detailed visual knowledge about word referents is not required to perform a LDT, such accounts predict no specific interference of visual noise with object word processing. If, however, sensory representations are activated automatically and have a functional contribution whenever spoken concrete words are heard (e.g., Gallese & Lakoff, 2005), the Visual Condition * Word Type interaction from Experiment 1 should persist.

Method

Participants

46 native Dutch participants with normal or corrected-to-normal vision and hearing completed the experiment. They gave written consent and were paid for participation. Sample size and stopping rule were determined in the same way as in Experiment 1.

Stimuli, Apparatus and Procedure

Stimuli, apparatus and procedure were identical to Experiment 1, with the following exceptions: 42 instead of 52 items per word type were used[5] and 168 pseudo-words added, which resulted in a total of 336 trials. We used the program Wuggy (Keuleers & Brysbeart, 2010) to create pseudo-words that differed in one or two phonemes from existing words making the lexical decision relatively difficult. Participants had to indicate as quickly and as accurately as possible whether the word they heard was an existing Dutch word or not by pressing one of two buttons.

Results and Discussion

Four participants had to be excluded because of error rates greater than 25%. The data of the remaining 42 participants were trimmed for errors (2.4%), timed-out responses (1.1%), and outliers (2.4%). The remaining data (a total of 1103/14.3% of trials was removed) were analyzed in the same way as in Experiment 1 with the exception that the model only included by-participants random slopes for Visual Condition, as the full model did not converge. There was a main effect of Visual Condition (β=.24, SE=.03; t=8.774, p<.001) reflecting slower responses in trials with visual noise but no main effect of Word Type (t<1) and no statistically significant interaction (t<1).

---

[5] We removed the error-prone items from Experiment 1 and some additional items to keep the duration of the experiment similar to Experiment 1 without sacrificing statistical power, at the same time making sure that the experimental conditions remained well-matched in terms of frequency etc.

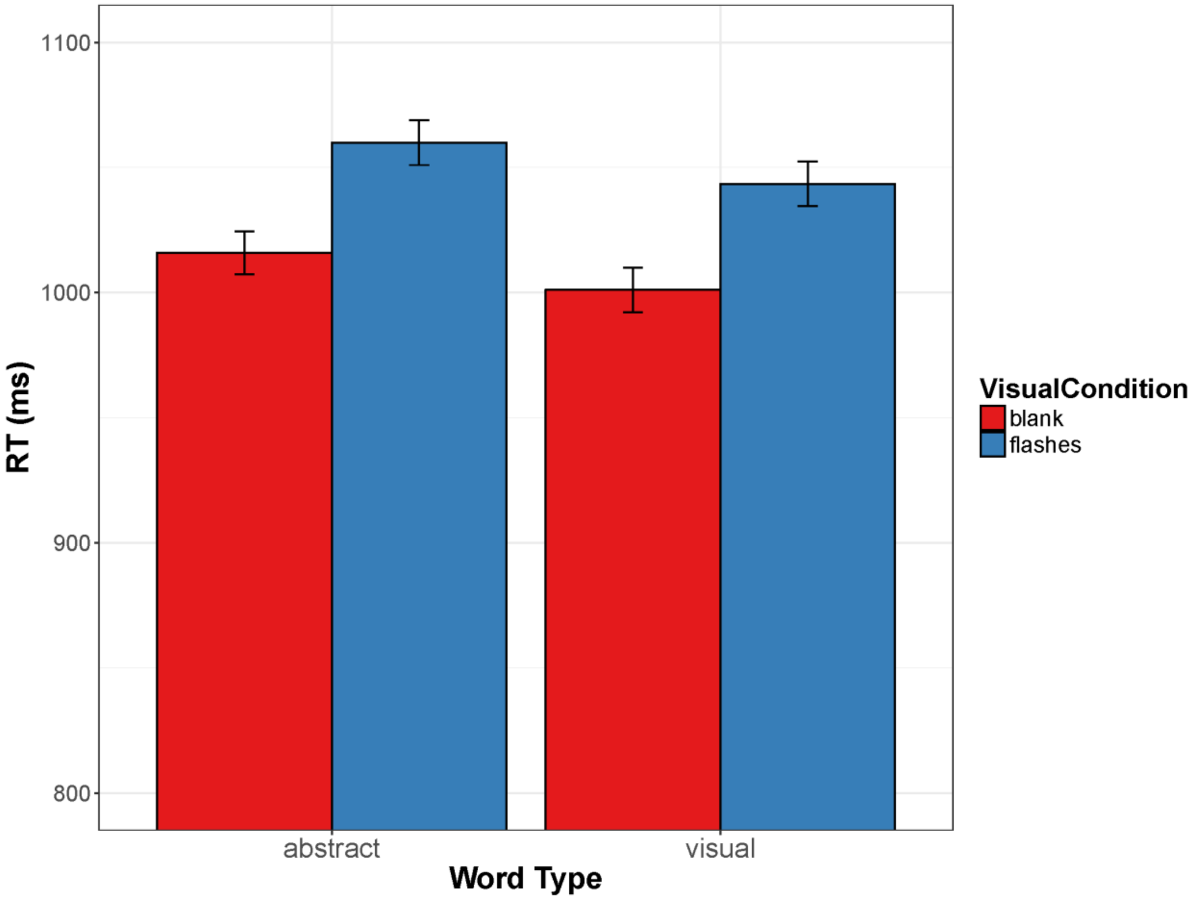


Figure 2: Experiment 2. Mean RTs of all four conditions, error bars indicate 95% confidence intervals.

These results suggest that in the context of a LDT, visual representations were not functionally involved in object word processing. Experiment 2 thus constrains theories of conceptual processing because it shows that the functional relevance of the visual system for language processing is contingent on task requirements. Our results of course do not conclusively rule out that low-level visual representations were activated in the present task. What we can conclude however is that if they were activated they did not have a detectable functional role. Note that the reason for this is unlikely to be lacking sensitivity of our technique, as Experiment 1 clearly shows that our paradigm can reveal an involvement of low-level visual processes when it is present.

Despite the clear evidence of semantic and (lexical) visual priming in many previous lexical decision studies one may question what information our participants activated in

Experiment 2. There is some evidence that words automatically activate non-associative semantic information, such as information about perceptual features. Some studies found priming effects among words with perceptually similar referents (*pizza-coin*) using lexical decision (d'Arcais, Schreuder, & Glazenborg, 1985; Schreuder, d'Arcais, & Glazenborg, 1984). On the other hand, Pecher, Zeelenberg, and Raaijmakers (1998) pointed out several methodological shortcomings with these data and, using an improved design, only observed this effect when participants performed a perceptual categorization task before the priming experiment but not in a standard LDT. These findings and our present data thus suggest that visual representations are not automatically retrieved whenever spoken words are heard.

Experiment 3

Since we cannot be sure that the LDT used in Experiment 2 involved semantic access we conducted a third experiment to further investigate the extent to which low-level visual representations are accessed in tasks requiring semantic access. To this end we used a word class decision task (is the word you are presented with a noun or adjective?). While there are certainly several ways by which the word class of a word can in principle be determined (Kemmerer, 2014a), it seems unlikely that semantic access is completely bypassed. At the same time, it is sufficient to access highly schematic conceptual representations that are largely shared among nouns to solve the task: Nouns are non-gradable, non-relational, temporally stable entities, whereas adjectives refer to gradable qualities and are used for modification (Croft, 2001; Kemmerer, 2014a). As such, there is no need to access detailed visual feature information. If low-level visual representations are automatically accessed and are functionally relevant during semantic spoken word processing regardless of contextual demands, visual noise should nevertheless interfere more with word class decisions to concrete nouns than abstract nouns (i.e. a significant visual noise by word type interaction). Alternatively, according to the flexible processing account, basic visual processes could be recruited specifically in situations where in-depth visual information needs to be retrieved, in which case concrete words should not be affected more by visual noise than abstract words.

Method

Participants

48 native Dutch participants with normal or corrected-to-normal vision and hearing completed the experiment. They gave written consent and were paid for participation. Sample size and stopping rule were determined in the same way as in Experiment 1 and 2.

Stimuli, Apparatus and Procedure

Stimuli, apparatus and procedure were identical to Experiment 1 and 2, with the following exceptions: We used 40 concrete and abstract nouns and 80 adjectives which were all presented twice, resulting in a total of 320 trials. Participants had to indicate as quickly and as accurately as possible whether the word they heard was a noun or an adjective by pressing one of two buttons. Similar to the nouns, adjectives were half concrete (e.g., *diagonaal* (diagonal), *bewolkt* (cloudy); mean concreteness rating: 3.6, SD=0.5) and half abstract (e.g., *populair* (popular), *ongewoon*, (unusual); mean=2.6, SD=0.8) and differed significantly on this dimension (p<.001), such that these factors were not confounded. Abstract and concrete adjectives were also matched for frequency (concrete: mean=24.6, SD=46.8; abstract: mean=30.0; SD=53.6; t<1), number of phonemes (concrete: mean=5.5, SD=2.0; abstract: mean=5.5, SD=2.0; t<1), and number of syllables (concrete: mean=1.8, SD=0.8; abstract: mean=2.0, SD=0.9; t<1).

Results

Two participants had to be excluded due to high error rates (>25%). The data from the remaining 46 participants were trimmed for errors (6.5%), time-outs (1.9%), and outliers (2.1%). A total of 1129 trials (14.8%) were thus removed. A linear mixed-effects model (equal in structure to the one used in Experiment 1) revealed a main effect of Visual Condition (β=.206; t=6.178; SE=.033; p<.001) with slower responses in visual noise trials. There was also a main effect of Word Type (β=-.511, SE=.083; t=-6.151, p<.001) with slower responses to abstract words, as in Experiment 1. Critically, the interaction between the two was not significant (t=-1.06). Thus, visual noise again affected concrete and abstract words to the same extent.

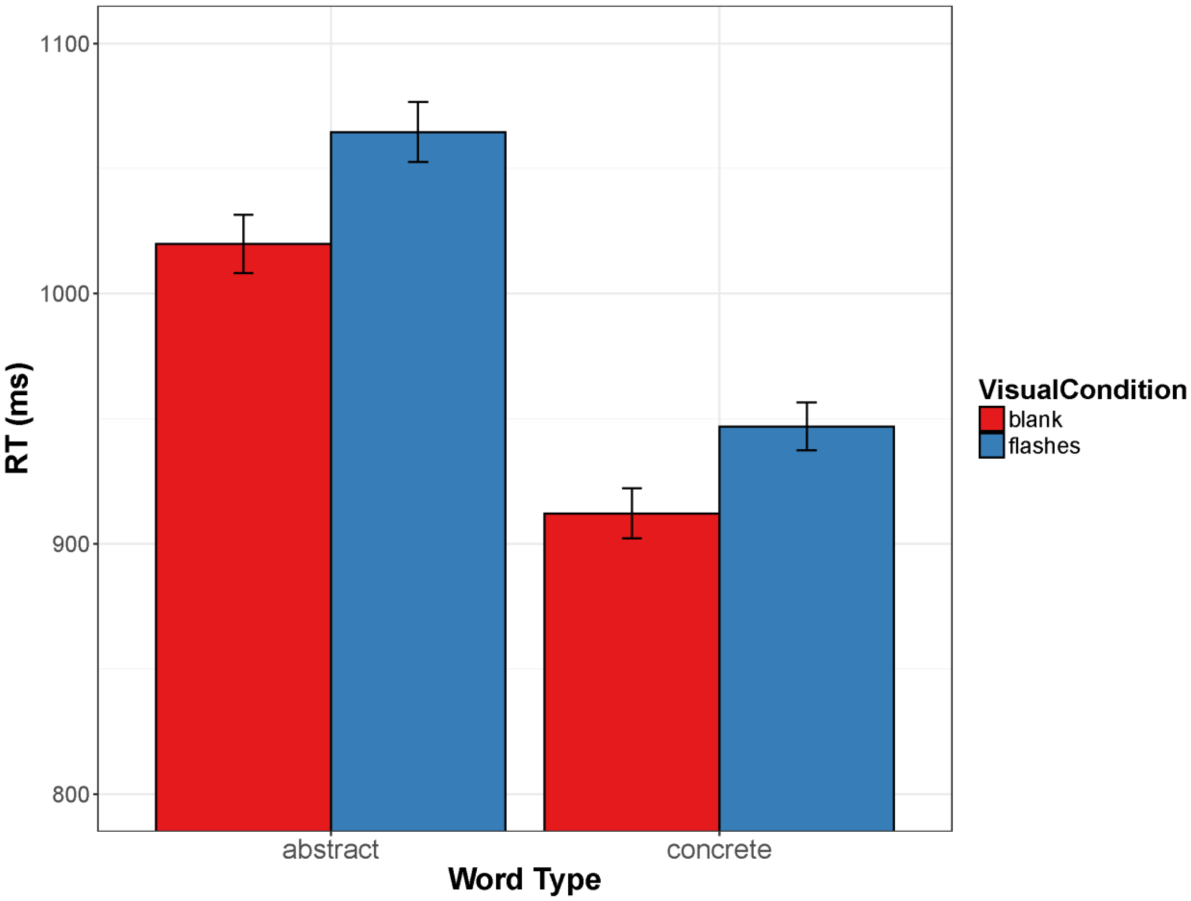


Figure 3: Experiment 3. Mean RTs of all four conditions, error bars indicate 95% confidence intervals.

Similar results were observed when analyzing the adjectives (see Figure 4): There was an effect of Visual Condition (β=.11; SE=.024; t=4.986; p<.001), but no effect of Word Type and no interaction (t<1).

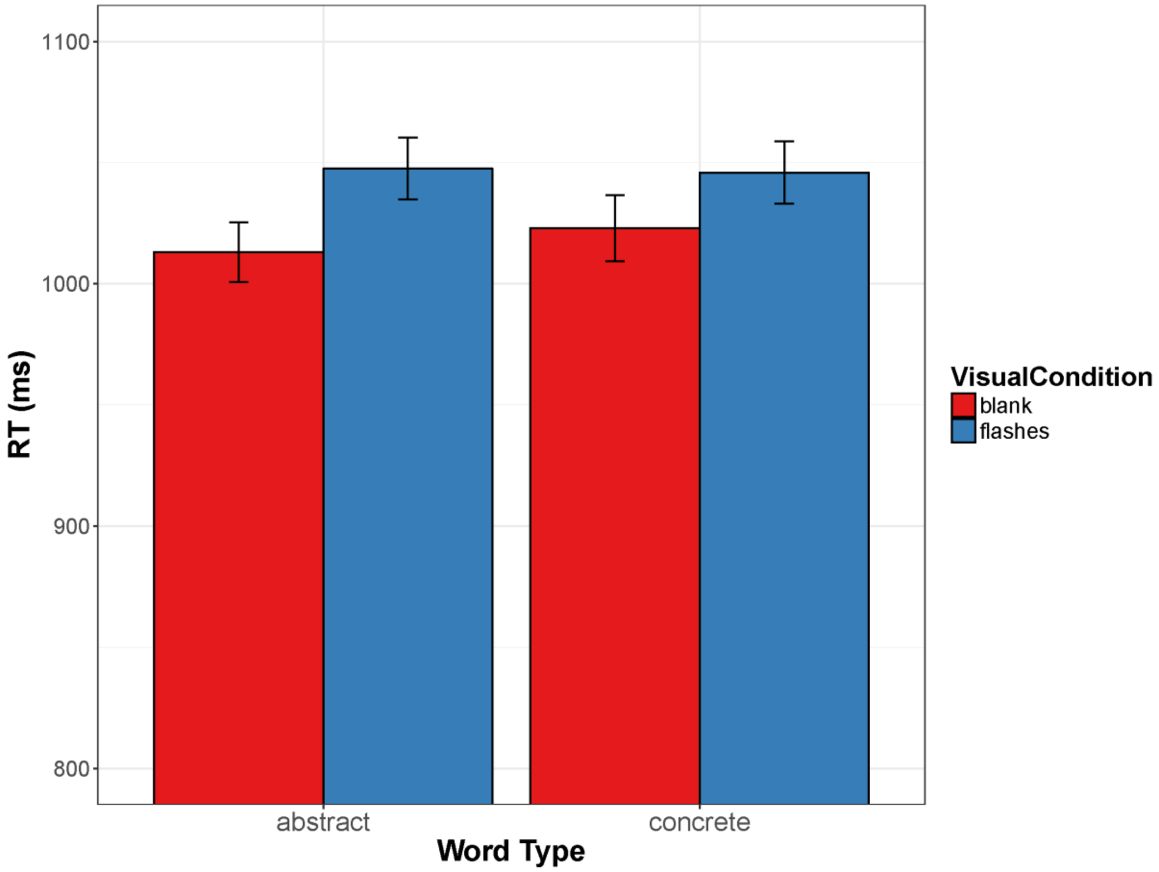


Figure 4: Experiment 3, adjectives. Mean RTs of all four conditions, error bars indicate 95% confidence intervals.

Follow-up analysis and Discussion

Our results across three experiments indicate that low-level visual representations are only functionally relevant in task situations that require visual information to be retrieved and that other semantic tasks do not rely on their contribution. As Table 4 indicates, visual noise interfered much stronger with concrete object word processing relative to abstract word processing in Experiment 1, but not in Experiments 2 and 3.

	Experiment 1	**Experiment 2**	**Experiment 3**
Concrete words	50ms	43ms	34ms
Abstract words	28ms	44ms	45ms
Interference score	**22ms**	**-1ms**	**-11ms**

Table 1: Overview of interference effects across experiments. Interference was calculated for each word type and experiment by subtracting mean RTs in blank screen trials from mean RTs in visual noise trials. The "interference score" represents the difference in interference between both word types (interference on concrete words - interference on abstract words).

To assess statistically whether the effect was specific to the concreteness task, we conducted a post-hoc mixed effects model with Experiment (1 vs. 2 vs. 3) as an additional factor. There was indeed a three-way interaction between Visual Condition, Word Type, and Experiment (likelihood-ratio test: $\chi^2(2)=8.412$; $p=0.0149$). The effect of visual noise on concrete words was larger in Experiment 1 compared to Experiment 2 ($\beta=0.117$; SE=.055; t=-2.145; p=.0319) and compared to Experiment 3 ($\beta=0.147$; SE=.055; t=-2.681; p=.007). The difference between Experiments 1 and 3 is particularly important. While it is possible that participants relied on phonological processing for the LDT, the noun vs. adjective decision task requires semantic access. This would also explain the stronger similarity of Experiment 1 and 3 in terms of RT distributions, as evidenced by a main effect of Word Type (the "concreteness effect") that was absent in Experiment 2. The absence of an interference effect of visual noise on concrete word processing in a task that requires semantic - but not specifically visual - information supports the view that low-level visual processes are not reflexively recruited for object word comprehension. Rather, they constitute an optional processing channel for situations that rely on the retrieval of visual features. It is important to note that Experiment 1 clearly shows that the interference technique we used is capable of revealing an involvement of low-level visual processes when it is present. Therefore, the absence of interference effects in Experiments 2 and 3 are unlikely due to a lack of sensitivity of the method but rather attributable to the lack of a functional role of low-level visual processes in these tasks.

General Discussion

In the present study, we used a visual interference paradigm to study the extent to which low-level visual representations are causally involved in spoken word comprehension. Our main finding is that dynamic Mondrian masks that are known to hinder basic low-level visual processing slowed down responses to high-imageability concrete words, as compared to low-imageability abstract words, but only in a task situation that required visual information to be accessed. This result is in line with embodied theories of conceptual processing only to the extent that they are embedded in a flexible processing framework (Hoenig et al., 2008; Lebois et al., 2015). Note that a central aspect of our study is the significant interaction between visual noise and word type (with a bigger effect of visual noise on visually imageable concrete than abstract words). This allows us to separate (potential) general distraction effects of the visual noise manipulation from 'functional interference' effects. Distraction effects likely account for the main effect of visual noise but cannot account for the visual noise by word type interaction observed in Experiment 1.

It is important to point out that we are agnostic with regard to whether the visual representations retrieved in the present study constitute mental imagery (cf. Kosslyn, Ganis, & Thompson, 2001; Pearson, Naselaris, Holmes, & Kosslyn, 2015) or what sometimes is referred to as 'simulations' (e.g., Barsalou, 1999; 2008). In our view there is currently no evidence available that conclusively shows that imagery and simulations are distinct. What our findings show is that low-level visual processes are causally involved in semantic processing but (crucially) only when the task requires it, regardless of whether one prefers to call these processes imagery or simulation.

We observed that visual interference affected object word processing more than abstract word processing in a concreteness task (Experiment 1), whereas it affected abstract and concrete words to the same extent in a lexical decision task (Experiment 2) and a word class decision task (Experiment 3). This suggests that task demands have a strong impact on the processing channels that are used for semantic processing. We are not the first to argue for flexible *activation* of modality-specific representations (Hoenig et al., 2008; Lebois et al., 2015; van Dam et al., 2012). What has not been shown before however is that the *functional role* of low-level visual representations is task-dependent. Our data suggest that visual processes are functionally relevant when in-depth visual information is required (e.g., about the physical properties of a word's referent, as in a concreteness task), while other sources of information are preferentially accessed for tasks that do not require such information. This view is supported by a recent study by Edmiston and Lupyan (2017) who found that visual interference impairs participants' ability to make use of visual but not encyclopedic knowledge during word processing.

Our paradigm specifically assessed whether visual representations were necessary for the tasks at hand, not whether they were activated. For instance, in a continuous flash suppression study we recently found evidence that passive listening of object words activates visual processes involved in a task as basic as visual detection (Ostarek & Huettig, 2017). Similarly, a recent magnetoencephalography study (Lewis & Peoppel, 2014) using an auditory LDT found a correlation between imageability and activity in visual association cortex within the first 200 ms after word onset. In the light of findings like these, we do not intend to rule out an involvement of the visual system in the LDT and word class decision experiments, but we can tell from the data that it did not have a functional role in these tasks.

Related to this, we do not argue that visual representations are secondary for object word comprehension. To illustrate, nobody would claim that blueness is a secondary feature of "blue", just because a Stroop task only leads to interference in certain conditions, but not others (Besner, Stolz, & Boutilier, 1997; Goldfarb, Aisenberg, & Henik, 2011; Jacoby, Lindsay, & Hessels, 2003; see Kemmerer, 2014b for discussion). What our study shows is that modality-specific visual representations are causally involved in word comprehension if the task situation is conducive for the use of visual information.

To further bolster this conclusion, future studies could investigate whether the amount of interference is predicted by the relevance of visual attributes, which can be assessed with attribute ratings (Binder et al., 2016). This seems to follow from recent fMRI studies establishing a relationship between visual aspects of word meaning and brain activity in visual brain areas (Fernandino et al., 2015a; 2015b). Furthermore, different kinds of interference could be used to target specific features, such as color, motion, or shape.

Our data fit within a framework of flexible and dynamical conceptual processing encompassing low-level modality-specific as well as high-level systems. It would be important for future work to characterize the exact roles, representational nature and interplay of these systems during online language processing. First steps have recently been made in a series of MEG studies that identified enhanced connectivity between anterior temporal lobe (ATL) and modality-specific sensory areas during semantic processing (van Ackeren, 2014). In the light of these findings, two exciting possibilities emerge: 1) Information related to the different multimodal aspects of the given word meaning is retrieved from multiple sensory areas and then converges in ATL where it might be bound (e.g., Patterson, Nestor, & Rogers, 2007; Pobric, Jefferies, & Lambon Ralph, 2010). In that case ATL would contain compressed multimodal information that may suffice for some semantic tasks. 2) ATL combines contextual factors and stimulus-driven input to pick out those distributed cortical areas that contain relevant information and activates them in concert, possibly binding them via theta-phase locking (Varela, Lachaux, Rodriguez, & Martinerie, 2001). In that case ATL might not contain semantic knowledge, but have a critical control function necessary for semantic access. Our results suggest that such long-range networks can span all the way into early visual regions, but that their functional role depends heavily on task requirements.

References

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Aguilar, M., ... & Raizada, R. D. (2016). Predicting Neural Activity Patterns Associated with Sentences Using a Neurobiologically Motivated Model of Semantic Representation. *Cerebral Cortex*.

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of vision*, *15*(7), 3-3.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and brain sciences*, *22*(04), 637-660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617-645.

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic bulletin & review*, 1-21.

Bates D, Maechler M, Bolker BM and Walker S (2014). “lme4: Linear mixed-effects models using Eigen and S4.” ArXiv e-print

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and clinical neurophysiology*, *60*(4), 343-355.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic bulletin & review*, *4*(2), 221-225.

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

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3-4), 130-174.

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

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767-2796.

Bracci, Stefania, and Hans Op de Beeck. "Dissociations and associations between shape and category representations in the two visual pathways." *The Journal of Neuroscience* 36.2 (2016): 432-444.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta psychologica*, *150*, 80-84.

Capitani, E., Laiacona, M., Pagani, R., Capasso, R., Zampetti, P., & Miceli, G. (2009). Posterior cerebral artery infarcts and semantic category dissociations: a study of 28 patients. *Brain*, awp013.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience*, *2*(10), 913-919.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex*, *12*(5), 545-551.

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience*, *34*(1), 332-338.

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1), 179-229.

d'Arcais, G. F., Schreuder, R., & Glazenborg, G. (1985). Semantic activation during recognition of referential words. *Psychological Research*, *47*(1), 39-49.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychological bulletin*, *136*(1), 87.

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M. S. (2009). Activation of sensory–motor areas in sentence comprehension. *Cerebral Cortex*, bhp115.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J., & Tyler, L. K. (2002). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia*, *40*(1), 54-75.

Devlin, J. T., Rushworth, M. F., & Matthews, P. M. (2005). Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia*, *43*(1), 69-74.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, *11*(8), 333-341.

Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, *92*, 281-292.

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

Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., & Binder, J. R. (2015b). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia*, *76*, 17-26.

Fischler, I. (1977). Associative facilitation without expectancy in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 18.

Gerlach, C. (2007). A review of functional imaging studies on category specificity. *Journal of Cognitive Neuroscience*, *19*(2), 296-314.

Goldfarb, L., Aisenberg, D., & Henik, A. (2011). Think the thought, walk the walk–Social priming reduces the Stroop effect. *Cognition*, *118*(2), 193-200.

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, 1-20.

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

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-2430.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and cognitive processes*, *5*(4), 281-312.

Hong, H., Yamins, D. L., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature neuroscience*, *19*(4), 613-622.

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

Huettig, F., Mishra, R. K., & Olivers, C. N. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology, 2,* 394.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785-813.

Hwang, K., Palmer, E. D., Basho, S., Zadra, J. R., & Müller, R. A. (2009). Category-specific activations during word generation reflect experiential sensorimotor modalities. *Neuroimage*, *48*(4), 717-725.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*(3), 638-644.

Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: Event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research*, *10*(1), 67-75.

Kemmerer, D. (2014a). 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*, 1-8.

Kemmerer, D. (2014b). Word classes in the brain: Implications of linguistic typology for cognitive neuroscience. *Cortex*, *58*, 27-51.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, *42*(3), 627-633.

Keuleers, E., Brysbaert, M. & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods, 42*(3), 643-650.

Kiefer, M., & Martens, U. (2010). Attentional sensitization of unconscious cognition: task sets modulate subsequent masked semantic priming. *Journal of Experimental Psychology: General,* 139(3), 464.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635-642.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14.

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects. *Cognitive Science*, *39*(8), 1764-1801.

Lewis, G., & Poeppel, D. (2014). The role of visual representations during the lexical access of spoken words. *Brain and language*, *134*, 1-10.

Mahon, B. Z. (2015). What is embodied about cognition?. *Language, cognition and neuroscience*, *30*(4), 420-429.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25-45.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, *37*(4), 547-559.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788-804.

Meteyard, L., Zokaei, N., Bahrami, B., & Vigliocco, G. (2008). Visual motion interferes with lexical decision on motion words. *Current Biology*, *18*(17), R732-R733.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *science*, *320*(5880), 1191-1195.

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *The Journal of Neuroscience*, *34*(20), 6721-6735.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience*, *18*(7), 1098-1111.

Noppeney, U., Price, C. J., Penny, W. D., & Friston, K. J. (2006). Two distinct neural mechanisms for category-selective responses. *Cerebral Cortex*, *16*(3), 437-445.

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible: Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*.

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

Paivio, A. (2013). *Imagery and verbal processes*. Psychology Press.

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.

Patterson, K., Nestor, P. J., & 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.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in cognitive sciences*, *19*(10), 590-602.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological research*, 66(3), 180-194.

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

Schreuder, R., d'Arcais, G. B. F., & Glazenborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychological Research*, 45(4), 339-354.

Schwanenflugel, P. (1991). Why are abstract concepts hard to understand? In. PJ Schwanenflugel (Ed.), The psychology of word meaning (pp. 223-250).

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499-520.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 82.

Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, *3*(1), e1394.

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. *Neuroimage*, *54*(3), 2418-2425.

Shtyrov, Y., Butorina, A., Nikolaeva, A., & Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences*, *111*(18), E1918-E1923.

Simanova, I., Hagoort, P., Oostenveld, R., & Van Gerven, M. A. (2014). Modality-independent decoding of semantic information from the human brain. *Cerebral cortex*, *24*(2), 426-434.

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring “how” from “where”. *Neuropsychologia*, *41*(3), 280-292.

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language*, *38*(4), 440-458.

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, *35*(10), 1319-1327.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature neuroscience*, *8*(8), 1096-1101.

Ueno, A., Abe, N., Suzuki, M., Hirayama, K., Mori, E., Tashiro, M., ... & Fujii, T. (2007). Reactivation of medial temporal lobe and occipital lobe during the retrieval of color information: A positron emission tomography study. *Neuroimage*, *34*(3), 1292-1298.

van Ackeren, M. J. (2014). *Oscillatory neuronal dynamics during lexical-semantic retrieval and integration* (Doctoral dissertation, University of York).

Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of memory and language*, *41*(2), 147-182.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience*, *2*(4), 229-239.

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, *24*(7), 1767-1777.

Vigliocco, G., Vinson, D. P., Lewis, W., & 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.

Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *NeuroImage*, *124*, 107-117.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (can) prime furiously. *Psychological science*, *23*(4), 364-369.

Yuval-Greenberg, S., & Heeger, D. J. (2013). Continuous flash suppression modulates cortical activity in early visual cortex. *The Journal of Neuroscience*, *33*(23), 9635-9643.

Appendix

List of stimuli in Experiment 1

Dutch Object words	**English Translation**	**Dutch Abstract words**	**English Translation**
kamer	room	advies	advice
straat	street	affaire	affair
school	school	aspect	aspect
schip	ship	schuld	debt
water	water	debuut	beginning
auto	car	bedrog	cheating
tafel	table	excuus	excuse
hoofd	head	factor	factor
lichaam	body	winst	profit
vogel	bird	spel	game
stoel	chair	kwestie	question
film	film	manier	way
boom	tree	maand	month
dokter	doctor	motief	motive
eekhoorn	squirrel	steun	support
ketel	boiler	begin	beginning
schild	shield	pardon	pardon
hengel	fishing rod	fase	phase
honing	honey	pleidooi	plea
laars	boot	gerucht	rumor
zebra	zebra	verhaal	story
tijger	tiger	theorie	theory
sneeuw	snow	comfort	comfort
pudding	pudding	deugd	virtue
dolfijn	dolphin	geluk	happiness
haai	shark	gemak	ease
foto	photo	hitte	heat
bloem	flower	honger	hunger
gitaar	guitar	moed	courage
papier	paper	troost	consolation
trein	train	energie	energy
winkel	shop	vrijheid	freedom
vliegtuig	plane	roddel	gossip
krant	newspaper	vreugde	joy
voet	foot	kennis	knowledge
vinger	finger	wet	law
slang	snake	verlies	loss
doos	box	liefde	love
citroen	lemon	leed	sorrow

Chapter 4

ARE VISUAL PROCESSES CAUSALLY INVOLVED IN "PERCEPTUAL SIMULATION" EFFECTS IN THE SENTENCE-PICTURE VERIFICATION TASK?

Based on: Ostarek, M., Ishag, A., Joosen, D., & Huettig, F. (under revision). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition*.

Abstract

Many studies have shown that sentences implying an object to have a certain shape produce a robust reaction time advantage for shape-matching pictures in the sentence-picture verification task. Typically, this finding has been interpreted as evidence for perceptual simulation, i.e., that access to implicit shape information involves the activation of modality-specific visual processes. It follows from this proposal that disrupting visual processing during sentence comprehension should interfere with perceptual simulation and obliterate the match effect. Here we directly test this hypothesis. Participants listened to sentences while seeing either visual noise that was previously shown to strongly interfere with basic visual processing or a blank screen. Experiments 1 and 2 replicated the match effect but crucially visual noise did not modulate it. When an interference technique was used that targeted high-level semantic processing (Experiment 3) however the match effect vanished. We conclude that the shape match effect in the sentence-picture verification paradigm does not rely on perceptual simulation.

Introduction

In theoretical and empirical efforts to understand conceptual processing during language comprehension recent work has focused on two main problems. The first is concerned with an accurate description of the informational content that is activated as we process language, whereas the second deals with the nature of the neural and cognitive mechanisms that are used to provide this information. Even though both are closely related, it is crucial to address both separately (Barsalou, 1999, 2016; Binder, 2016; Borghesani & Piazza, 2017; Mahon, 2015; Mahon & Caramazza, 2008).

Regarding conceptual content, an overwhelming body of evidence suggests that language processing involves the contextualized retrieval of a multitude of conceptual features that, together, constitute their meanings (Anderson et al., 2016; Binder et al., 2016; Binder & Desai, 2011; Collins & Loftus, 1975; Cree & McRae, 2003; Fernandino, Humphries, Conant, Seidenberg, & Binder, 2016; Fernandino et al., 2016; Huettig & McQueen, 2007; Vigliocco, Meteyard, Andrews, & Kousta, 2009; Vigliocco, Vinson, Lewis, & Garrett, 2004). This view is theoretically appealing because it nicely accounts for the high degree of conceptual flexibility (Barsalou, 1993; Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008b; Ostarek & Huettig, 2017a; van Dam, van Dijk, Bekkering, & Rueschemeyer, 2012; Yee & Thompson-Schill, 2016) by conceiving of conceptual processing as a form of ad hoc sampling from a feature space that is constrained by both long-term memory and immediate context.

Recent behavioral and neuroimaging studies have begun to unravel the underlying mechanisms and started painting a multifaceted picture of a widely distributed system that includes modality-specific processes (Fernandino et al., 2016; Hauk, Johnsrude, & Pulvermüller, 2004; Lewis & Poeppel, 2014; Ostarek & Huettig, 2017a, 2017b; Vukovic, Feurra, Shpektor, Myachykov, & Shtyrov, 2017), different stages of convergence possibly culminating in a modality-independent central hub (Bruffaerts et al., 2013; Fernandino et al., 2016; Patterson, Nestor, & Rogers, 2007; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017), and flexible retrieval mechanisms (Kan & Thompson-Schill, 2004).

The present study focuses on one particular semantic feature; object shape. Visual world eye-tracking studies indicate that processing nouns referring to concrete objects activates information about their typical shape (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007). As many objects can occur in multiple shapes, listeners often need to incorporate contextual information in order to retrieve the appropriate shape representations. Using the sentence-picture verification task, Zwaan, Stanfield, and Yaxley (2002) provided evidence that contextually appropriate shape information is readily activated during sentence comprehension. In that paradigm, participants read or listen to sentences about objects that are implied to have a certain shape (e.g., *The ranger saw the eagle in the sky;* implying outstretched wings). Shortly after sentence offset, in the critical conditions a picture appears of the mentioned object either in matching (e.g., an eagle with outstretched wings) or mismatching shape (an eagle with closed wings). Participants then have to indicate as quickly and accurately as possible whether the object was mentioned in the sentence by pressing one of two buttons. The critical finding (Zwaan & Pecher, 2012; Zwaan et al., 2002) is shorter response latencies in the matching condition, suggesting that the sentences activate information about object shape that is specific enough to produce a priming effect on the verification judgement. Although there has been some debate about the replicability of congruency effects of this type (Papesh, 2015; Rommers, Meyer, & Huettig, 2013; Wagenmakers et al., 2016), the shape match advantage, at least in the sentence-picture verification paradigm, has proven to be very robust and reproducible (Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Rommers, Meyer, & Huettig, 2013; Zwaan & Pecher, 2012).

Previous studies have implicitly or explicitly gone further and suggested that the reaction time advantage in the match condition indicates the kind of process that provides shape information, namely the process of perceptual simulation (Engelen et al., 2011; Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009; Yaxley & Zwaan, 2007; Zwaan & Pecher, 2012; Zwaan et al., 2002). According to that account, accessing conceptual shape information (e.g., about a flying eagle) involves the approximate re-instatement of sensory processes that are active during visual perception of relevant objects (e.g., of a flying eagle).

However, one does not need to invoke simulation in order to explain the behavioral pattern, as studies using the sentence-picture verification paradigm can only tell us something about what kind of information is accessed, but not about the kinds of processes and representations involved. One way to get at the latter question is to study the neural correlates of the shape match effect. Using magnetoencephalography, Hirschfeld, Zwitserlood, and Dobel (2011) found a stronger positivity to pictures following shape matching vs. mismatching sentences in occipital cortex at ca. 120ms after picture onset (M1), suggesting a top-down modulation of early visual processing as a function of shape match vs. mismatch. However, this approach still cannot answer whether visual processes were involved in sentence comprehension, as, similar to RT paradigms, what is measured is the effect of the comprehension process on picture verification that happens only after sentence comprehension is accomplished (Mahon & Caramazza, 2008).

One direct way of testing the hypothesis that visual processes are functionally involved in visual information retrieval is to interfere with visual processing during language comprehension and assess whether visual information retrieval is impaired. Recent studies have demonstrated that dynamic low-level visual noise patterns can selectively interfere with the retrieval of visual information during auditory single word processing (Ostarek & Huettig, 2017a) and in a property verification task (Edmiston & Lupyan, 2017), and they can strongly diminish the effectiveness of a word cue on a subsequent picture discrimination task (Edmiston & Lupyan, 2017). Here, we used the visual noise technique to interfere with visual processing while participants were listening to sentences to directly probe the functional role of perceptual simulation in the sentence-picture verification task.

Experiment 1

The basic rationale for this experiment was that interfering with basic visual processing while participants were listening to sentences should significantly reduce the usually observed shape-match effect if it relies on perceptual simulation. Conversely, if the match effect is independent of visual simulation, visual interference should not have an impact on the match advantage. Experiment 1 used the same kind of visual interference that was recently shown to impair access to visual information during semantic processing (Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017a), consisting of dynamically changing Mondrian-type masks that are usually used for continuous flash suppression and are designed to maximally interfere with basic visual processing (Tsuchiya & Koch, 2005).

Based on the fact that the match effect pertains to visual shape information, whose processing in early visual cortex has been shown to be influenced by processing sentences about matching vs. mismatching object shape (Hirschfeld et al., 2011), and the intuitive proposal that contextually embedded language tends to engage more specific representations and might thus be more likely to activate modality-specific processes (Kurby & Zacks, 2013; Zwaan, 2014), we predicted that visual interference would decrease the match advantage.

Method

Participants

We recruited 115 healthy participants with normal or corrected-to-normal vision and normal hearing from the local MPI subject database. Four had to be excluded due to technical failure, resulting in 111 participants that were used for analysis. We opted for a higher number of participants compared to previous studies using this paradigm based on the fact that our design included the additional factor of Visual Condition (visual noise vs. blank screen) and the conviction that high-powered studies are needed in the field of experimental psychology (Pashler & Wagenmakers, 2012). Participants received a payment of 6 euros. The study was covered by ethics approval from Radboud University Nijmegen.

Materials, Set-up, and Design

We used the materials from the original Zwaan et al. (2002) study that were provided by Rommers et al. (2013). They included 40 quadruplets of pairs of sentences implying shape A or shape B and corresponding pairs of pictures of the mentioned objects in shape A or shape B, and there were 40 filler sentences paired with target pictures that are not mentioned in the sentence. In the original design, every participant saw one of four sentence-picture combinations, resulting in four lists. In the present study, the additional factor of Visual Condition (visual noise vs. blank screen) was added such that every sentence-picture pair was still only shown once to each participant, but across participants every pair occurred equally often in the visual noise and blank screen condition, resulting in eight lists.

Participants were seated 60 centimetres from the screen and placed their head on a chin rest. Presentation (Neurobehavioral Systems) was used to control the display of target pictures and visual noise as well as the sentences that were played back on headphones. Auditory sentences were used instead of written sentences to be able to interfere with visual processing during sentence comprehension. The task was to listen to the sentences and to decide as quickly and accurately as possible by pressing one of two buttons (left/right on a house-built button box, counterbalanced across participants) whether the subsequently displayed picture represented an object that was mentioned in the sentence or not.

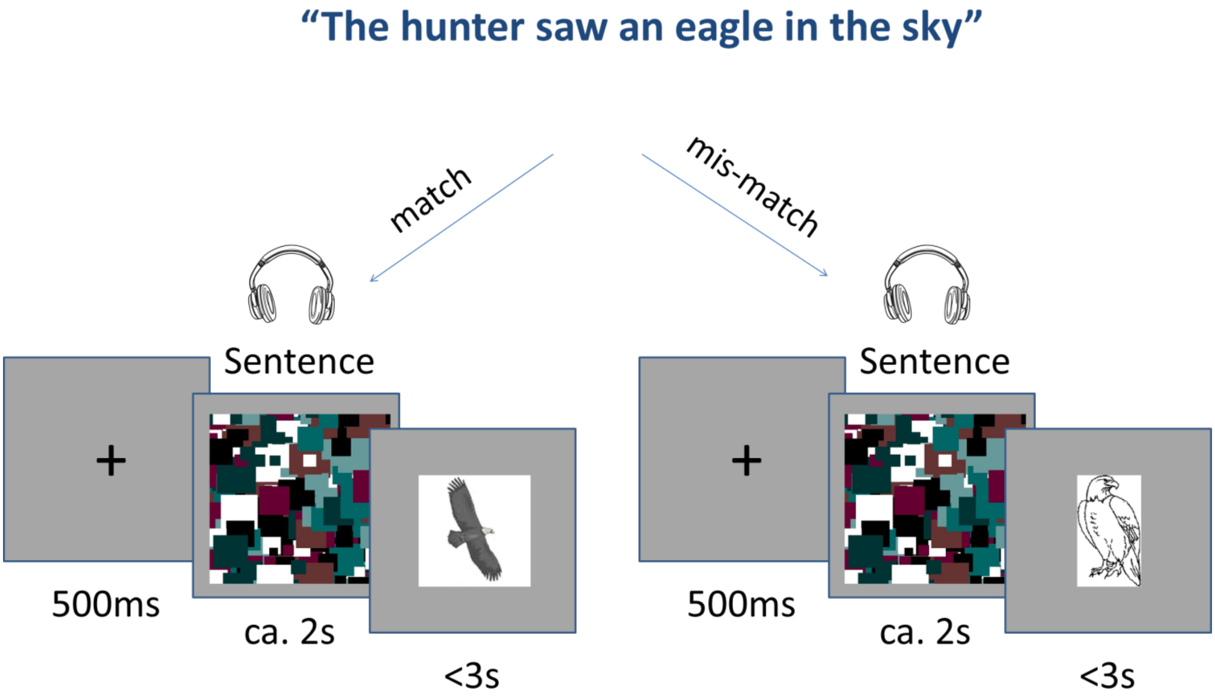


Figure 1: Illustration of the design and trial structure with one of the sentences in the visual noise condition.

Every trial (see Figure 1) started with a fixation cross at the centre of the screen (500ms) followed by an auditory sentence (ca. 2s on average). Sentences were accompanied by visual noise in half of the trials. It consisted of 80 masks that were all generated by randomly superimposing 1000 rectangles of different colors and sizes (similar to Hesselmann, Hebart, & Malach, 2011). For every trial a random order was generated for the 80 masks and they were displayed at a rate of ca. 10 Hz each until 250ms after sentences offset, at which point the target picture was presented. Once a button was pressed or 3 seconds elapsed, the next trial started. After half of the filler trials, a comprehension question appeared on the screen to encourage participants to listen closely to the sentences.

Analysis

Prior to analysis, fillers and trials with incorrect responses or with RTs faster than 300ms or slower than 2500ms were excluded. We then removed trials with RTs 2.5 SDs or higher from the grand condition means. The resulting dataset was analysed using linear mixed effects modelling as implemented in the R package lme4 (Bates, Mächler, Bolker, & Walker, 2014). The full model included Match Condition (match vs. mismatch) and Visual Condition (visual noise vs. blank screen) and their interaction as fixed effects and by-participant and by-sentence random intercepts and slopes for Match Condition, Visual Condition, as well as the interaction term. We used the get_pvalues function from the psy811 package to obtain p-values. Of main interest was whether we would find a reduction of the match effect in the visual noise condition, as reflected in the interaction between Match Condition and Visual Condition. We additionally performed planned follow-up analyses looking at the effect of Match Condition in the visual noise condition and the blank screen condition separately by splitting the dataset in two halves and running a model with Match Condition as fixed effect and random by-participant and by-sentence intercepts and slopes.

Results and Discussion

The results showed a significant main effect of Visual Condition (estimate=0.047, SE=0.014, t=3.447, $p<0.001$) with slower responses in visual noise trials (M=741ms, SD=230ms) compared to blank screen trials (M=721ms, SD=224ms). We also obtained a main effect of Match Condition (estimate=0.074, SE=0.03, t=2.463, p=0.014) with shorter RTs in the shape-matching condition (M=718ms, SD=217ms) compared to the mismatching condition (M=745ms, SD=236ms), thus replicating the match effect. As can be seen in Figure 2, there was, however, no evidence for an interaction between the two factors (t<1). The match effect was present both in blank screen trials (estimate=0.077, SE=0.034, t=2.267, p=0.023) and in visual noise trials (estimate=0.068, SE=0.029, t=2.363 p=0.018).

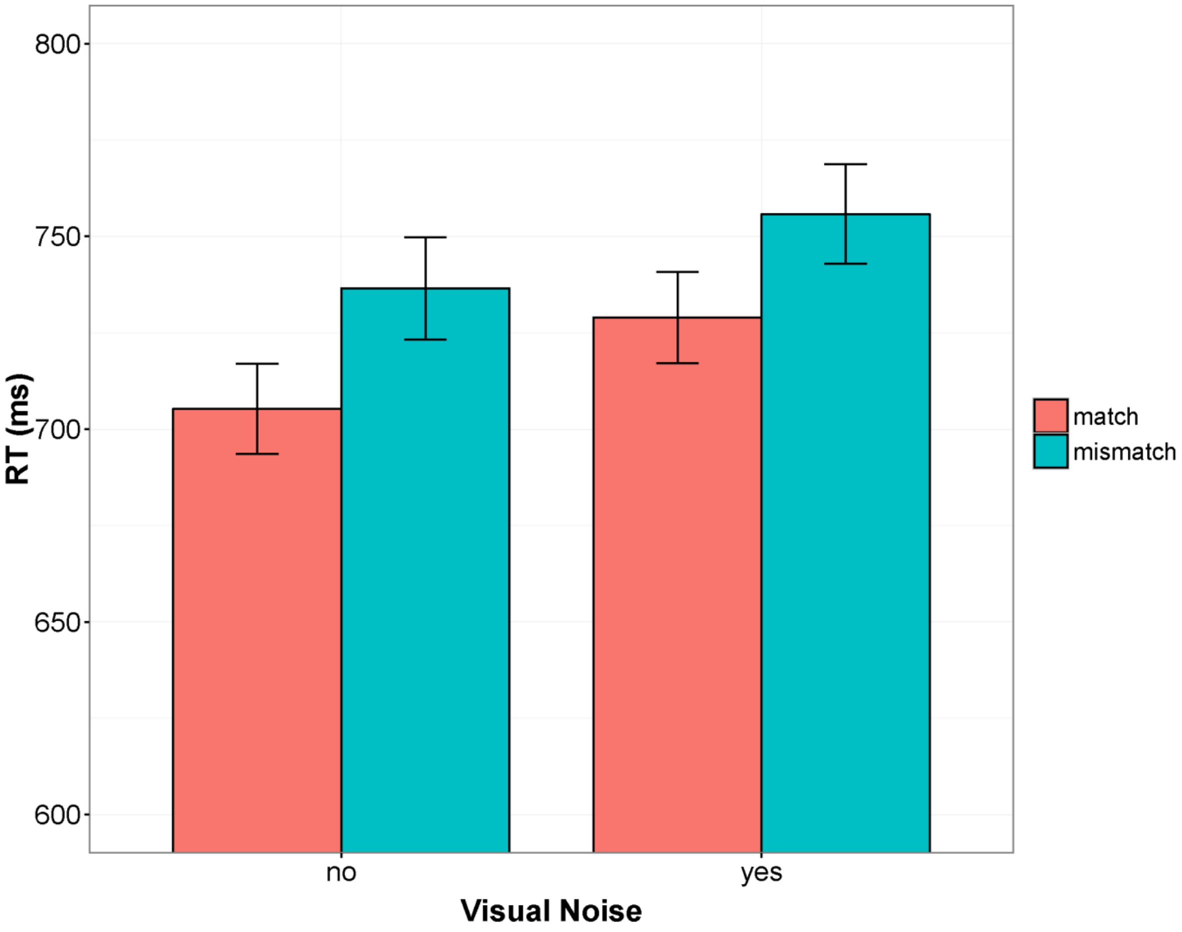


Figure 2: Experiment 1. Mean RTs to shape-matching vs. mismatching pictures (color-coded) with and without visual noise during auditory sentence presentation. Error bars indicate 95% confidence intervals.

Thus, our results suggest that shape information was activated (as reflected by the match effect), but low-level visual processes were not necessary for it (as reflected by the irrelevance of visual interference for the match effect). In the context of two recent studies that reported disrupted access to visual information due to the same type of visual noise in paradigms using single words (Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017a), it seems implausible that the visual noise technique did not sufficiently interfere with basic visual processing. By extension, higher level processes can be assumed to have provided the implicit shape information. Regarding the study by Hirschfeld et al. (2011), this result speaks against the possibility that the occipital M1 modulation they observed reflected low-level visual simulations activated during comprehension. This is consistent with the increased positivity observed by Hirschfeld et al. (2011), as opposed to a decrease that would be expected in a priming-via-re-activation explanation (due to repetition suppression).

Experiment 2

Experiment 1 did not provide evidence for the hypothesis that the shape match effect in the Sentence-Picture Verification task relies on low-level perceptual simulation. It is important to note, however, that we used a type of visual noise that selectively interferes with the most basic computations related to local color, edge, and orientation detection involving only horizontal and vertical components. Thus, it remains possible that visual processes higher up in the hierarchy are causally involved in providing conceptual shape information. To test this possibility, we replaced the low-level visual noise with what we might call *mid-level* visual noise. 80 mid-level noise masks were generated by superimposing 30-40 images of random objects (from De Groot, Koelewijn, Huettig, & Olivers, 2016) and distorting them with the Adobe Photoshop functions "shear", "ripple", and "crystalize" such that they could no longer be recognised (see Figure 3). The resulting images were at least as visually complex as real objects but were unlikely to evoke consistent semantic associations. Our rationale for using these masks as visual noise was that they should interfere with neuron populations with larger receptive fields that are tuned to complex conjunctions of multiple visual features (Peirce, 2015; Vernon, Gouws, Lawrence, Wade, & Morland, 2016). As such, the mid-level visual noise allowed us to test whether mid-level visual processes are involved in the shape match effect.

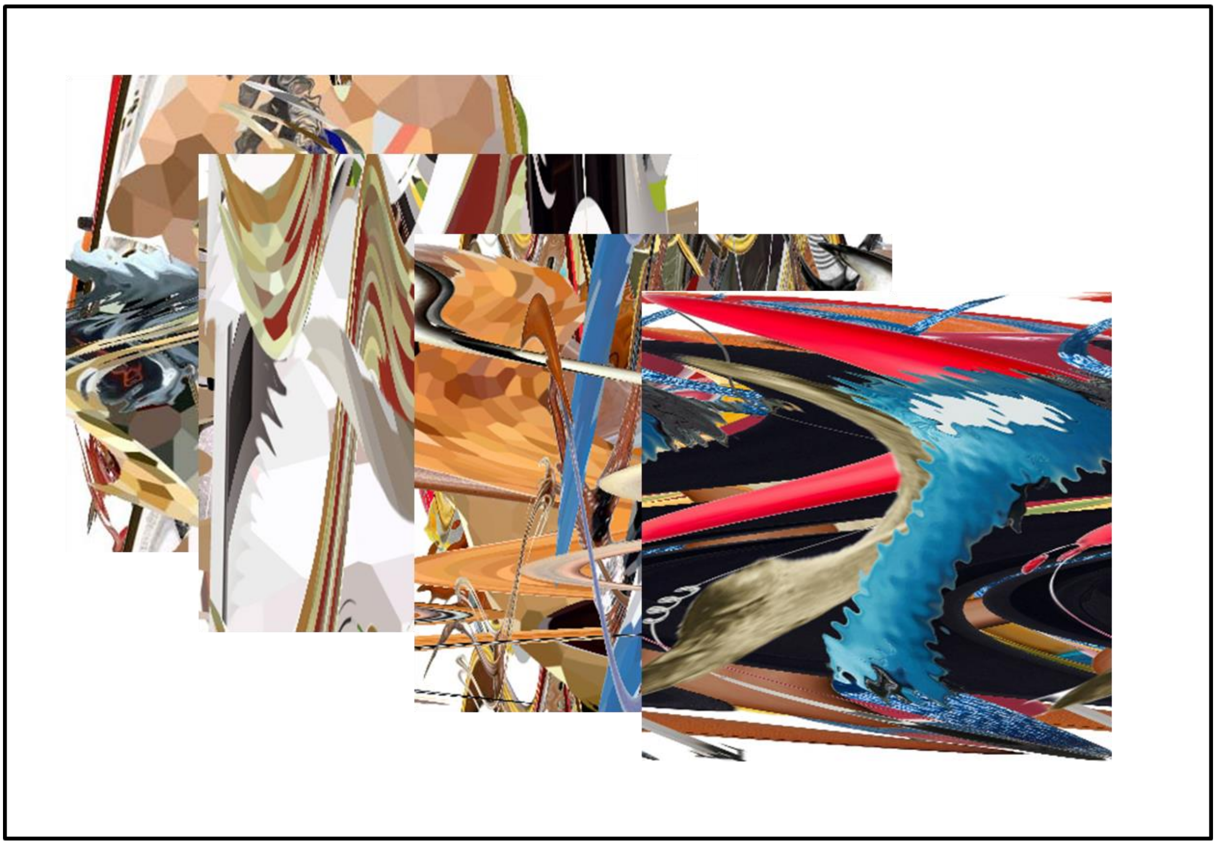


Figure 3: Examples of the mid-level visual noise masks used in Experiment 2.

Method

Participants

We recruited 114 participants from the MPI subject database, one of which was excluded due to an error rate higher than 15%.

Materials, Set-up, Design, and Analysis

Everything was identical to Experiment 1 except for the use of mid-level visual noise that was presented at ca. 10 Hz in visual noise trials.

Results and Discussion

There was again a significant main effect of Visual Condition (estimate=0.065, SE=0.014, t=4.729, p<0.001) with slower RTs in the visual noise condition (M=755ms, SD=233ms) compared to the blank screen condition (M=729ms, SD=229ms), and a main effect of

Match Condition (estimate=0.069, SE=0.033, t=2.116, p=0.034) with shorter RTs in the match (M=728ms, SD=223ms) compared to the mismatch condition (M=756ms, SD=238ms). Again, as Figure 4 shows, there was no evidence for an interaction (t<1). The size of the match effect was similar in the blank screen condition (estimate=0.077, SE=0.033, t=2.307, p=0.021) and in the visual noise condition (estimate=0.06, SE=0.035, t=1.726, p=0.084).

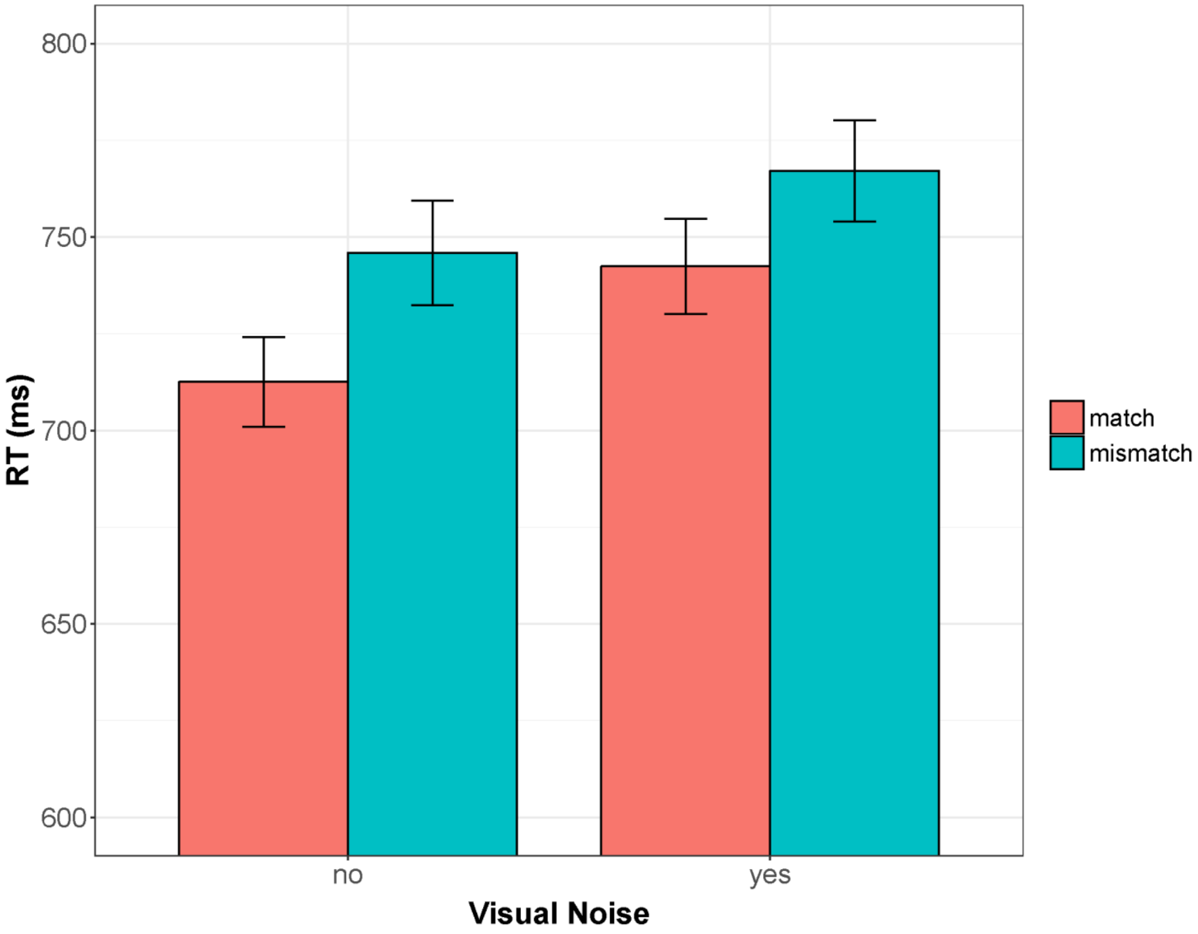


Figure 4: Experiment 2. Mean RTs to shape-matching vs. mismatching pictures (color-coded) with and without visual noise during auditory sentence presentation. Error bars indicate 95% confidence intervals.

Thus, we again replicated the match effect but found no evidence that mid-level visual processes were functionally involved.

Experiment 3

The results of the first two experiments that the shape match effect is largely independent of visual processes are surprising in the context of its prior interpretation (Zwaan, 2003; Zwaan et al., 2002). Given the recent demonstrations of the involvement of low-level visual processes in semantic processing using the same interference technique (Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017a), it is implausible that the visual noise we used was not capable of impeding simulation. Nevertheless, Experiment 3 was designed to ascertain that dynamic visual noise can in principle reduce the match effect in the sentence-picture verification paradigm. To that end, the meaningless visual noise masks were replaced with pictures of intact objects that were again displayed at ca. 10 Hz (henceforth *semantic noise*). The critical difference to the previous two experiments was that pictures activate semantic representations, even when they are presented for very short durations and in rapid succession (Potter, 1976; Potter, Wyble, Hagmann, & McCourt, 2014; Thorpe, Fize, & Marlot, 1996). Recent studies have also shown that visual object recognition involves the rapid activation of object names in adults (McQueen & Huettig, 2014), and even in toddlers (Mani & Plunkett, 2010). As such, semantic noise can be expected to interfere with the access to conceptual shape information implicit in sentences.

Method

Participants

We recruited 111 participants with normal or corrected-to-normal vision and normal hearing from the MPI subject database.

Materials, Set-up, Design, and Analysis

Everything was identical to Experiment 1 and 2 except for the use of object pictures as dynamic visual noise (again ca. 10 Hz). To that end, 80 pictures were randomly selected from the de Groot database with the constraint that they did not represent items mentioned in any of the sentences.

Results and Discussion

Due to a large number of participants with very high mean RTs and large SDs, the outlier removal procedure resulted in very few observations per condition in some participants and the mixed effects model did not converge. We therefore excluded participants who had mean RTs larger than 1000ms and SDs higher than 400ms in at least one condition to reduce noise. The data from the remaining 74 participants are summarised in Figure 5. As in the previous experiments, there was a main effect of Visual Condition (estimate=0.083, SE=0.018, t=4.67, p<0.001) with shorter responses in the blank screen condition (M=718 ms, SD=209 ms) compared to the visual noise condition (M=751 ms, SD=206 ms). There was no significant main effect of Match Condition (estimate=0.046, SE=0.034, t=1.37, p=0.172), but crucially, the interaction was significant (estimate=0.035, SE=0.016, t=2.152, p=0.031), reflecting the match effect (of 31 ms) in the blank screen condition (estimate=0.079, SE=0.036, t=2.165, p=0.03) compared to the absent match effect (5 ms) in the visual noise condition (estimate=0.009, SE=0.035, t=0.271, p>0.7).

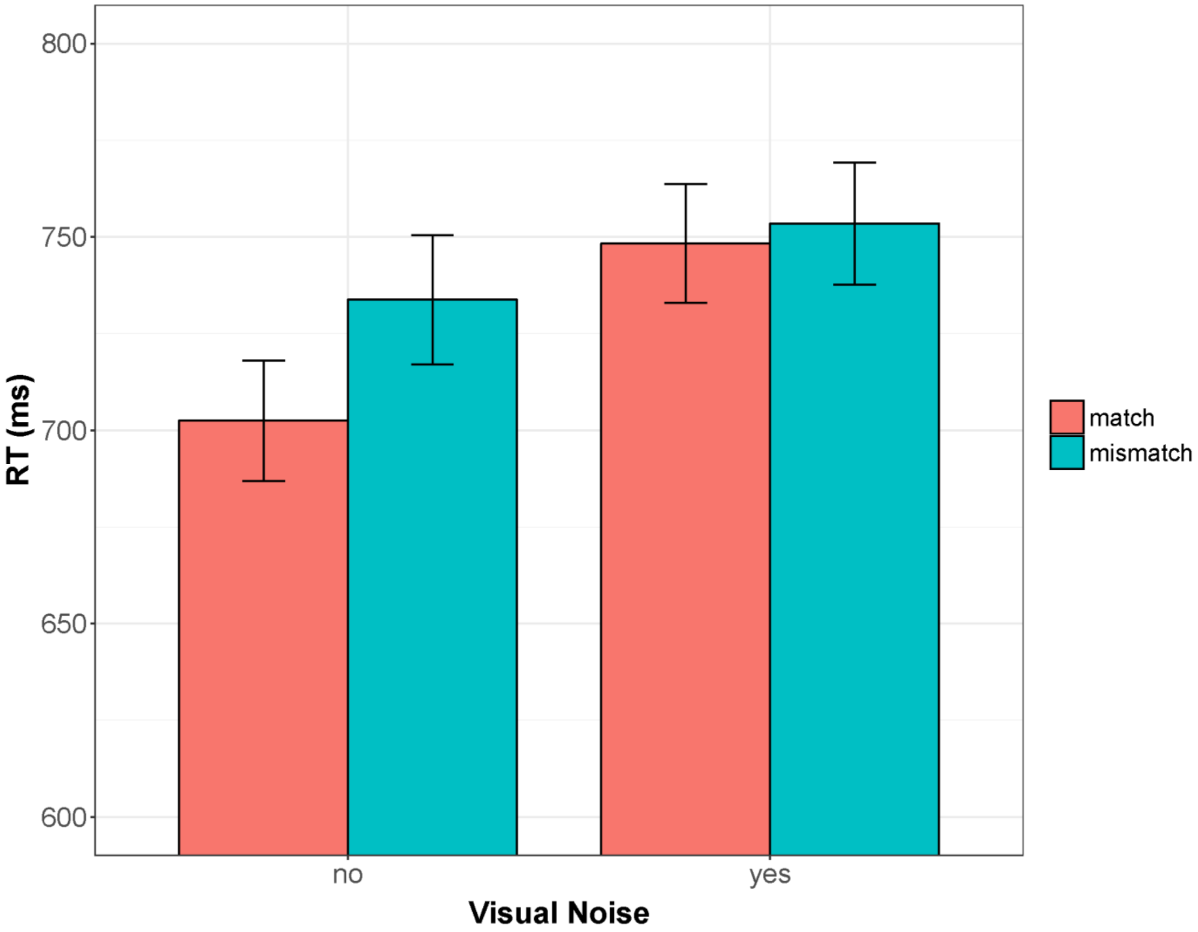


Figure 5: Experiment 3. Mean RTs of to shape-matching vs. mismatching pictures (color-coded) with and without visual noise during auditory sentence presentation. Error bars indicate 95% confidence intervals.

Thus, semantic noise strongly interfered with the access to shape information implicit in sentences, demonstrating that dynamic visual stimuli can be effective at interfering with the retrieval of conceptual shape information if the relevant system is targeted by the noise. This confirms that the absent impact of visual noise on the match effect in Experiments 1 and 2 was unlikely due to an inability of the present interference technique to diminish it, but rather due to its independence of modality-specific visual processes.

General Discussion

The sentence-picture verification task has been of great value for the field of language processing, as it is a versatile tool to reveal the contents of conceptual representations that are activated as listeners/readers comprehend sentences. The key insight was that a match vs. mismatch in a feature of interest between the sentence and the following picture modulates response latencies in the verification task to the extent that the feature was activated during sentence processing. Previous studies have shown that listeners activate object shape information implicit in sentences by demonstrating a reaction time advantage in subsequent sentence-picture verification for target pictures that matched the implied object shape (Hirschfeld et al., 2011; Rommers, Meyer, & Huettig, 2013; Zwaan & Pecher, 2012; Zwaan et al., 2002).

Here, we asked what processes enable the retrieval of shape information during online sentence comprehension. The match effect is typically interpreted as indexing perceptual simulation in sentence comprehension (Engelen et al., 2011; Zwaan, 2003; Zwaan & Pecher, 2012; Zwaan et al., 2002). In particular, the idea is that modality-specific visual processes relevant for shape perception are recruited during language comprehension to provide conceptual shape information. Consequently, when a subsequent shape matching picture appears it is processed more efficiently due to the pre-activation of relevant visual processes. However, while the sentence-picture verification paradigm in its basic form is well-suited to uncover the contents of representations activated during sentence comprehension, it does not allow inferences about how this content is represented. As Hirschfeld et al. (2011) pointed out, besides perceptual simulation the match effect is consistent with a top-down effect based on amodal semantic representations (Mahon, 2015) and with task-based expectations/predictions about the target (Rabagliati, Doumas, & Bemis, 2017; Rommers, Meyer, Praamstra, & Huettig, 2013).

To directly probe the functional role of perceptual simulation for the shape match effect, in two high-powered experiments we employed visual noise to interfere with visual processing during sentence comprehension and observed no evidence for a decrease in the match effect. This was despite the use of a visual noise technique that was developed to maximally interfere with visual processing (Tsuchiya & Koch, 2005) and that has recently been shown to hinder access to visual information in single word processing (Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017a). A third experiment using visually presented semantic noise obliterated the match effect, suggesting that conceptual processes that do not depend on the visual system are driving it. Our results thus suggest that the shape match effect is largely independent of low and mid-level visual processes.

This is a theoretically important finding as it constitutes strong evidence against a simulation-based explanation of the shape match effect. Our findings are compatible with accounts of amodal representation that assume conceptual priming effects to arise in high-level systems with a non-modality-specific representational system (e.g., Fodor, 1975). They are also consistent with the grounding-by-interaction model (Mahon, 2015; Mahon & Caramazza, 2008), which hypothesises amodal representations that are connected to the sensory systems, but only to the extent that sensory states do not measurably affect processing in the amodal system. This model can account for our behavioral results and the MEG data reported by Hirschfeld et al. (2011) by postulating a congruency effect in an amodal conceptual system that optionally interacts with visual processing. One prediction following from this account that could be tested in future studies is that visual interference diminishes the modulation of early visual processing (as observed by Hirschfeld et al. (2011)) without affecting the size of the match effect.

It is important to stress that we are not denying that simulation contributes to language comprehension, given the large body of evidence for this view (Barsalou, 2008; Correia et al., 2014; Fernandino et al., 2016; Hauk et al., 2004; Lewis & Poeppel, 2014; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Ostarek & Huettig, 2017a, 2017b; Pulvermüller, 2005; Vukovic et al., 2017). Other paradigms have recently provided compelling evidence that low-level visual processes (likely related to shape) are engaged in the comprehension of concrete object words (Edmiston & Lupyan, 2017; Lewis & Poeppel, 2014; Ostarek & Huettig, 2017a, 2017b). However, the match effect in the sentence-picture verification task seems to depend on higher-level processes.

This is a striking result not only because the shape match effect is considered a hallmark finding for theories of embodied cognition, but also because the paradigm seems such a good candidate for perceptual simulation (which may partly explain why the match effect has usually been interpreted the way it has). Edmiston and Lupyan (2017) observed a clear effect of visual noise in single word-picture-verification where word cues are followed by a matching or mismatching picture in correct and inverted orientation and participants have to indicate which the correctly oriented picture is. These results in combination with ours suggest that single words activate low-level visual processes (likely reflecting typical object features), whereas implicit shape information derived from event-level representations involves abstraction away from the sensory systems. This is consistent with neuroimaging studies that implicated anterior temporal regions with high-level semantic processing and conceptual combination, whereas visual regions are linked to individual features, such as size, color, and shape (Borghesani et al., 2016; Coutanche & Thompson-Schill, 2014). It will be crucial for future studies to precisely delineate what determines the involvement and role of sensory processes in conceptual processing.

The arguments and method presented in this paper can readily be applied to other paradigms relying on congruency between sensory-motor content evoked by linguistic input and a task involving sensory-motor processing: Congruency effects do not provide evidence that sensory-motor systems produce them unless the task which language is found to have an effect on *only* involves sensory-motor processes (Lupyan & Ward, 2013; Ostarek & Huettig, 2017b). Therefore, match effects are a useful first step after which further investigations are required to reveal the underlying mechanisms. Interference techniques are appealing because they go beyond correlational approaches by directly testing causality.

References

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Aguilar, M., … Raizada, R. D. (2016). Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cerebral Cortex*.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. *Theories of Memory*, *1*, 29–31.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, *23*(4), 1122–1142.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*.

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

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.

Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage*, *143*, 128–140.

Borghesani, V., & Piazza, M. (2017). The neuro-cognitive representations of symbols: the case of concrete words. *Neuropsychologia*.

Bruffaerts, R., Dupont, P., Peeters, R., De Deyne, S., Storms, G., & Vandenberghe, R. (2013). Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. *Journal of Neuroscience*, *33*(47), 18597–18607.

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

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *Journal of Neuroscience*, *34*(1), 332–338.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. *Cerebral Cortex*, *25*(9), 2584–2593.

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*(2), 163.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459.

De Groot, F., Koelewijn, T., Huettig, F., & Olivers, C. N. (2016). A stimulus set of words and pictures matched for visual and semantic similarity. *Journal of Cognitive Psychology*, *28*(1), 1–15.

Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, *92*, 281–292.

Engelen, J. A., Bouwmeester, S., de Bruin, A. B., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*(4), 659–675.

Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., & Binder, J. R. (2016). Heteromodal Cortical Areas Encode Sensory-Motor Features of Word Meaning. *Journal of Neuroscience*, *36*(38), 9763–9769. https://doi.org/10.1523/JNEUROSCI.4095-15.2016

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., … Seidenberg, M. S. (2016). Concept Representation Reflects Multimodal

Abstraction: A Framework for Embodied Semantics. *Cerebral Cortex*, *26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

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

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

Hesselmann, G., Hebart, M., & Malach, R. (2011). Differential BOLD activity associated with subjective and objective reports during "blindsight" in normal observers. *Journal of Neuroscience*, *31*(36), 12936–12944.

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing–MEG dissociates early perceptual and late N400 effects. *Brain and Language*, *116*(2), 91–96.

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

Huettig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985–1018.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Kan, I. P., & Thompson-Schill, S. L. (2004). Selection from perceptual and conceptual representations. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(4), 466–482.

Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, *126*(3), 338–349.

Lambon Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Lewis, G., & Poeppel, D. (2014). The role of visual representations during the lexical access of spoken words. *Brain and Language*, *134*, 1–10.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 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*(1), 59–70.

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science*, *21*(7), 908–913.

McQueen, J. M., & Huettig, F. (2014). Interference of spoken word recognition through phonological priming from visual objects and printed words. *Attention, Perception, & Psychophysics*, *76*(1), 190–200.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804.

Ostarek, M., & Huettig, F. (2017a). A Task-Dependent Causal Role for Low-Level Visual Processes in Spoken Word Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1215. https://doi.org/http://dx.doi.org/10.1037/xlm0000375

Ostarek, M., & Huettig, F. (2017b). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 499.

Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General*, *144*(6), e116.

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530.

Patterson, K., Nestor, P. J., & 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. https://doi.org/10.1038/nrn2277

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology*, 62(6), 1108–1114.

Peirce, J. W. (2015). Understanding mid-level representations in visual processing. *Journal of Vision*, *15*(7), 5–5.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, *76*(2), 270–279.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.

Rabagliati, H., Doumas, L. A., & Bemis, D. K. (2017). Representing composed meanings through temporal binding. *Cognition*, *162*, 61–72.

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, *24*(11), 2218–2225.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096.

van Dam, W. O., van Dijk, M., Bekkering, H., & Rueschemeyer, S.-A. (2012). Flexibility in embodied lexical-semantic representations. *Human Brain Mapping*, *33*(10), 2322–2333.

Vernon, R. J., Gouws, A. D., Lawrence, S. J., Wade, A. R., & Morland, A. B. (2016). Multivariate Patterns in the Human Object-Processing Pathway Reveal a Shift from Retinotopic to Shape Curvature Representations in Lateral Occipital Areas, LO-1 and LO-2. *Journal of Neuroscience*, *36*(21), 5763–5774.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247.

Vigliocco, G., Vinson, D. P., Lewis, W., & 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.

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, 96, 222–229.

Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams Jr, R. B., … others. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, *11*(6), 917–928.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*(1), 229–236.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027.

Zwaan, R. A. (2003). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, *44*, 35–62.

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PloS One*, *7*(12), e51382.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Chapter 5

SACCADE TRAJECTORIES REVEAL DYNAMIC INTERACTIONS OF SEMANTIC AND SPATIAL INFORMATION DURING THE PROCESSING OF IMPLICITLY SPATIAL WORDS

Based on: Ostarek, M., Ishag, A., Joosen, D., & Huettig, F. (in press). Saccade trajectories reveal dynamic interactions of semantic and spatial information during the processing of implicitly spatial words. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Abstract

Implicit up/down words, such as *bird* and *foot*, systematically influence performance on visual tasks involving immediately following targets in compatible vs. incompatible locations. Recent studies have observed that the semantic relation between prime words and target pictures can strongly influence the size and even the direction of the effect: Semantically related targets are processed faster in congruent than in incongruent locations (location-specific priming), whereas unrelated targets are processed slower in congruent locations. Here, we used eye-tracking to investigate the moment-to-moment processes underlying this pattern. Our reaction time results for related targets replicated the location-specific priming effect and showed a trend towards interference for unrelated targets. We then used growth curve analysis to test how up/down words and their match vs. mismatch with immediately following targets in terms of semantics and vertical location influences concurrent saccadic eye movements. There was a strong main effect of spatial association on linear growth with up words biasing changes in y-coordinates over time upwards relative to down words (and vice versa). Similar to the RT data, this effect was strongest for semantically related targets and reversed for unrelated targets. Intriguingly, all conditions showed a bias in the congruent direction in the initial stage of the saccade. Then, at around halfway into the saccade the effect kept increasing in the semantically related condition, and reversed in the unrelated condition. These results suggest that online processing of up/down words triggers direction-specific oculomotor processes that are dynamically modulated by the semantic relation between prime words and targets.

Introduction

Language comprehension involves the rapid and efficient retrieval of relevant multimodal information about word referents (Binder & Desai, 2011; Fernandino, Humphries, Conant, Seidenberg, & Binder, 2016; Fernandino et al., 2016b). A question of intense debate in recent years has been whether and to what extent modality-specific sensory-motor systems are used as a source of information during conceptual processing (Barsalou, 2016; Binder, 2016; Binder & Desai, 2011; Mahon, 2015; Mahon & Caramazza, 2008; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012). Regarding concrete concepts, whose referents possess a reliable set of perceptual features, there is evidence suggesting that information about object shape (Lewis & Poeppel, 2014; Ostarek & Huettig, 2017a, 2017b), motion (Meteyard, Bahrami, & Vigliocco, 2007; Meteyard, Zokaei, Bahrami, & Vigliocco, 2008; van Dam, Speed, Lai, Vigliocco, & Desai, 2017), color (Simmons et al., 2007), and associated actions (Beauchamp & Martin, 2007; Chao, Haxby, & Martin, 1999; Hauk et al., 2004; Martin, 2007; Pulvermüller, 2005; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014; Vukovic et al., 2017) is retrieved from the corresponding modal systems. Recent studies have indicated that access to sensory processes during conceptual processing is highly task-dependent (Hoenig et al., 2008b; Kemmerer, 2015; Lebois, Wilson-Mendenhall, & Barsalou, 2015; Ostarek & Huettig, 2017a; Rommers, Meyer, & Huettig, 2013; van Dam et al., 2012; Yee & Thompson-Schill, 2016), a property that is likely owed to the highly context-dependent mapping between words and conceptual representations (Barsalou, 1983, 1993; Thomson & Tulving, 1970) and to a division of labor between low-level and high-level systems (Barsalou, 2016; Binder, 2016; Borghesani et al., 2016; Borghesani & Piazza, 2017; Chen, Lambon Ralph, & Rogers, 2017; Fernandino et al., 2016a; Fernandino et al., 2016b; Pobric, Lambon Ralph, & Jefferies, 2009; Lambon Ralph, 2014; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017; Rogers et al., 2004).

The rigorous examination of the mechanisms involved in promising experimental paradigms tapping conceptual representation proved to be critical to allow the field to ask more and more detailed questions, but it can also dampen initial excitement when the story is more complicated than a first look suggested. A point in case is the study of verbs implying upward or downward motion (such as *rise* and *fall*). Initially, Meteyard et al. (2007) showed that such verbs induce increased sensitivity at detecting near-threshold motion of dot patterns in congruent direction. A subsequent study went a step further by showing that, conversely, near-threshold motion patterns interfered with lexical decisions on incongruent verbs, suggesting a causal role of low-level visual processes in verb comprehension (Meteyard et al., 2008). However, recent functional magnetic resonance imaging (fMRI) studies localized the direction-specific congruency effect of verbs on motion perception to the left middle temporal gyrus implicated with high-level conceptual processing and found no evidence that these verbs activated motion-sensitive visual areas (Francken, Kok, Hagoort, & De Lange, 2014; Francken, Meijs, Hagoort, Van Gaal, & De Lange, 2015). Moreover, the congruency effect persisted when prime words were presented subliminally, suggesting that it does not rely on feedback to visual areas (Francken, Meijs, Ridderinkhof, et al., 2015). At the same time, studies looking at sentences (Saygin, McCullough, Alac, & Emmorey, 2010) and narratives (Wallentin et al., 2011) about motion did find increased activation levels in the motion-selective region MT/V5, suggesting that richer contexts may be more likely to recruit featural information from sensory cortex. Thus, more research is required to elucidate when motion language recruits motion-sensitive visual processes.

Research in a related paradigm investigating the effects of words/sentences with up vs. down associations on visual categorization and detection has gone through a similar refinement process. Bergen, Lindsay, Matlock, and Narayanan (2007) first reported that up/down sentences (such as *the ground/roof shook*) interfered with the subsequent identification of a circle vs. a square in congruent location. Further studies extended the interference effect to single nouns (Estes, Verges, & Barsalou, 2008; Verges & Duffy, 2009) and verbs (Verges & Duffy, 2009). These findings were interpreted as evidence that location-specific visual simulations of the denoted referents interfered with concurrent perceptual processing of the targets, as proposed by Barsalou and colleagues (Barsalou, 1999, 2008; Barsalou, Simmons, Barbey, & Wilson, 2003).

Several subsequent studies observed facilitation instead of interference for targets in compatible location using very similar paradigms (Dudschig, Lachmair, de la Vega, De Filippis, & Kaup, 2012; Gozli, Chasteen, & Pratt, 2013; Gozli, Pratt, Martin, & Chasteen, 2016; Zhang et al., 2013). In a series of experiments, Gozli et al. (2013) singled out the factors driving the direction of the effect and concluded that interference is likely to be observed at short stimulus onset asynchronies (SOAs) between cue word and target (< 400 ms) and when discrimination (as opposed to detection) tasks are used, whereas facilitation is typically obtained at longer SOAs or when detection tasks are used. Finally, recent studies showed that the usually observed interference in compatible location at short SOAs can be turned into facilitation when the target is semantically related to the prime word (Estes, Verges, & Adelman, 2015; Ostarek & Vigliocco, 2017).

One account that has been brought forth to explain this set of findings rests on the notion of situated conceptualization. According to that framework, even single words evoke rich knowledge of events in which their referents are typically perceived (Barsalou, 2003; Barsalou, 2009). The role of SOA can be captured by a basic simulation explanation, as interference is only expected during semantic processing when visual simulation is expected to hinder simultaneous visual processing of the target, whereas residual activation of spatial representations can account for late facilitation. Critically, the account can explain the location-specific facilitation effect for semantically related targets by assuming that related targets benefit from the prime because they can be integrated in the activated (location-specific) event, whereas unrelated targets cannot (Ostarek & Vigliocco, 2017).

Two accounts have recently been proposed to complement the situated conceptualization explanation. The perceptual matching account proposes that participants engage in a form of automatic visual search behavior and implicitly code targets for congruence vs. incongruence in terms of their identity and location, and that conflicting codes produce interference (Estes et al., 2015). It is based on evidence from eye-tracking studies that words trigger a visual search for a corresponding (or related) target (see Huettig, Olivers, & Hartsuiker, 2011, for discussion), which has been suggested to be automatic because it even occurs with word cues which are irrelevant for the task (Spivey, Tyler, Eberhard, & Tanenhaus, 2001). The rationale is that inconsistent codes constitute ambiguous evidence for whether the target is the word referent or not: An unrelated target appearing in a compatible location provides evidence that it may be the word referent, which then has to be ruled out before the identity of the actual target is established. Conversely, when an unrelated target appears in an incompatible location, there is no evidence that it is the referent to begin with and its identity can be processed immediately. Finally, when a related target appears in a compatible location, a match in terms of both location and object constitutes useful evidence that the target is the referent. Apart from being quite post-hoc, the perceptual matching account does not predict the finding that semantically related, but perceptually dissimilar, targets elicit facilitation at short SOAs (Ostarek & Vigliocco, 2017). Moreover, it does not speak to the mechanisms by which location and object codes are generated. In fact, it is conceivable that matching of location and identity codes is done by comparing location and object features of the visual targets with the content of simulated events, in which case perceptual matching can be considered as a paradigm-specific consequence of situated conceptualization.

Two recent studies (Amer, Gozli, & Pratt, 2017; Gozli et al., 2013) suggested an alternative account based on the theory of event coding (TEC) whose central tenet is that perceptual and action-related features belonging to the same event are temporarily bound into a short-lived event file (Amer et al., 2017; Hommel, 1998, 2004; Rothermund, Wentura, & De Houwer, 2005). Assuming that conceptual processing involves sensorimotor systems, the TEC can indeed be successfully applied to the spatial cuing paradigm: It correctly predicts interference at short SOAs when a stimulus feature that is hypothesized to be bound (e.g., the semantic feature "up" of a word) is required for the visual task (e.g., when the target appears at the top of the screen), and it correctly predicts facilitation at long SOAs (due to pre-activation) when features are expected to no longer be bound. Similarly, the finding that semantically related words shorten RTs to targets in compatible location only (Estes et al., 2015; Ostarek & Vigliocco, 2017) can be ascribed to the ability to integrate the picture features with the event file activated by the word.

In the present study, we used eye-tracking to investigate how the cognitive mechanisms underlying the location-specific priming vs. interference effects unfold from the moment the prime word is presented until the target picture is fixated. Previous studies used RT measures that are blind to the online cognitive processes that unfold as congruence vs. incongruence on multiple levels interacts and only give one piece of discrete information about their final outcome. Eye-tracking, by contrast, can reveal the dynamically emerging processes with the highest temporal resolution.

Recent eye-tracking studies found that processing up/down words can shorten launch times for saccades in compatible direction (Dudschig, Souman, Lachmair, de la Vega, & Kaup, 2013; Dunn, 2016). However, these studies used a paradigm in which participants performed a lexical decision task by moving their eyes up vs. down, which makes the link of the results to semantic processing unclear because, by the very nature of the task, semantic processing was likely completed by the time participants started moving their eyes. This is because in order to make a lexical decision one must have first processed the word. The crucial test would be a situation in which semantic processing of up/down words occurs simultaneously with eye-movements. In the present study, we achieved this by minimizing the delay between word and visual target allowing us to study the direct effect of semantic processing on concurrent eye movements. We focused on saccade trajectories instead of launch times to investigate how the effects of implicitly spatial words unfold throughout the saccade and how they interact with spatial and semantic properties of the targets across time.

The present study

In the present study, we used a design similar to a recent study (Ostarek & Vigliocco, 2017) that looked into the role of the semantic relation between cue words and target pictures on the spatial congruency effect. The design is different from the one usually used in spatial cuing paradigms in that the targets are not meaningless symbols (such as X vs. O), but drawings of objects that are either semantically related or unrelated to the prime word. The motivation for choosing this paradigm was three-fold: 1) Both the situated conceptualization and TEC account assume that sensory processes are activated to provide semantic information about spatial location, however to date there is no strong evidence for that claim. Previous studies observed shortened saccade launch times right after up/down words were processed (Dudschig et al., 2013; Dunn, 2016). We go one step further by testing whether up/down word processing biases *concurrent* saccade trajectories up/down, as this would indicate that semantic processing of up/down words involves direction-sensitive representations in the oculomotor network. 2) On the behavioral level, we wanted to see whether we could replicate the finding that semantic relatedness of cue and target can turn the usually observed interference effect at short SOAs into facilitation (Ostarek & Vigliocco, 2017). This pattern is predicted by situated conceptualization accounts because processing of semantically related targets in compatible locations is expected to benefit from the situated simulation triggered by the prime word. TEC would also be compatible with this outcome if one assumes that prime-target pairs with semantic feature overlap can be integrated in a single event file whereas prime-target pairs with non-overlapping features cannot. 3) Most importantly, we wanted to investigate to what extent this pattern is reflected in saccade trajectories, i.e. to observe potential facilitation or interference effects emerge and unfold on a moment-to-moment basis. One main hypothesis was that saccade trajectories would be biased towards the direction associated with the prime word at early stages of the saccade, in line with simulation-based accounts of situated conceptualization. Furthermore, TEC predicts additional modulations of saccade trajectories depending on the semantic relation of the prime word and target. In particular, saccades towards spatially and semantically congruent targets should be facilitated, whereas saccades towards targets that are only congruent on one domain (spatial or semantic) should be hindered. This is because spatial and semantic features activated by the prime words are assumed to be temporarily bound in an event file. Therefore, when one of the features is required for processing of the target it first has to be separated from the event file evoked by the word and then incorporated into a new event file (Hommel, 1998, 2004).

Method

Participants

We tested 57 healthy participants (35 female, 10 left-handed, mean age: 26.8) from the local MPI subject database. Two participants had to be excluded because their vision was impaired, one because of a technical error, and two because calibration failed. All analyses were performed on the data from the remaining 52 participants who all had normal or corrected-to-normal vision (contact lenses were allowed). They gave written consent and were paid six euros for their participation. We had ethics approval for the study from the faculty for Social Sciences of Radboud University Nijmegen.

Set-up, Materials, and Procedure

We recorded eye-movements with a tower-mount Eye-Link 1000 eye-tracker (SR Research) with 1000 Hz temporal resolution. The experiment was programmed and run in Experiment Builder (SR Research). The session began with a standard nine-point calibration procedure. Participants placed their head on a chin rest 70 cm from the screen (1920x1080, 60Hz) and were asked to keep their head still during the experiment.

There were a total of 224 trials (see Figure 1) broken down into eight blocks such that participants had a chance to take a brief break every 28 trials. At the beginning of every block, a fixation dot appeared at the center of the screen that served as drift correction. Once the dot was fixated by the participant, the researcher confirmed correct fixation by pressing enter which resets the current fixation location as center of the screen. Each trial began with a central fixation cross displayed for 500ms, followed by a central prime word (100ms), a blank screen (50ms), and finally the target. This quick succession of events made it possible to study the effect of online word processing on saccade trajectories. In particular, the visual target appeared just 150ms after word onset (and 50ms after word offset) and was expected to rapidly trigger a saccade that should land on the target some hundreds of milliseconds later. Thus, the saccade was expected to occur in the time window typically associated with semantic processing.

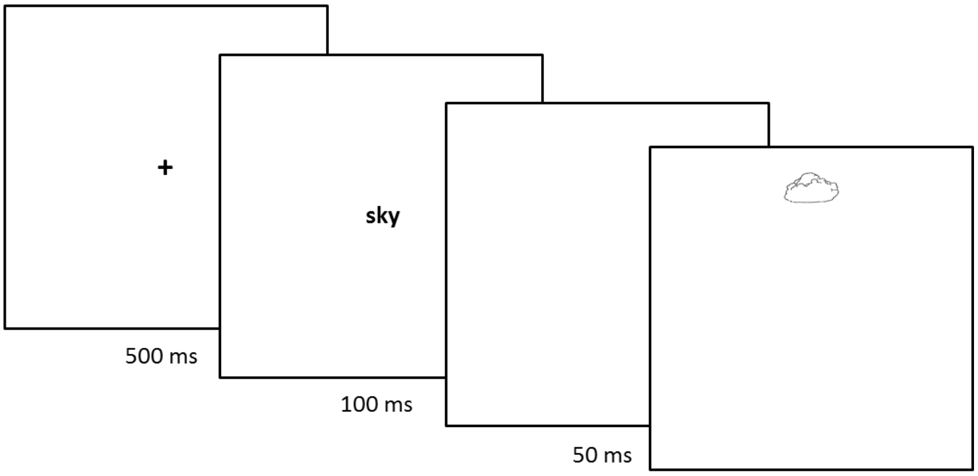


Figure 1: Trial structure illustrated with an up word and a semantically related target in congruent location.

The targets were line drawings of objects (n=28) or geometric shapes (n=28) and were presented four times each; twice at the top of the screen and twice at the bottom. Every line drawing was paired with one semantically related up/down word, one unrelated up word, one unrelated down word, and one spatially neutral control word, which were all presented once[6]. The geometric shapes were paired with the same set of words. We counterbalanced target location (top vs. bottom) between subjects such that every word was followed by the corresponding targets at the top vs. bottom of the screen equally often.

22 separate subjects that did not participate in the main experiment rated all words on a 7-point Likert scale ranging from 1 (object is always seen low in the visual field) to 7 (object is always seen high in the visual field). Non-parametric Wilcoxon signed rank tests showed that up words (median=5.25, SD=0.66) were rated to have a significantly stronger up association (estimate=3.50, 95% CIs: 3.12, 3.88, $p<0.001$) compared to down words (median=2, SD=0.54), and both differed significantly from the neutral control words (median=4, SD=0.50; up vs. control: estimate=1.87, 95% CIs: 1.50, 2.13, $p<0.001$; down vs. control: estimate=1.75, 95% CIs: 1.38, 2.00, $p<0.001$). The different prime types (semantically related up words, semantically related down words, unrelated up words, unrelated down words, and spatially neutral control words) were matched (all pairwise comparisons; $p>0.05$) for frequency (using log word frequency in the SUBTLEX databse; Keuleers, Brysbaert, & New, 2010), age of acquisition (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014), number of letters, and number of syllables (see Appendix). Latent semantic similarity analysis (Landauer, Foltz, & Laham, 1998) was used to quantify semantic similarity between the prime words and target pictures across conditions and showed that the semantically related condition had significantly higher cosines than the control condition ($p<0.001$) and the unrelated condition ($p<0.001$), whereas there was no significant difference between the control and unrelated condition ($p>0.2$). Finally, up-down ratings were not different in the related vs. unrelated condition ($p>0.05$) and thus effects of prime type cannot be attributed to confounds in spatial association.

Participants had to decide as quickly and as accurately as possible whether they saw an everyday object or a geometric shape by pressing the left or right button of a button-box

---

[6] with the exception of one unrelated up word and one unrelated down word that were presented twice

(counterbalanced across participants). As soon as participants responded, the trial ended with a 500ms blank screen after which the next fixation cross appeared. The everyday object vs. geometric shape task was chosen for two reasons: 1) Previous data indicate that congruency effects in the spatial cueing paradigm may depend on target features, especially the extent to which semantic features are shared between prime and target. At short SOAs, up/down words tend to have an interference effect on the identification of simple meaningless targets (such as single letters) in compatible location (Estes et al., 2015, 2008; Gozli et al., 2013), whereas a facilitation effect was observed on semantically related pictures (Estes et al., 2008; Ostarek & Vigliocco, 2017), and no location-specific effect was observed on unrelated pictures (Ostarek & Vigliocco, 2017). Our design allowed us to test the role of the type of prime-target relation on the spatial congruency effect within a single experiment. 2) As we used a variety of geometric shapes, participants were required to look at the targets in order to accurately perform the task.

Analysis

RT data

RT data were trimmed by removing erroneous trials, responses faster than 300ms or slower than 2.5s (which are unlikely to be voluntary responses related to the task), and responses that were more than 2.5 standard deviations from the grand mean per condition. RTs were mean-centered and divided by the standard deviation.

We assessed the effects of Prime Target Relation (semantically related, unrelated, geometric shape) and Spatial Congruence (congruent vs. incongruent) as well as their interaction using a linear mixed effects model (lme4 package) with per-participants and per-target item intercepts as well as random by-participants and by-target item slopes for Prime Target Relation and Spatial Congruence. A model including random slopes for the interaction term did not converge. Likelihood ratio tests were used to test for significant main effects and interactions by comparing a model including the factor/interaction of interest with one that did not include it but was otherwise identical. To follow up significant interaction effects, we used lsmeans for pairwise comparisons of the effect of Spatial Congruence at all Prime-Target Relation levels (lsmeans package; Lenth, 2016).

Eye-tracking data

With respect to the eye-tracking data, we hypothesized that up/down words differentially influence saccade trajectories and that this effect should interact with the relation between prime word and target picture. To test this hypothesis we used growth curve analysis (GCA; Mirman, Dixon, & Magnuson, 2008) implemented in the psy811 package to analyze changes in y-coordinates over time. In particular, the dataset was organized such that for every participant there was one data point (y-coordinate in pixels) per condition per ms to which a first-degree polynomial (linear term) was fitted, which was then submitted to a linear mixed effects model including Spatial Association of Prime (up vs. down), Prime-Target Relation (semantically related vs. unrelated vs. geometric shape), the linear term, and all interaction terms as fixed effect and per-participants random intercepts (more complex models with random slopes did not converge).

We analyzed the data from the moment the prime word appeared (500ms into the trial) until the saccade landed on the target (1250ms)[7]. There were two key predictions: 1) An effect of Spatial Association of Prime on the linear term. This would indicate that up vs. down words have differential effects on the slopes of saccade trajectories, meaning that differences in y-coordinates due to the words' spatial associations grow stronger over time (see Mirman et al., 2008). We expected that up words would bias the slope of the trajectory upwards relative to down words (and vice versa). As the slopes describe the rate of change in y-coordinates over time, they directly reflect the speed of saccades. Effects on the linear term can therefore be thought of as consistent effects on saccade velocities across time. 2) An interaction effect of Spatial Association of Prime and Prime-Target Relation on the linear term such that words would speed up saccades towards the associated location when the target is semantically related and slow down saccades towards the associated location when the target is semantically unrelated.

---

[7] This number was derived from visual inspection of Figures 3 and 4 and reflects the point of inflection where on average the trajectories stopped rising or falling.

Results

RT data

The linear mixed effects model on the RT data (see Figure 2) showed no main effect of Spatial Congruence and a small effect of Prime-Target Relation with slightly faster responses to geometric shapes compared to the unrelated condition (estimate=-0.071, SE=0.035, t=-2.012, p=0.044). Importantly, there was an interaction of Spatial Congruence and Prime-Target Relation (Chisq(2)=9.873, p=.007) that was characterized by the pattern that in the semantically related condition RTs were significantly shorter for spatially congruent vs. incongruent targets (lsmeans estimate=-.130, SE=.052, t-ratio=-2.491, p=.015) whereas in the unrelated and geometric shape conditions RTs were slower for spatially congruent targets. However, this effect was only marginally significant for unrelated pictures (estimate=.059, SE=.035, t-ratio=1.701, p=.091), and non-significant for geometric shapes (estimate=.043, SE=.029, t-ratio=1.482, p=.142).

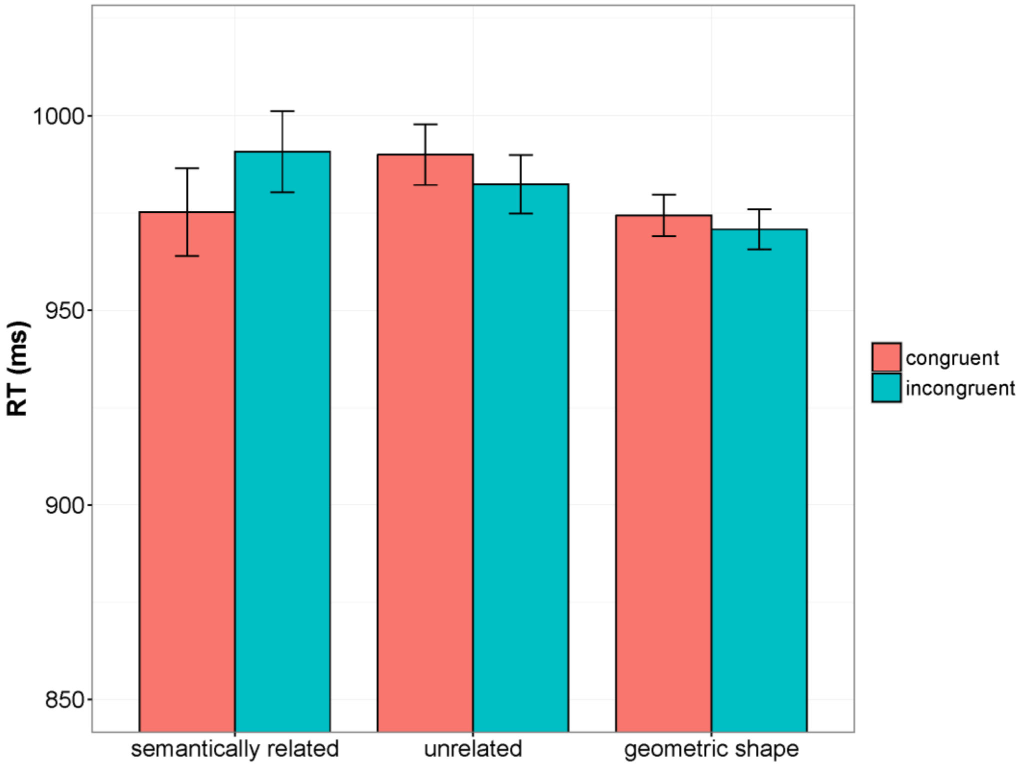


Figure 2: Mean reaction times in all conditions; the x-axis indicates the Prime-Target Relation conditions, Spatial Congruence is color-coded. Error bars indicate 95% confidence intervals.

We thus replicated the location-specific priming effect (Ostarek & Vigliocco, 2017); RTs to trials of the type ["sky" --> picture of a cloud at the top] are fastest, whereas RTs to ["sky" --> picture of a cloud at the bottom] are slowest despite the semantic relationship between prime and target that typically yields facilitation when the spatial location is not manipulated. Furthermore, as in a previous study (Ostarek & Vigliocco, 2017), we only obtained a non-significant trend towards interference for unrelated targets in congruent location. One possible explanation is that studies that found the interference effect used very simple meaningless targets (Estes et al., 2015, 2008; Gozli et al., 2013) and that the RT data reflect a mixture of facilitation due to spatial congruence and interference due to semantic incongruence, a point to which we will return in the Discussion.

Eye-tracking data

In terms of the saccade trajectories, the main question was whether up/down word processing would have a direction-specific impact on simultaneous saccades and how this effect would be modulated by the semantic relation with the target. Figures 3 and 4 show the effects of Spatial Association of Prime and Prime-Target Relation on upward and downward saccades respectively. The lines were corrected for small baseline differences by subtracting the mean y-coordinates in the 500ms time window before the prime word appeared per condition. The vertical dotted lines represent prime word onset (at 500ms into the trial) and target picture onset (at 650ms). The plots suggest a large congruency effect in the semantically related condition, a smaller effect in the geometric shape condition, and no visible effect in the unrelated condition. Note that small differences may not be detectable in these plots as they show the entire eye-movement trajectories from the center to the top/bottom of the screen.

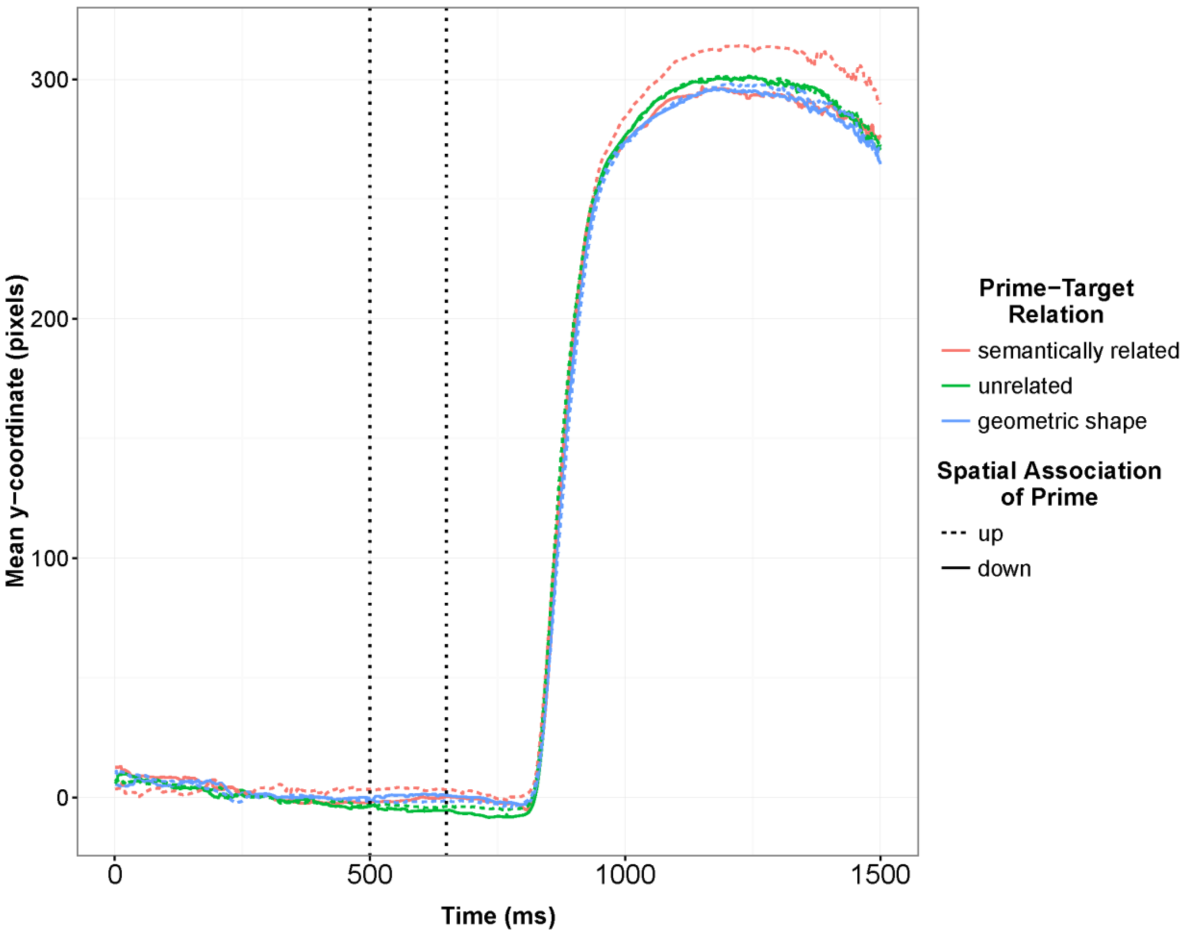


Figure 3: Mean y-coordinates in pixels for up vs. down words at all levels of Prime-Target Relation for all trials with targets at the top of the screen. The vertical dotted lines indicate the appearance of the prime word and the target, respectively.

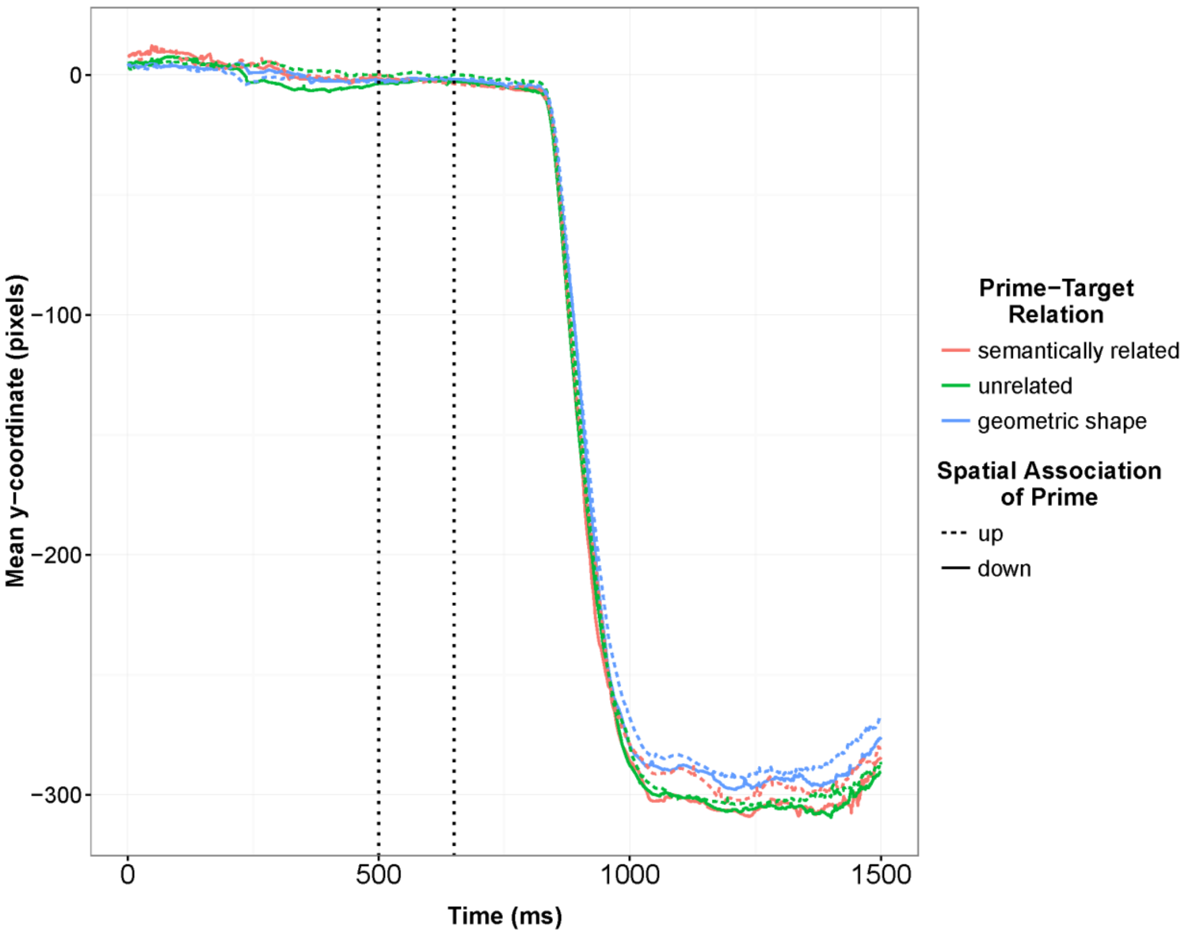


Figure 4: Mean y-coordinates in pixels for up vs. down words at all levels of Prime-Target Relation for all trials with targets at the bottom of the screen. The vertical dotted lines indicate the appearance of the prime word and the target, respectively.

To better visualize smaller effects over time, Figure 5 collapses across upwards and downwards saccades and shows the spatial congruency effect (i.e. the extent to which up words induce an upward bias in y-coordinates relative to down words (or vice versa)) in the three Prime-Target Relation conditions. As has been repeatedly reported in the literature, upwards saccades are launched and executed slightly faster than downwards saccades. To adjust for that we computed one grand mean y-coordinate per ms (the mean y-coordinate across all conditions) and subtracted it from the observed y-coordinates in all conditions (henceforth "normalized y-coordinate"), so as to conserve the relative differences between conditions whilst eliminating the bump created by overall quicker upward saccades. This was done to make the plot easier to read: If there was no effect the lines would be expected to stay flat at around zero. Instead, the figure suggests an increasing bias in the direction of the prime words' spatial associations as the saccade unfolds that seems to be strongest for semantically related targets and reversed for unrelated objects.

GCA was used to statistically quantify effects of Spatial Association of Prime and Prime-Target Relation on vertical saccade trajectories. It showed a strong effect of Spatial Association of Prime on the linear term with up words biasing saccades upwards relative to down words (estimate=26.693, SE=1.133, t=23.570, p<.001). Similar to the behavioral data, Spatial Association of Prime interacted with Prime-Target Relation (Chisquare(2)=568.08, p<.001) with a larger effect of Spatial Association of Prime on the linear term in the semantically related condition compared to the unrelated condition (estimate=31.111, SE=1.602, t=19.424, p<.001) and compared to the geometric shape condition (estimate=3.637, SE=1.602, t=2.271, p=0.023).

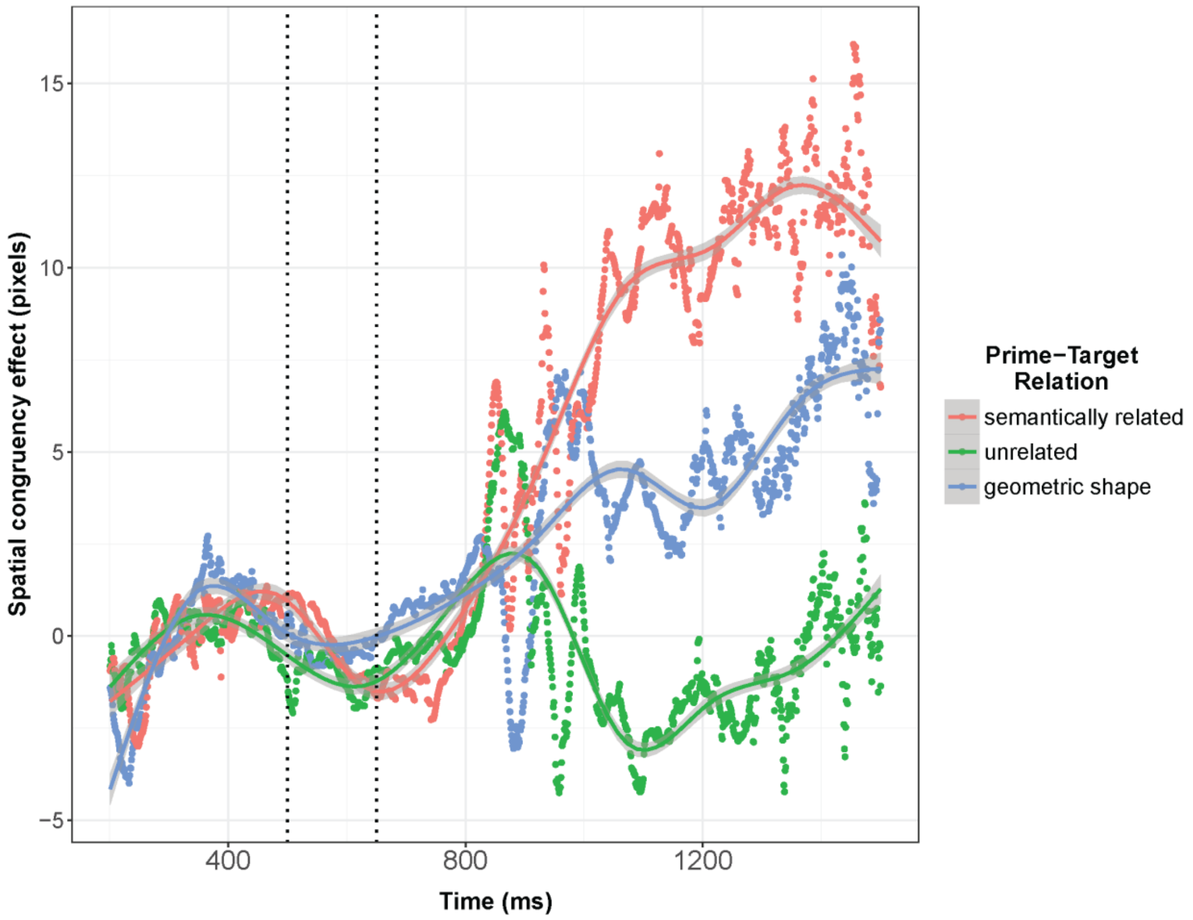


Figure 5: Spatial congruency effect (in pixels) calculated by subtracting normalized mean y-coordinates in down word trials from up word trials for all levels of Prime-Target Association (color-coded) per millisecond. Positive values indicate a bias congruent with the words' spatial associations; negative values indicate a bias in the opposite direction. The vertical dotted lines indicate the appearance of the prime word and the target, respectively.

Post-hoc tests looking at the effect of Spatial Association of Prime in all Prime-Target Relation conditions showed a large effect in the semantically related condition (estimate=61.440, SE=2.049, t=29.990, p<.001), a large effect in the geometric shape condition (estimate=23.056, SE=1.791, t=12.876, p<.001), and a small effect in the opposite direction in the unrelated condition (estimate=-4.418, SE=1.924, t=-2.296, p=.022). Figure 5 suggests that in the unrelated condition prime words initially have an effect in the congruent direction in the first third/half of the saccade that turns into a bias in the incongruent direction in the second half. To assess this statistically, we performed an exploratory follow-up analysis replacing the linear term with a quadratic term that should capture the bidirectional effect in the form of a significant negative effect of Spatial Association of Prime on the quadratic term, which is precisely what we found (estimate=-8.020, SE=1.892, t=-4.24, p<.001). An equivalent analysis for the geometric shape condition showed a non-significant trend towards a negative effect of Spatial Association of Prime on the quadratic term (estimate=2.483, SE=1.767, t=1.406, p=0.16), whereas in the semantically related condition there was a positive effect on the quadratic term (estimate=39.174, SE=4.12, t=9.509, p<.001) likely reflecting the initial dip followed by a growing increase.

These results suggest that implicitly spatial words have an initial automatic effect on concurrent saccades biasing them towards the congruent location, effectively speeding up saccades in the congruent direction and slowing down saccades in the incongruent direction. This effect can be either enhanced or reversed as the saccade unfolds depending on the relation of the word with the target. When the word and target are semantically related the initial congruency effect grows stronger over time until the saccade reaches its target. By contrast, when the target is an unrelated object, saccades in the congruent direction are slowed down as the eyes move towards the target.

It could be argued that including data points from the prime word onset until the saccade lands on the target might have contaminated the analysis with baseline differences before the saccades were initiated and with differences in target fixation after the saccade was completed. However, as can be seen in Figure 5, differences between conditions are very small in the short time-window prior to saccade onset. Nevertheless, we conducted an additional analysis restricted to the mean saccade onset until mean saccade offset (ca. 300ms after word onset until 700ms after word onset, 800-1200ms on Figures 3, 4, and 5). We found that the effect of Spatial Association of prime on the linear term remains significant (estimate=10.048, SE=1.245, t=8.07, p<.001), as does the interaction with Prime-Target Relation (Chisquare(2)=242.95, p<.001); again there was a larger effect of up vs down words on the linear term in the semantically related relative to the unrelated condition (estimate=22.607, SE=1.760, t=12.84, p<.001), however the comparison with the geometric shape condition was not significant (estimate=2.181, SE=1.760, t=1.12, p>0.4).

We next addressed a potential confound pertaining to the interaction effect. Because of the difficulty to come up with pictures that are semantically related to both up and down words (or vice versa), in our design, targets in the semantically related condition following up words ("up targets") are not the same as the ones following down words ("down targets"). It is thus possible that there are differences in the pictures themselves that influence saccade trajectories in a systematic way. Therefore we conducted a follow-up analysis that included the interaction of target type (up target vs. down target) and the linear term as additional fixed effect to account for the effect of low-level differences in the target images on saccade trajectories (we dropped geometric shape both in the Prime-Target Relation condition as well as in the Target Type condition because they could not be orthogonally crossed). The results showed that up targets indeed had a larger effect on the linear term compared to down targets (estimate=34.567, SE=0.792, t=43.648, p<.001). Crucially, the effect of Spatial Association of Prime on the linear term persisted (estimate=53.748, SE=4.747, t=11.32, p<.001), as did the interaction with Prime-Target Relation; related words have a much bigger effect on the linear term compared to unrelated words (estimate=63.327, SE=5.705, t=11.10, p<.001). This suggests that semantically related words exert a stronger influence on saccade trajectories congruent with their spatial association because of their relation with the target, not because of a low-level confound.

Discussion

In this study we set out to test the effect of words whose referents are typically observed up or down in space on concurrent eye movements. We did so by investigating whether processing such words impacts on saccade trajectories that have to be carried out in parallel, and we probed to what extent the relation between prime word and target picture has a modulatory effect, similar to what previous behavioral studies have observed (Estes et al., 2015; Ostarek & Vigliocco, 2017). Our two main findings were that 1) up/down words have a highly consistent effect on saccadic eye movements in the direction of their spatial associations and 2) the closer the eye gaze gets to the targets the stronger the semantic relation between primes and targets weighs in, an effect we attribute to increasingly foveal processing of the targets and the resulting increasing availability of semantic information.

More specifically, our results suggest that processing implicitly spatial words initially biases vertical saccade trajectories toward the direction congruent with the words' spatial association, regardless of what the target is. From about a third to halfway into the saccade, the nature of the target plays a major role such that semantically related targets trigger a further boost of the initial congruency effect, whereas unrelated targets turn the effect around and show a bias in the opposite direction until the saccade lands on the target. We suggest that this pattern is the eye-movement correlate of the location-specific priming effect observed in the RT data and that this is a prime example of eye-tracking permitting a glimpse into the moment-to-moment workings of the brain that RT data do not provide: Possibly the interference effect for unrelated targets in congruent location is small and therefore only sometimes observed (Estes et al., 2015, 2008; Gozli et al., 2013; Ostarek & Vigliocco, 2017) because it reflects two contrastive effects unfolding in time: Initial facilitation due to spatial congruence and subsequent interference due to semantic incongruence.

What is particularly striking is the short time-scale at which these complex cognitive processes occur. It has been clear from previous research that saccades are not entirely ballistic and can be modulated by the presence of distractors and attentional factors (Van der Stigchel, 2010; Van der Stigchel, Meeter, & Theeuwes, 2006). However, this study is the first to report the incorporation of such a complex set of high-level cognitive factors: During the eye movement from the center of the screen to the target, spatial aspects of word meanings are integrated with oculomotor programs that are then dynamically adapted over time according to a match vs. mismatch between primes and targets on the semantic level.

The bias in saccade trajectories in the first couple of hundreds of milliseconds after word onset can be interpreted as evidence that semantic processing of up/down words involves direction-specific processes in brain regions that control goal-directed eye movements. One possibility is that oculomotor representations are engaged directly for semantic processing, as part of a situated sensory re-instatement process (Barsalou, 2003; Barsalou, 2009; Barsalou, Simmons, Barbey, & Wilson, 2003). Indeed, accounts of situated conceptualization predict that comprehension of a word (such as *bird*) involves the activation not only of features related to the item itself but also of features related to situations in which it is typically perceived (such as looking up to see the bird somebody is talking about in its nest). Our results that reading up/down words quickly activates direction-specific oculomotor processes and that semantic priming is contingent on the target appearing in its typical location are thus highly compatible with this view. An alternative is that semantic spatial information is retrieved from a high-level spatial system that interfaces with oculomotor regions and activates direction-specific processes therein. Anatomically, this is plausible because systems for covert and overt spatial attention are largely overlapping (Beauchamp, Petit, Ellmore, Ingeholm, & Haxby, 2001; Corbetta et al., 1998; Gitelman et al., 1999; Thompson, Biscoe, & Sato, 2005). As neurons responsible for oculomotor function and covert spatial attention allocation may be intermingled in the same cell populations and are likely to strongly interact (Petersen & Posner, 2012; Schafer & Moore, 2007; Thompson et al., 2005), these two possibilities are very difficult to tease apart. Future studies could investigate to what extent the activation of oculomotor processes by implicitly spatial words is dependent on a secondary task involving eye movements.

Another interpretation of our results is that the modulation of eye movements is the consequence of a location and object matching process (Estes et al., 2015), to the extent that this process directly interacts with oculomotor processes. Under this account the facilitation-to-interference effect for unrelated targets in compatible locations and the strong interference for semantically related targets in incompatible locations can be considered to reflect a clash of the location and object codes. However, it incorrectly predicts the same pattern for geometric shapes that are assumed to be coded as a mismatch for object identity in the same way as unrelated pictures. Moreover, as it stands, facilitation is only predicted for perceptually matching targets, whereas our results suggest that semantically related, but perceptually dissimilar, targets show facilitation in the compatible location. A matching explanation thus could only succeed with the amendment that object codes depend on semantic overlap and that the amount of semantic similarity maps non-linearly onto the outcome: High similarity (semantically related target) produces strong facilitation, no similarity (geometric shapes) produces moderate facilitation, but small degrees of similarity (unrelated pictures that are inanimate object items as the prime words' referents) produce interference. Our results are thus incompatible with the perceptual matching account as initially formulated as the ease of processing of a visual target in compatible location seems to depend on the semantic, not perceptual, relation with the prime word.

Both the location-specific priming effect for semantically related targets and the interference effect for unrelated targets suggest that congruence on one level and incongruence on another is more detrimental than incongruence on both. Previous studies showed that this pattern is only observed at short SOAs (Gozli et al., 2013; Ostarek & Vigliocco, 2017). This is reminiscent of, and can be accounted for, by the theory of event coding according to which sensory-motor features belonging to the same event are temporarily bound into an "event file" (Amer et al., 2017; Hommel, 1998, 2004; Rothermund et al., 2005). As discussed in the Introduction, TEC predicts interference for unrelated targets in compatible locations at short SOAs when spatial features of words are bound and thus unavailable for the visual task, and it predicts facilitation at long SOAs when spatial features are expected to be pre-activated but no longer be bound. Our data, as well as Ostarek and Vigliocco's (2017), suggest that items with partial semantic overlap can be integrated in one event file. An open question is what the boundary conditions are for successful integration, i.e., what kind and amount of features need to overlap for facilitation to occur.

Related to the last point are the somewhat surprising results in the geometric shape condition. Whereas the eye-tracking and RT effects in the semantically related and unrelated conditions pair up nicely, the two measures show a marked difference in the geometric shape condition: The RT data suggest a trend towards interference, but the eye movements show a strong congruency effect. To speculate, it seems possible that the complete lack of feature overlap makes it easier to overcome the inability to integrate the target with the activated event (or event file). This could explain the smaller interference effect in terms of RT and lead to the initial up/down bias in the eye movements to persist.

To conclude, the present study adds to the growing literature on the activation of concepts' typical location in vertical space and, for the first time, shows an interaction of implicit up/down words with direction-specific oculomotor processes during online word processing. Semantic processing of implicitly spatial nouns was found to induce a bias in early stages of concurrently executed saccades that can grow stronger over time or reverse depending on the semantic relation between the prime words and target pictures.

Acknowledgements

We would like to thank all participants for their participation and Agnieszka Konopka for a very helpful introduction to growth curve analysis.

References

Amer, T., Gozli, D. G., & Pratt, J. (2017). Biasing spatial attention with semantic information: an event coding approach. *Psychological Research*, 1–19.

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

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. *Theories of Memory*, *1*, 29–31.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

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

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, *23*(4), 1122–1142.

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

Beauchamp, M. S., & Martin, A. (2007). Grounding object concepts in perception and action: evidence from fMRI studies of tools. *Cortex*, *43*(3), 461–468.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., & Haxby, J. V. (2001). A parametric fMRI study of overt and covert shifts of visuospatial attention. *Neuroimage*, *14*(2), 310–321.

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

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.

Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage*, *143*, 128–140.

Borghesani, V., & Piazza, M. (2017). The neuro-cognitive representations of symbols: the case of concrete words. *Neuropsychologia*.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919.

Chen, L., Lambon Ralph, M. A. L., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, *1*, 0039.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., … Shulman, G. L. (1998). A Common Network of Functional Areas for Attention and Eye Movements. *Neuron*, *21*(4), 761–773. https://doi.org/10.1016/S0896-6273(00)80593-0

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012). From top to bottom: spatial shifts of attention caused by linguistic stimuli. *Cognitive Processing*, *13*(1), 151–154.

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. *PloS One*, *8*(2), e56872.

Dunn, B. M. (2016). *Which way is up? Grounded mental representations of space*. University of Glasgow.

Estes, Z., Verges, M., & Adelman, J. S. (2015). Words, objects, and locations: Perceptual matching explains spatial interference and facilitation. *Journal of Memory and Language*, *84*, 167–189.

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down object words orient attention to the objects' typical location. *Psychological Science*, *19*(2), 93–97.

Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., & Binder, J. R. (2016). Heteromodal Cortical Areas Encode Sensory-Motor Features of Word Meaning. *Journal of Neuroscience*, *36*(38), 9763–9769. https://doi.org/10.1523/JNEUROSCI.4095-15.2016

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., … Seidenberg, M. S. (2016). Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. *Cerebral Cortex*, *26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Francken, J. C., Kok, P., Hagoort, P., & De Lange, F. P. (2014). The behavioral and neural effects of language on motion perception. *Journal of Cognitive Neuroscience*.

Francken, J. C., Meijs, E. L., Hagoort, P., Van Gaal, S., & De Lange, F. P. (2015). Exploring the automaticity of language-perception interactions: Effects of attention and awareness. *Scientific Reports*, *5*.

Francken, J. C., Meijs, E. L., Ridderinkhof, O. M., Hagoort, P., de Lange, F. P., & van Gaal, S. (2015). Manipulating word awareness dissociates feed-forward from feedback models of language-perception interactions. *Neuroscience of Consciousness*, *2015*(1), niv003.

Gitelman, D. R., Nobre, A. C., Parrish, T. B., LaBar, K. S., Kim, Y.-H., Meyer, J. R., & Mesulam, M.-M. (1999). A large-scale distributed network for covert spatial attention. *Brain*, *122*(6), 1093–1106.

Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit spatial cues for visual attention. *Journal of Experimental Psychology: General*, *142*(4), 1028.

Gozli, D. G., Pratt, J., Martin, K. Z., & Chasteen, A. L. (2016). Implied spatial meaning and visuospatial bias: Conceptual processing influences processing of visual targets and distractors. *PloS One*, *11*(3), e0150928.

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

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

Hommel, B. (1998). Automatic stimulus–response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(5), 1368.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500.

Huettig, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–150.

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*(4), 1068–1075.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650.

Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1634), 20120392.

Lambon Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284.

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*(8), 1764–1801.

Lenth, R. V., & others. (2016). Least-squares means: the R package lsmeans. *J Stat Softw*, 69(1), 1–33.

Lewis, G., & Poeppel, D. (2014). The role of visual representations during the lexical access of spoken words. *Brain and Language*, *134*, 1–10.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 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*(1), 59–70.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804.

Meteyard, L., Zokaei, N., Bahrami, B., & Vigliocco, G. (2008). Visual motion interferes with lexical decision on motion words. *Current Biology*, *18*(17), R732–R733.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494.

Ostarek, M., & Huettig, F. (2017a). A Task-Dependent Causal Role for Low-Level Visual Processes in Spoken Word Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1215.

Ostarek, M., & Huettig, F. (2017b). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 499.

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 579.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89.

Pobric, G., Lambon Ralph, M. A. L., & Jefferies, E. (2009). The role of the anterior temporal lobes in the comprehension of concrete and abstract words: rTMS evidence. *Cortex*, *45*(9), 1104–1110.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.

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.

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, *24*(11), 2218–2225.

Rothermund, K., Wentura, D., & De Houwer, J. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 482.

Saygin, A. P., McCullough, S., Alac, M., & Emmorey, K. (2010). Modulation of BOLD response in motion-sensitive lateral temporal cortex by real and fictive motion sentences. *Journal of Cognitive Neuroscience*, *22*(11), 2480–2490.

Schafer, R. J., & Moore, T. (2007). Attention governs action in the primate frontal eye field. *Neuron*, *56*(3), 541–551.

Shtyrov, Y., Butorina, A., Nikolaeva, A., & Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences*, *111*(18), E1918–E1923.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, *12*(4), 282–286.

Thompson, K. G., Biscoe, K. L., & Sato, T. R. (2005). Neuronal basis of covert spatial attention in the frontal eye field. *Journal of Neuroscience*, *25*(41), 9479–9487.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *J. Exp. Psychol*, *86*(2), 255–262.

van Dam, W. O., Speed, L. J., Lai, V. T., Vigliocco, G., & Desai, R. H. (2017). Effects of motion speed in action representations. *Brain and Language*, *168*, 47–56.

van Dam, W. O., van Dijk, M., Bekkering, H., & Rueschemeyer, S.-A. (2012). Flexibility in embodied lexical-semantic representations. *Human Brain Mapping*, *33*(10), 2322–2333.

Van der Stigchel, S. (2010). Recent advances in the study of saccade trajectory deviations. *Vision Research*, *50*(17), 1619–1627.

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience & Biobehavioral Reviews*, *30*(5), 666–679.

Verges, M., & Duffy, S. (2009). Spatial representations elicit dual-coding effects in mental imagery. *Cognitive Science*, *33*(6), 1157–1172.

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, *96*, 222–229.

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). BOLD response to motion verbs in left posterior middle temporal gyrus during story comprehension. *Brain and Language*, *119*(3), 221–225.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027.

Zhang, E., Luo, J., Zhang, J., Wang, Y., Zhong, J., & Li, Q. (2013). Neural mechanisms of shifts of spatial attention induced by object words with spatial associations: an ERP study. *Experimental Brain Research*, 227(2), 199–209.

Appendix

Stimulus characteristics

The table displays means and standard deviations per prime word condition

	Related up	Related down	Unrelated up	Unrelated down	Control
Log-Frequency	2.57 (0.70)	2.35 (0.81)	2.59 (0.63)	2.75 (0.59)	2.61 (0.63)
Age of Acquisition	6.98 (2.52)	6.63 (1.87)	6.7 (2)	6.46 (1.46)	6.17 (1.37)
Syllables	1.79 (0.8)	1.79 (0.7)	1.79 (0.88)	1.64 (0.68)	1.82 (0.77)
Letters	5.86 (2.28)	6.86 (2.71)	6.21 (2.3)	5.57 (1.37)	6 (1.56)
Phonemes	5 (2.15)	6 (2.42)	5.29 (1.94)	4.79 (1.34)	5.25 (1.62)

Names of target pictures and corresponding prime words

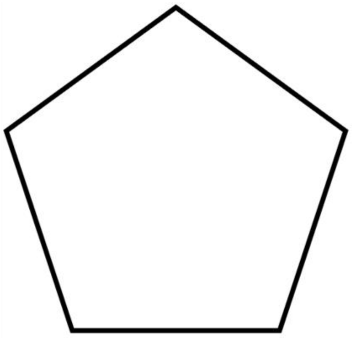
Targets	**Related up**	**Related down**	**Unrelated up**	**Unrelated down**	**Control**
pet	hoed		gordijnen	vloer	nagel
vliegtuig	cockpit		valk	octopus	glas
dak	zolder		vlinder	kakkerlak	telefoon
hoofd	oog		satelliet	anker	handdoek
engel	vleugel		toren	duikboot	kaars
koning	kroon		loopbaan	hoeven	taart
maan	ster		bladeren	rivier	schaar
nest	duif		vlieger	kelder	waterkoker
raket	projectiel		licht	dolfijn	kaas
vlag	pijler		voorhoofd	oceaan	horloge
vogel	uil		wolkenkrabber	grot	zwaard
wolk	hemel		lamp	slang	meloen
zon	regenboog		top	duiker	wortel
paraplu	parachute		planeet	aarde	envelop
voet		teen	adelaar	tram	doek
egel		muis	lamp	zand	koren
fiets		motorfiets	raam	stoep	hout
gras		wortels	vlieg	sok	voetbal
kikker		vijver	spreeuw	walvis	fles
schoen		sandaal	ooglid	modder	cirkel
schip		boot	helm	broek	citroen
tapijt		kleedje	astronaut	valstrik	tonijn
tent		schuilplaats	piek	straat	rubber
trein		spoor	boom	steen	potlood
tunnel		metro	ballon	graf	kameel
tractor		aanhangwagen	kasteel	slang	papier
wiel		huifkar	kraag	been	kokosnoot
bed		kussen	neus	asfalt	muur

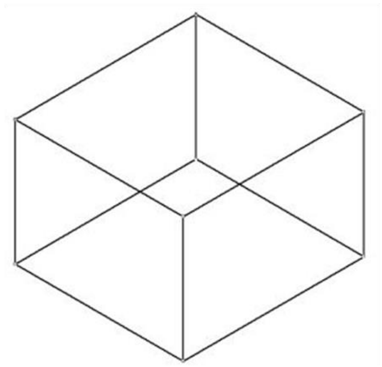
English translation of stimuli

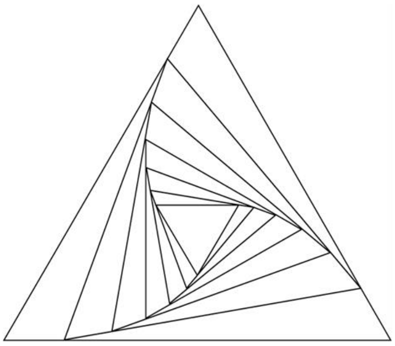
Targets	Related up	Related down	Unrelated up	Unrelated down	Control
cap	hat		curtains	floor	nail
airplane	cockpit		falcon	octopus	glass
roof	attic		butterfly	cockroach	telephone
head	eye		satellite	anchor	towel
angel	wing		tower	submarine	candle
king	crown		orbit	hooves	cake
moon	star		leaves	river	scissors
nest	pigeon		kite	cellar	kettle
rocket	missile		light	dolphin	cheese
flag	pillar		forehead	ocean	watch
bird	owl		skyscraper	cave	sword
cloud	sky		lamp	snake	melon
sun	rainbow		summit	diver	carrot
umbrella	parachute		planet	earth	envelope
foot		toe	eagle	tram	cloth
hedgehog		mouse	lamp	sand	choirs
bicycle		motorbike	window	pavement	wood
grass		roots	fly	sock	football
frog		pond	starling	whale	bottle
shoe		sandal	eyelid	mud	circle
ship		boat	helmet	pants	lemon
carpet		rug	astronaut	snare	tuna
tent		shelter	peak	street	rubber
train		rail	tree	stone	pencil
tunnel		metro	balloon	grave	camel
tractor		trailer	castle	snake	paper
wheel		carriage	collar	leg	coconut
bed		pillow	nose	asphalt	wall

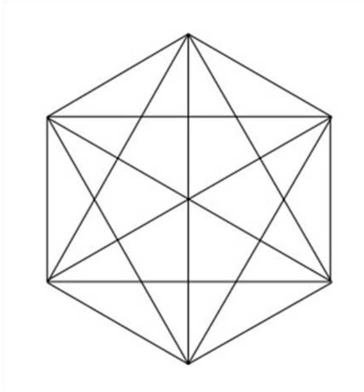
Examples of the geometric shape stimuli

The geometric shapes ranged from more 'standard' ones (top two) to more complex ones (bottom two)









Chapter 6

CROSS-DECODING REVEALS SHARED BRAIN ACTIVITY PATTERNS BETWEEN SACCADIC EYE-MOVEMENTS AND SEMANTIC PROCESSING OF IMPLICITLY SPATIAL WORDS

Based on: Ostarek, M., van Paridon, J., Evans, S., & Huettig, F. (manuscript in preparation). Cross-decoding reveals shared brain activity patterns between saccadic eye movements and semantic processing of implicitly spatial words.

Abstract

Processing words with referents that are typically observed up or down in space (up/down words) influences the subsequent identification of visual targets in congruent locations. Eye-tracking studies have shown that up/down word comprehension shortens launch times of subsequent saccades to congruent locations and modulates concurrent saccade trajectories. This can be explained by a task-dependent interaction of semantic processing and oculomotor programs or by a direct recruitment of direction-specific processes in oculomotor and spatial systems as part of semantic processing. To test the latter possibility, we conducted a functional magnetic resonance imaging experiment and used multi-voxel pattern analysis to assess 1) whether the typical location of word referents can be decoded from the fronto-parietal spatial network and 2) whether activity patterns are shared between up/down words and up/down saccadic eye movements. In line with these hypotheses, significant decoding of up vs. down words and cross-decoding between up/down saccades and up/down words were observed in the frontal eye field region in the superior frontal sulcus and the inferior parietal lobule. Beyond these spatial attention areas, typical location of word referents could be decoded from a set of occipital, temporal, and frontal areas, indicating that interactions between high-level regions typically implicated with lexical-semantic processing and spatial/oculomotor regions constitute the neural basis for access to spatial aspects of word meanings.

Introduction

The capacity of the adult human brain to virtually effortlessly derive abstract meaning from language input poses an enormous challenge for contemporary cognitive neuroscience. One influential proposal has been that language comprehension is made possible by grounding it in the sensory and motor systems of the brain (Barsalou, 1999; Barsalou, Simmons, Barbey, & Wilson, 2003). According to this proposal, conceptual processing of spoken and written words involves a recycling process: It necessitates the re-recruitment of sensory-motor processes that were consistently activated during previous encounters with the 'real world' referents the words denote. In the past decades, our understanding of the cognitive and neural mechanisms underlying conceptual processing has advanced considerably (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017). Recent evidence suggests that the brain's solution to the complex problem of knowledge storage and conceptual processing has indeed not been to evolve a new anatomically defined self-sufficient semantic module but to employ a widely distributed network that seems to encompass most of the cerebral cortex (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Mitchell et al., 2008).

When it comes to the question of how conceptual knowledge is represented in neural codes, most studies to-date have targeted concepts with concrete referents that are straightforwardly described in terms of their sensory-motor features. For instance, action words have often been found to activate motor areas implicated with executing or planning movements required for the action (Hauk, Johnsrude, & Pulvermüller, 2004; Vukovic, Feurra, Shpektor, Myachykov, & Shtyrov, 2017). Similarly, words referring to objects with salient visual characteristics recruit the corresponding visual processes/areas (shape: Correia et al., 2014; Lewis & Poeppel, 2014; Ostarek & Huettig, 2017a, 2017b; color: Simmons et al., 2007; speed of motion: van Dam, Speed, Lai, Vigliocco, & Desai, 2017). Recent studies using multivariate pattern analyses (MVPA) suggest that activation patterns in areas associated with high level semantic processing (Binder, Desai, Graves, & Conant, 2009) at least partly reflect conjunctive coding of multiple sensory-motor attributes (Fernandino, Humphries, Conant, Seidenberg, & Binder, 2016; Fernandino et al., 2015), allowing encoding models based solely on sensory-motor attributes to predict voxel-wise activity patterns in response to concrete - but not abstract - words

(Fernandino et al., 2016). Thus, the emerging picture is a hierarchical organization of semantic knowledge derived from experiential sensory-motor experience (Fernandino et al., 2016) that may culminate in holistic multimodal concept-level neural codes in high-level areas (Binder & Desai, 2011; Coutanche & Thompson-Schill, 2014; Fernandino et al., 2016; Lambon Ralph et al., 2017).

We thus have a growing understanding of both the architecture and the processes related to item-dependent concepts with referents that map onto a set of sensory-motor features. However, it is still largely unclear what neurocognitive mechanisms are recruited for item-*independent* conceptual dimensions. These include abstract concepts, such as *justice*, which do not have tangible referents with consistent physical properties, and spatial dimensions, such as *up* and *down*. Researchers have so far struggled to explain how this important aspect of our conceptual knowledge is enabled on the cognitive and neural level. At least partly, this is due to the fact that researchers have thus far failed to get a good grasp on the kinds of relevant features because of the lacking consistency of features across instantiations of most abstract concepts (Borghi et al., 2017).

Here, we tested the hypothesis that the dorsal pathway provides semantic information about vertical space, a dimension largely orthogonal to features of individual items (Lambon Ralph et al., 2017). More specifically, we examined the processing of words that differ in the vertical spatial location where their referents are typically observed (e.g., *bird* vs. *foot*, henceforth 'up/down words') in order to address how the abstract features 'up' and 'down' are reflected in neural population codes using multi-voxel pattern analysis (MVPA). The vertical dimension is not only abstract because it is item-independent, but also because it can only be indirectly experienced by virtue of being defined through varying reference frames. This makes the dimension a great testbed for the scope of the conceptual grounding hypothesis; if a dimension that is in principle an ideal candidate for abstract representation in a high-level system is nevertheless anchored in sensory mechanisms, this would indicate that grounding may be the primary route to meaning.

If this were the case, it would make sense to capitalize on to the most reliable way of engaging with up/down items. The best candidate is the cortical substrate for vertical looking behavior, primarily the one related to saccadic eye movements. This is because looking behavior is a prime attentional gate keeper that largely determines which entities we consciously perceive. Therefore, in a typical situation in which you perceive a bird (perhaps because somebody says "Look at that big bird") you move your eyes up to look at it.

Numerous studies have shown that processing up/down words can influence performance at visual tasks involving compatible vs. incompatible location (Estes, Verges, & Adelman, 2015; Estes, Verges, & Barsalou, 2008; Gozli, Chasteen, & Pratt, 2013; Ostarek & Vigliocco, 2017). Behavioral work has established that processing up/down words enhances spatial attention to the compatible location which often leads to facilitated detection and identification of visual targets in the primed location (Dudschig, Lachmair, de la Vega, De Filippis, & Kaup, 2012; Gozli et al., 2013; Ostarek & Vigliocco, 2017). Interference in the compatible location is typically observed when stimulus onset asynchrony between prime words and visual targets is short (< 400ms) and when they are semantically unrelated (Estes et al., 2015, 2008; Gozli et al., 2013). Recent eye-tracking studies found that up/down words speed up subsequent spatially compatible saccades (Dudschig et al., 2013; Dunn, 2016) and influence concurrent saccades in the congruent direction (Ostarek, Ishag, Joosen, & Huettig, in press), suggesting that processing these words pre-activates specific motor programs in saccade-related brain areas. However, as these studies relied on paradigms with visual targets in up/down locations, inferences about processes directly activated by the words (in the absence of a visual-spatial task) are limited.

To directly address this question, we assessed whether conceptual processing of up/down words activates direction-specific patterns in the cortical network for saccadic eye movements and spatial attention (Corbetta et al., 1998; Grosbras, Laird, & Paus, 2005; Paus, 1996; Vernet, Quentin, Chanes, Mitsumasu, & Valero-Cabré, 2014). In particular, we predicted that a linear classifier could decode whether a word is 'up' or 'down' from multi-voxel activity patterns in the frontal eye field in the premotor cortex, the supplementary eye field in the supplementary motor area, the intraparietal sulcus as well as surrounding cortex in the superior and inferior parietal lobules. We further predicted shared activity patterns with up/down saccadic eye movements in these regions. A previous decoding study has reported that spatial judgements about visually presented geometric shapes (slightly above vs. below the centre) and spatial judgements about up/down words (whether the word referent's typical location is up vs. down) trigger similar patterns in the inferior parietal lobule (Quadflieg et al., 2011). Although this finding is consistent with a shared neural substrate for spatial judgements based on different inputs, it remains unclear whether semantic processing of these words activates direction-specific patterns in the brain networks for spatial attention. Here, we used a task (concreteness judgement) that did not allude to the words' spatial properties (and this dimension went entirely un-noticed by the participants). Thus, we were able to directly test the involvement of the oculomotor network in semantic processing.

Methods

Participants

We recruited 20 native Dutch participants (11 female, mean age: 22) from the local Radboud University database. One participant was excluded because she fell asleep during the experiment and for another participant no data were collected because the scanner stopped working. All analyses we report were based on the remaining 18 participants (10 female, mean age: 21). Participants gave written consent and were paid 12.50 euros. We had ethical approval from the local ethics committee (dossier CMO 2014/288), assigned by the accredited local ethics board CMO Arnhem-Nijmegen.

MRI Acquisition

Data acquisition was performed on a Siemens Prisma Fit 3T scanner at the Donders Center for Cognitive Neuroimaging using a 32-channel head coil. Functional images were acquired using a multiband sequence with 68 slices, acceleration factor = 4, spatial resolution = 2.0x2.0x2.0 mm, TR = 1.5 s, TE = 39.40 ms. There were two functional runs (ca. 11 minutes each) for the language task and one run for the saccadic eye movement task (ca. 15 minutes). To obtain a high-resolution anatomical image, a t1-weighted MPRAGE sequence was run (192 volumes, spatial resolution = 1mm).

Stimuli, Procedure, and Design

After receiving task instructions, participants were accompanied into the scanning room where they were asked to put earphones on after which their head was stabilised with cushions and the head coil was put in place. Before the experiment started, we made a sound check by playing a word stimulus repeatedly with the scanner noise on, giving the participants an opportunity to adjust the volume until they could hear the words well. Then, the two language runs were carried out, followed by the saccadic eye movement run, and finally by the anatomical scan.

The main experiment used 24 concrete and 24 abstract filler words. Of the concrete words (see Table 1), 12 referred to objects that are usually observed in the lower visual field (shoe), and 12 referred to objects that are usually up (bird). These two word classes were closely matched for frequency (SUBTLEX: Keuleers, Brysbaert, & New, 2010), age of acquisition (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014), concreteness (Brysbaert et al., 2014) number of syllables, number of letters, and number of phonemes (all p-values > 0.4). The abstract filler words differed strongly in concreteness ($p<0.001$), but were similar in frequency, length, number of letters, phonemes, and syllables ($p>0.05$).

	Frequency	AoA	Concreteness	Letters	Phonemes	Syllables
up words	26.90 (21.19)	6.09 (1.27)	4.63 (0.28)	4.42 (1.38)	4.25 (1.22)	1.25 (0.45)
down words	26.86 (27.34)	5.60 (1.78)	4.69 (0.24)	4.83 (1.19)	4.50 (1.17)	1.33 (0.49)

Table 1: Stimulus characteristics. Means per word category and standard deviations in parentheses. AoA = Age of acquisition.

A different group of 38 participants were asked to rate the up/down words in terms of whether they refer to objects that are typically perceived up vs. down in space. On a Likert scale ranging from 1="always down" to 7="always up", down words were given a mean rating of 1.84 (SD=1.15) compared to a mean of 6.25 (SD=0.99) for up words ($p<0.001$).

There was an auditory and a written version of each item intended to avoid low-level confounds in classification. This precautionary measure was taken because classifiers are very sensitive and might otherwise pick up on imperfectly matched low-level stimulus characteristics that differ slightly between word types (even though this was unlikely given that the word type were closely matched for these). The spoken and written word versions were both repeated three times, amounting to six repetitions per word across the whole language experiment and to 288 trials in total (ca. 22 minutes). The order of trials was fully randomized per run. Trials lasted 4.5 seconds (3 TRs). Every spoken word trial began with a spoken word (average: 600ms) accompanied with a fixation dot that remained on the screen until the end of the trial. In written word trials, the written word appeared centrally for 600ms and was then replaced by the fixation dot for the remaining 3900ms. Participants were asked to respond by pressing the correct button (left or right index finger) as quickly as possible without sacrificing accuracy right after word presentation. The task was to decide whether the word was concrete or abstract (concreteness task).

After the language task, participants performed 71 mini-blocks of saccadic eye-movements in which they had to follow a fixation cross with their eyes as closely as possible (the method was based on Knops, Thirion, Hubbard, Michel, & Dehaene, 2009). On every trial, the fixation cross was first presented centrally for 3s and then moved either to the top or bottom of the screen in four steps that were slightly jittered on the vertical and horizontal axis: On the vertical axis the visual angle at each step randomly varied between 1.71-2.14° and the horizontal location was randomly jittered between 0 and 0.64° to the left or right of the centre on the horizontal axis. The speed at which the four steps occurred was varied randomly across mini-blocks (the fixation cross remained at each location for 750ms, 1125ms, or 1500ms). After the fourth step the fixation cross remained at the last position for 6750ms, 5650ms, or 4500ms depending on the speed of the previous steps such that the total time per mini-block was kept constant at 12s.

Eye-tracking pre-test

As recent studies indicated that processing up/down words can modulate subsequent and concurrent saccades, we conducted a pre-test to rule out that participants move their eyes up/down during central fixation when performing a semantic task on these words. We recruited a separate group of 36 participants for an eye-tracking experiment using an Eye-Link 1000 tracker (SR Research) in which they performed the same task (concreteness judgment) on the same set of words we used in the MRI experiment. Spoken words were presented via headphones and participants were asked to maintain central fixation. A linear mixed effects model was used to analyse possible differences in vertical looking behavior as a function of the words' spatial associations. The mode included spatial association (up vs. down) as fixed effect and had per-participant intercepts and slopes. Our results showed that central fixation was maintained remarkably well and there was no effect of the spatial association of words on y-coordinates (up words: $M=5.65$, $SD=2.30$; down words: $M=5.65$, $SD=2.25$; $t<1$).

Preprocessing

Preprocessing of the volumes was done using the SPM12 fMRI toolkit for MATLAB. Volumes were slice time corrected to the middle slice and spatially realigned to the first volume of the first run. Volumes were co-registered to the t1-weighted structural scan and the structural scan was segmented by tissue type using SPM12's segmentation subroutine. Both the functional volumes and the structural scans were normalized to the MNI template.

General linear model

Functional volumes were entered into a general linear model in which each trial was modelled as a separate regressor (the six motion parameters and per-run parameters were included as regressors of no interest). The regressors corresponding to word events were convolved with the canonical hemodynamic response function and the model parameters were estimated. This procedure yielded 144 beta maps. Beta maps were then aggregated by word and t-value maps were computed (with each t-map being informed by six beta maps of the same word, three of which corresponding to spoken and three to written presentation of the word). This procedure yielded 24 t-maps per run and 48 t-maps in total (i.e. one per item).

Searchlight analysis

From the 48 word t-maps, the 24 maps associated with concrete words were selected, discarding the 24 maps associated with abstract words. The 24 concrete word maps were further split into a group of 12 maps associated with up words and a group of 12 maps associated with down words.

To identify brain regions whose patterns distinguish up words from down words, a whole-brain searchlight procedure was used, as proposed by Kriegeskorte, Goebel, and Bandettini (2006) and implemented in the PyMVPA toolkit (Hanke et al., 2009). For each voxel in the map, a sphere with a radius of 3 voxels around that voxel was selected and a linear kernel support vector machine classifier was trained on that sphere of voxels in twenty of the t-maps and then tested on the four remaining t-maps. This procedure was cross-validated by repeating it on 10 different balanced folds of the dataset per sphere to ensure a stable estimate of prediction accuracy.

To identify clusters with significant prediction accuracy, a combination of training set permutation and bootstrapping followed by cluster-size control and correction for multiple testing was used, as proposed by Stelzer, Chen, and Turner (2013), as implemented in the PyMVPA toolkit (Hanke et al., 2009): To allow statistical inferences based on the searchlight prediction accuracy, an empirical null distribution was estimated based on a permutation-bootstrapping scheme. In the permutation step, a chance accuracy map was generated by repeating the searchlight and cross-validation procedure with randomly shuffled t-map labels. This procedure was repeated 100 times per participant. A group-level null distribution was then estimated using a bootstrap procedure by selecting one chance accuracy map per participant and computing an averaged group chance map 100,000 times.

Statistically significant clusters were identified by first clustering contiguous voxels significant at the .001-level, computing the cluster-wise p-value and then correcting these p-values using the Benjamini and Hochberg method for false discovery rate correction (Benjamini & Hochberg, 1995).

To assess whether brain activity patterns were shared between looking up vs. down and processing an up vs. down word, we performed a cross-classification analysis. Cross-decoding from saccades (upwards/downwards) to words (up/down) was done by performing a searchlight analysis (radius = 3 voxels) in which a linear support vector machine classifier was trained on the patterns related to the saccadic eye-movement data and then tested on the patterns corresponding to the 24 up/down words. To ensure stable estimates, for every searchlight we used a cross-validation scheme where the classifier was trained on all but six randomly chosen saccade trials and tested on all up/down words. This procedure was repeated 10 times and the average accuracy across the 10 folds was taken as the central voxel's cross-decoding accuracy value. Cluster-based permutation testing was performed by again repeating this analysis 100 times per participant with randomly shuffled labels in the training set to then estimate a group-level null distribution as described above for up/down word decoding.

Results

Behavioral results

Participants performed the concreteness task accurately and, importantly, achieved highly similar accuracies for up words (93.85%) and down words (94.01%) that were statistically indistinguishable ($t<1$). Reaction times to up words ($M=1105$ ms, $SD=435$ ms) and down words ($M=1123$ ms, $SD=462$ ms) were also very similar ($t<1$).

MVPA Results

As a first step, we used searchlight analysis to identify brain regions whose multi-voxel patterns can discriminate up from down words (see Figure 1). Cluster-based permutation tests revealed successful decoding in a set of regions linked to saccadic eye-movements and spatial attention, including the posterior portion of the superior frontal sulcus, the middle frontal gyrus, the inferior and superior parietal lobule, posterior temporal lobe, the cerebellum, and portions of the occipital cortex including primary visual cortex.

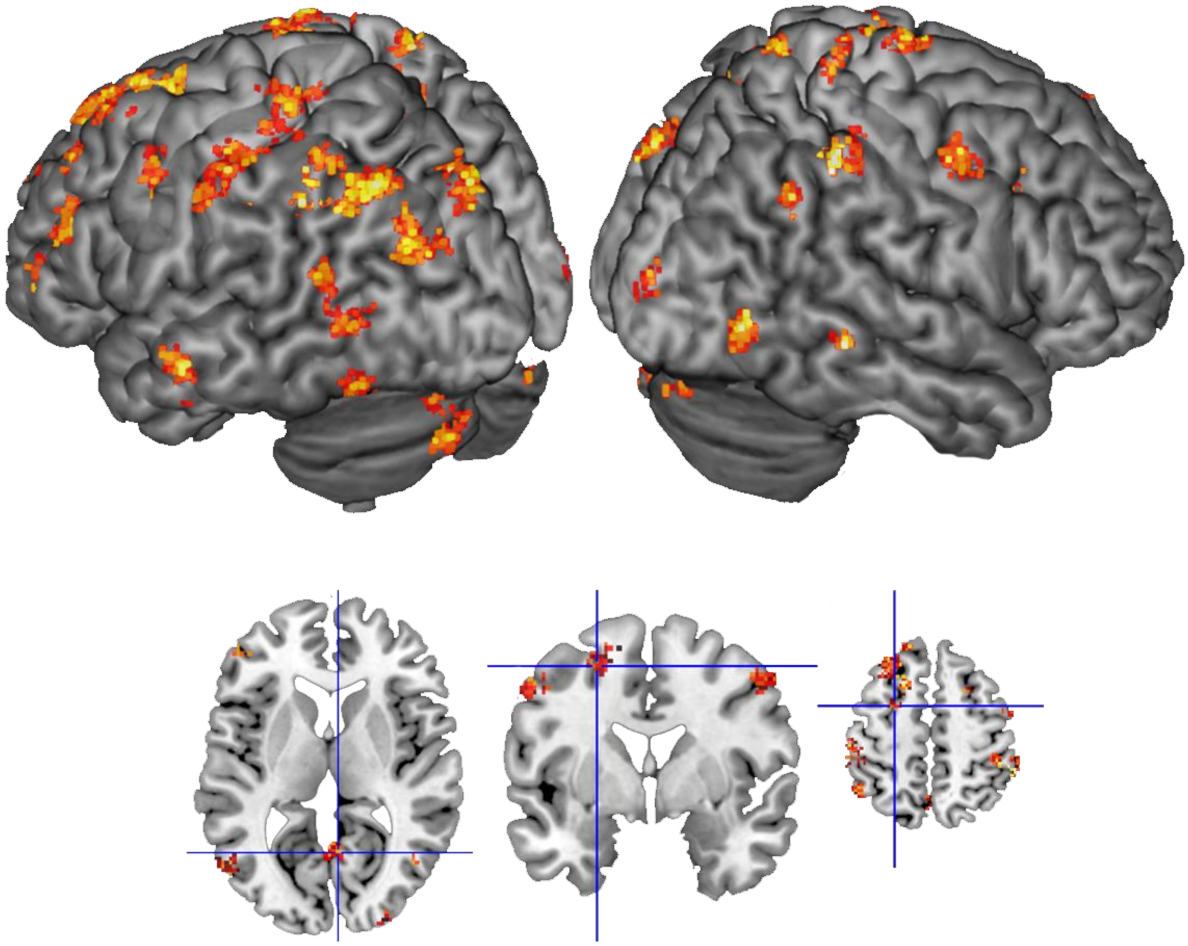


Figure 1: Up/down word decoding results. Top: Significant clusters on the cortical surface. Bottom (from left to right): Axial slice with crosshairs on the right primary visual cortex; coronal and axial slice with crosshairs on the posterior superior frontal sulcus. Brighter colors indicate higher classification accuracies.

To provide direct evidence that up/down words engage processes similar to actual up/down shifts of attention, we sought to train a classifier to distinguish up vs. down saccades and tested it on the up/down words without further training. Decoding saccade direction is expected to be possible based on activity patterns in a network of occipital, parietal, and frontal areas. However, we obtained above-chance decoding in virtually the entire brain and there was a prominent peak with decoding accuracies close to 1 in the white matter behind the eyes. This suggests that the eye movements induced small motion artefacts. We carried out cross-decoding nevertheless based on the following grounds: 1) Peaks in saccade decoding (looking up vs. down) were observed in the occipital cortex, posterior parietal cortex, and the frontal eye field region at the intersection of the superior frontal sulcus and the precentral sulcus, as expected (Figure 2 shows regions with decoding accuracies higher than 60%). 2) It seems implausible that motion artefacts induce fine-grained activation patterns that resemble the ones evoked by up/down words. Instead, generalization of patterns from the saccade to the word domain should only succeed based on brain activity related to saccade planning and execution.

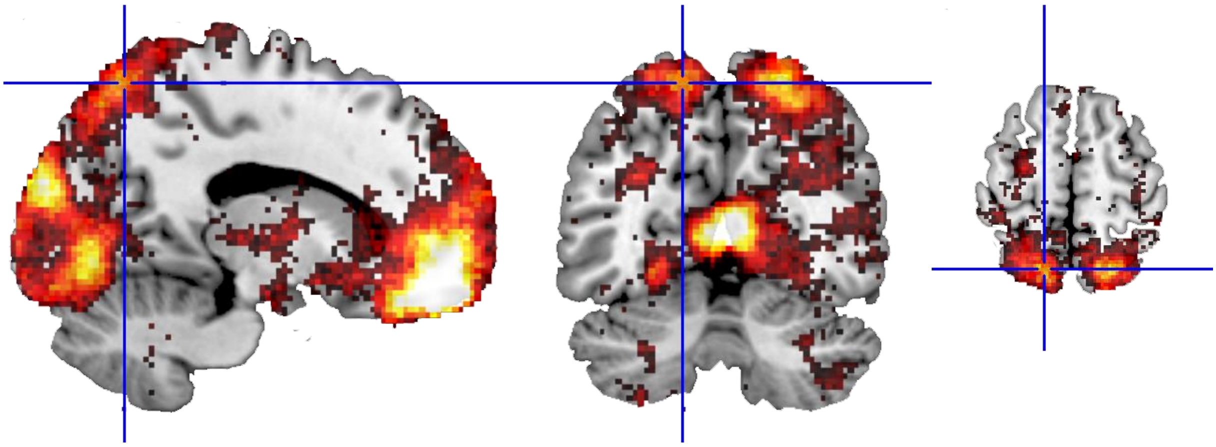


Figure 2: Saccade direction decoding. Sagittal (left), coronal (middle), and axial (right) slices plotting decoding accuracies > 60%. Brighter colors indicate higher classification accuracies.

In line with this rationale, no significant cross-decoding was observed in the white matter behind the eyes that appeared to be most susceptible to motion artefacts. Instead, the cross-decoding approach revealed very robust clusters in the intraparietal sulcus, superior parietal lobule, inferior parietal lobule, the intersection of the superior frontal sulcus and the precentral sulcus (FEF), supplementary motor area, prefrontal cortex, precuneus, posterior middle and superior temporal lobe, posterior inferior temporal cortex, cingulate gyrus, and in the occipital lobe including primary visual cortex (see Figure 3).

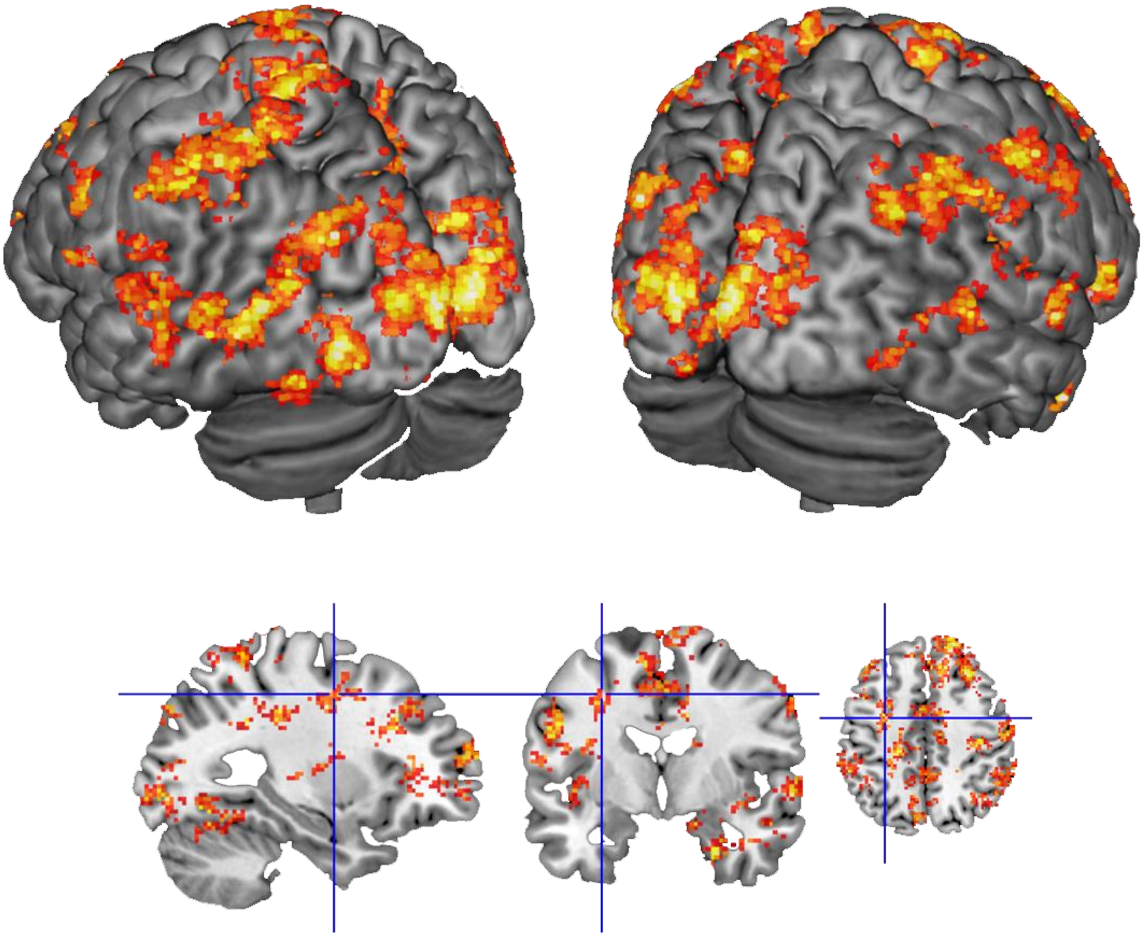


Figure 3: Cross-decoding results. Top: Significant cross-decoding from saccades to words in visual, temporal, and parietal regions. Bottom (left to right): Sagittal, coronal, and axial slices with crosshairs on the FEF. Brighter colors indicate higher classification accuracies.

Overlap between word decoding and cross-decoding (Figure 4) was observed in the inferior parietal lobule (angular gyrus and left supramarginal gyrus), precuneus, middle and superior occipital gyrus, posterior middle temporal gyrus, supplementary motor area, superior frontal gyrus, and superior frontal sulcus towards the precentral sulcus.

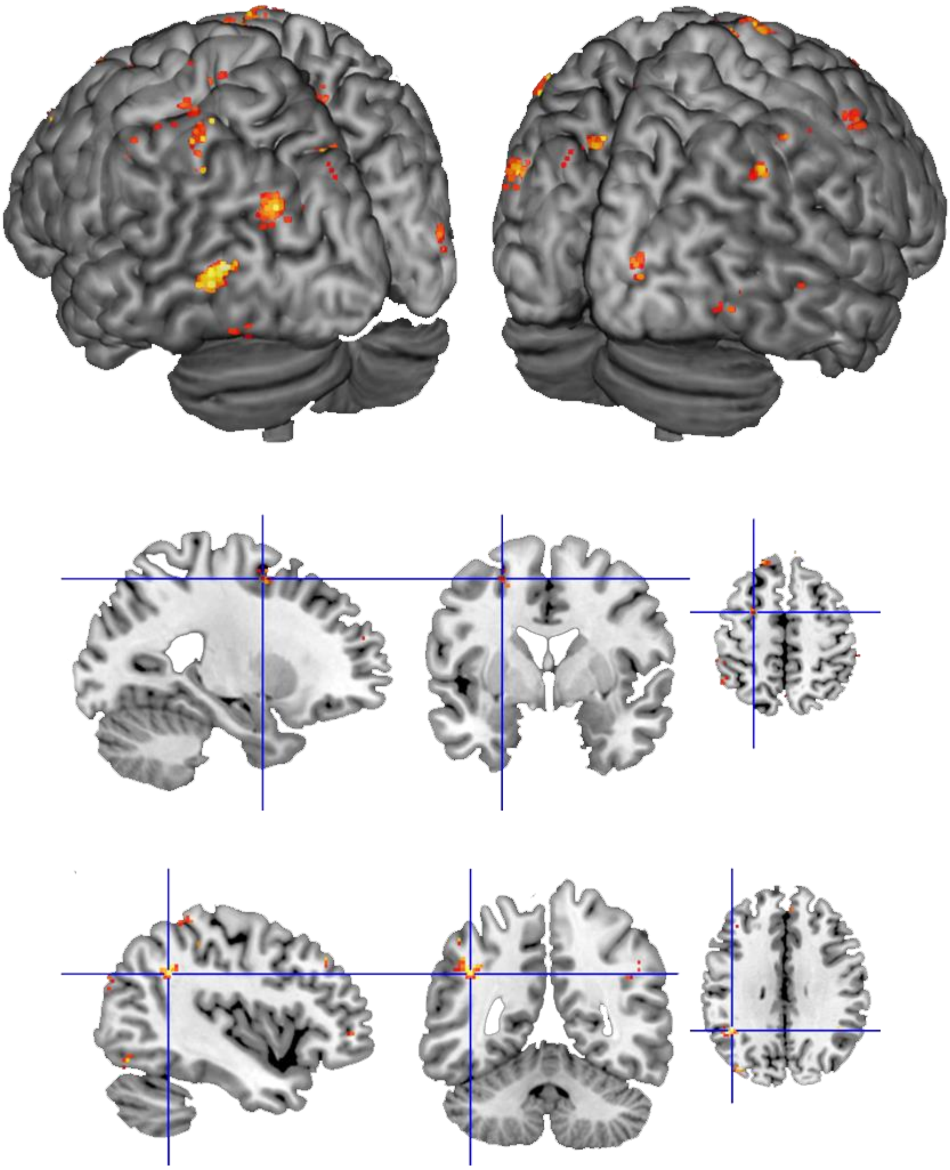


Figure 4: Overlap between cross-decoding and word decoding. The cortical surface mainly shows overlap in the middle occipital gyri, posterior inferior and middle temporal lobes (top panel). Overlap was also observed in the posterior superior frontal sulcus (middle panel) and the inferior parietal lobule (bottom panel). Brighter colors indicate higher classification accuracies.

Discussion

Previous behavioral and eye-tracking studies suggested that words with implicit spatial (up/down) associations can subtly influence performance on simple visual identification tasks involving up/down targets (Gozli et al., 2013). Up/down words have further been reported to increase the speed of subsequent saccadic eye movement initiation (Dudschig et al., 2013; Dunn, 2016), and to modulate the trajectories of concurrently performed saccades (Ostarek et al., in press). Here, we set out to test the hypothesis that semantic processing of up/down words in the absence of eye movements recruits the oculomotor system. In line with this hypothesis, we found that multi-voxel patterns in several key areas implicated in programming and executing eye movements allowed reliable decoding of up vs. down words and cross-decoding from saccades to words.

Two related lines of research make this finding particularly interesting. The first capitalizes on the insight that covert shifts of spatial attention have a largely shared neural substrate with covert spatial attention (Corbetta et al., 1998; Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Grosbras et al., 2005). This can be accounted for by the premotor theory of attention (Rizzolatti, Riggio, Dascola, & Umiltá, 1987) holding that the same neurons are active for saccades and covert shifts of attention, or by models assuming overlapping but distinct interacting cell populations (Petersen & Posner, 2012; Schafer & Moore, 2007; Thompson, Biscoe, & Sato, 2005). Distinguishing between the two is not possible in fMRI because of limitations in spatial resolution, but we can conclude that on a systems level, similarly to covert shifts of attention, semantic information about the typical location of an item is retrieved from the cortical network for goal-directed eye movements.

The second line of research is the work on the mental number line and the neural correlates of arithmetic processing that has established a strong link with the cortical systems for orientation in horizontal space. For instance, one study reported successful cross-decoding from left vs. right saccades to subtraction vs. addition (Knops et al., 2009). Furthermore, when trying to bisect numerical intervals patients with left neglect show an error pattern with a shift towards larger numbers (Zorzi, Priftis, & Umiltà, 2002), similar to the bisection of physical lines where a shift to the right relative to the midpoint was observed (Schenkenberg, Bradford, & Ajax, 1980). This had led to the view that cultural acquisitions, such as numerical cognition, recycle evolutionary older systems that perform useful computations for the new task at hand (Dehaene, 2005; Dehaene & Cohen, 2007).

In the context of these considerations, it thus seems reasonable to propose that a similar phenomenon occurs in the domain of language comprehension. In particular, our results indicate that the dorsal-frontal system for overt and covert shifts of attention is recycled to provide semantic information about the typical vertical spatial location of word referents. Previous studies have shown that explicit language about spatial relationships (Kemmerer & Tranel, 2003; Tranel & Kemmerer, 2004) and judgments about word referents' typical location (Quadflieg et al., 2011) recruit parietal areas involved in spatial cognition. The novel contribution of our data is that they implicate the orienting network with conceptual processing of implicitly spatial words in a semantic task orthogonal to the spatial domain.

Besides the spatial processing/oculomotor areas, information about typical location was present in temporal and prefrontal areas that are part of the 'core' semantic network (Binder et al., 2009). This suggests that access to semantic information about space relies on interactions between high-level areas typically associated with lexical-semantic processing on the one hand, and spatial processing/oculomotor areas on the other. Future studies benefiting from higher temporal resolution or using functional connectivity analysis could investigate the flow of information between these systems.

Acknowledgments

We would like to thank Paul Gaalman and José Marques for their help with the preparation of the MRI sequences and data collection.

References

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., … Shulman, G. L. (1998). A Common Network of Functional Areas for Attention and Eye Movements. *Neuron*, *21*(4), 761–773. https://doi.org/10.1016/S0896-6273(00)80593-0

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *Journal of Neuroscience*, *34*(1), 332–338.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. *Cerebral Cortex*, 25(9), 2584–2593.

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. *From Monkey Brain to Human Brain, Ed. S. Dehaene, J.-R. Duhamel, MD Hauser* & *G. Rizolatti*, 133–57.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384–398.

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012). From top to bottom: spatial shifts of attention caused by linguistic stimuli. *Cognitive Processing*, *13*(1), 151–154.

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. *PloS One*, *8*(2), e56872.

Dunn, B. M. (2016). *Which way is up? Grounded mental representations of space*. University of Glasgow.

Estes, Z., Verges, M., & Adelman, J. S. (2015). Words, objects, and locations: Perceptual matching explains spatial interference and facilitation. *Journal of Memory and Language*, *84*, 167–189.

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down object words orient attention to the objects' typical location. *Psychological Science*, *19*(2), 93–97.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage*, *26*(2), 471–479. https://doi.org/10.1016/j.neuroimage.2005.02.004

Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., & Binder, J. R. (2016). Heteromodal Cortical Areas Encode Sensory-Motor Features of Word Meaning. *Journal of Neuroscience*, *36*(38), 9763–9769. https://doi.org/10.1523/JNEUROSCI.4095-15.2016

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., … Seidenberg, M. S. (2016). Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. *Cerebral Cortex*, 26(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., & Binder, J. R. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia*, *76*, 17–26.

Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit spatial cues for visual attention. *Journal of Experimental Psychology: General*, *142*(4), 1028.

Grosbras, M.-H., Laird, A. R., & Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Human Brain Mapping*, *25*(1), 140–154. https://doi.org/10.1002/hbm.20145

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics*, *7*(1), 37–53. https://doi.org/10.1007/s12021-008-9041-y

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

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Kemmerer, D., & Tranel, D. (2003). A double dissociation between the meanings of action verbs and locative prepositions. *Neurocase*, *9*(5), 421–435.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650.

Knops, A., Thirion, B., Hubbard, E. M., Michel, V., & Dehaene, S. (2009). Recruitment of an Area Involved in Eye Movements During Mental Arithmetic. *Science*, *324*(5934), 1583–1585. https://doi.org/10.1126/science.1171599

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Lewis, G., & Poeppel, D. (2014). The role of visual representations during the lexical access of spoken words. *Brain and Language*, *134*, 1–10.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, *320*(5880), 1191–1195. https://doi.org/10.1126/science.1152876

Ostarek, M., & Huettig, F. (2017a). A Task-Dependent Causal Role for Low-Level Visual Processes in Spoken Word Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1215. https://doi.org/http://dx.doi.org/10.1037/xlm0000375

Ostarek, M., & Huettig, F. (2017b). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 499.

Ostarek, M., Ishag, A., Joosen, D., & Huettig, F. (in press). Saccade trajectories reveal dynamic interactions of semantic and spatial information during the processing of implicitly spatial words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/DOI: 10.1037/xlm0000536

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 579.

Paus, T. (1996). Location and function of the human frontal eye-field: A selective review. *Neuropsychologia*, *34*(6), 475–483. https://doi.org/10.1016/0028-3932(95)00134-4

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89.

Quadflieg, S., Etzel, J. A., Gazzola, V., Keysers, C., Schubert, T. W., Waiter, G. D., & Macrae, C. N. (2011). Puddles, parties, and professors: Linking word categorization to neural patterns of visuospatial coding. *Journal of Cognitive Neuroscience*, *23*(10), 2636–2649.

Lambon Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1), 31–40.

Schafer, R. J., & Moore, T. (2007). Attention governs action in the primate frontal eye field. *Neuron*, *56*(3), 541–551.

Schenkenberg, T., Bradford, D. C., & Ajax, E. T. (1980). Line bisection and unilateral visual neglect in patients with neurologic impairment. *Neurology*, *30*(5), 509–509.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage*, *65*, 69–82.

Thompson, K. G., Biscoe, K. L., & Sato, T. R. (2005). Neuronal basis of covert spatial attention in the frontal eye field. *Journal of Neuroscience*, *25*(41), 9479–9487.

Tranel, D., & Kemmerer, D. (2004). Neuroanatomical correlates of locative prepositions. *Cognitive Neuropsychology*, *21*(7), 719–749.

van Dam, W. O., Speed, L. J., Lai, V. T., Vigliocco, G., & Desai, R. H. (2017). Effects of motion speed in action representations. *Brain and Language*, *168*, 47–56.

Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., & Valero-Cabré, A. (2014). Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Frontiers in Integrative Neuroscience*, *8*. https://doi.org/10.3389/fnint.2014.00066

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, *96*, 222–229.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Brain damage: Neglect disrupts the mental number line. *Nature*, *417*(6885), 138–139. https://doi.org/10.1038/417138a

Appendix

Critical stimulus words, their spatial associations, and English translations

Stimuli	Association	Translation
afgrond	down	abyss
bodem	down	bottom
graf	down	tomb
gras	down	grass
kelder	down	cellar
metro	down	subway
plas	down	puddle
tapijt	down	carpet
tunnel	down	tunnel
vloer	down	floor
voet	down	foot
worm	down	worm
berg	up	mountain
dak	up	roof
komeet	up	comet
kroon	up	crown
palm	up	palm tree
plafond	up	ceiling
ster	up	star
top	up	top
toren	up	tower
vlieger	up	kite
vogel	up	bird
vuurwerk	up	fireworks

Chapter 7

MULTI-VOXEL PATTERN ANALYSIS REVEALS HIGH CONCEPTUAL FLEXIBILITY IN HIGH AND LOW LEVEL BRAIN SYSTEMS

Based on: Ostarek, M., van Paridon, J., Hagoort, P., & Huettig, F. (manuscript in preparation). Multi-voxel pattern analysis reveals high conceptual flexibility in high and low level brain systems.

Abstract

Conceptual processing is characterized by a striking degree of flexibility on the one hand, and a remarkable capacity for abstraction on the other. Using multi-voxel pattern analysis, we investigated how task demands shape semantic processing of concrete nouns, focusing particularly on size and animacy information. Representational similarity analysis revealed highly flexible processing of size and animacy in the left anterior temporal lobe and a network of primary sensory and association areas. Surprisingly, we found no evidence for task invariant processing of size or animacy anywhere in the brain. To further probe flexibility, split half analyses comparing within and between task pattern consistency similarly showed task-specific patterns in occipital areas. They also revealed some evidence for task invariant processing in the intraparietal sulcus and surrounding cortex (including the angular gyrus). On balance, these results suggest a high level of flexibility: Whereas a distributed network encompassing the anterior temporal lobe and sensory areas seem to flexibly adapt to task demands to provide contextually relevant information, we obtained only limited evidence for task invariant semantic processing in a parietal region in and around the intraparietal sulcus.

Introduction

The nature of the human conceptual system remains a topic of intense debate and several central questions remain unclear. One particularly tricky issue is that conceptual processing relies on two core capacities that, at first sight, appear to be incompatible: On the one hand, theorists have long stressed the remarkable human ability to abstract away from the perceptual details of any particular exemplar to build abstract conceptual representations (e.g., Fodor, 1975). For instance, mountains can be experienced from various standpoints and vary widely in their features (i.e., flora, fauna, steepness), most of which need to be abstracted away from in order to arrive at a representation that fits the concept 'mountain' in general, as opposed to any specific exemplar. On the other hand, a defining feature of the human mind is its high degree of flexibility (e.g., Barsalou, 1993), allowing us to grasp immediately what a *miniature mountain* is, even if it requires us to override the aspect of our conceptual knowledge that mountains are usually very big.

Many behavioral and electroencephalography studies have provided evidence for conceptual flexibility (Nieuwland & Van Berkum, 2006; Ostarek & Huettig, 2017; Rommers, Meyer, & Huettig, 2013; Van Berkum, Brown, & Hagoort, 1999; Yee, Ahmed, & Thompson-Schill, 2012; Yee & Thompson-Schill, 2016) by showing that contextual demands can strongly influence which semantic features are retrieved and which processing systems are involved. To date, relatively little high spatial resolution neuroimaging work has addressed flexibility in language comprehension. This would be crucial for a better understanding of the neuro-cognitive mechanisms underlying this striking human capacity. Univariate functional magnetic resonance imaging (fMRI) studies have shown that task relevance of motor or visual features of words leads to higher activation in brain regions involved in movement planning and visual perception (Hoenig, Sim, Bochev, Herrnberger, & Kiefer, 2008; van Dam, van Dijk, Bekkering, & Rueschemeyer, 2012), and to increased functional connectivity between auditory cortex and these areas (Van Dam, Van Dongen, Bekkering, & Rueschemeyer, 2012). Hoenig et al. (2008) reported stronger activation in posterior temporal cortex for words referring to artefacts when the task probed visual features and stronger activation in precentral and parietal cortex for words referring to natural objects when the task probed action features. Van Dam et al. (2012) contrasted activation levels evoked by words with motor and color associations (e.g., *tennis ball*) in a color-verification task vs. an action-verification task and observed increased activation in the left inferior parietal lobule (which is associated with motor planning) when the task focused on action features, but did not find context-related differences regarding processing of color features.

Due to its focus on net between-condition differences in activation in contiguous clusters of voxels, univariate MRI has clear limitations in terms of sensitivity; for instance, it is blind to information in distributed population codes if the activation level of the population as a whole does not change (Kriegeskorte & Kievit, 2013). Multi-voxel pattern analysis (MVPA) capitalizes on machine learning methods to probe the informational content present in fine-grained patterns of activation (Norman, Polyn, Detre, & Haxby, 2006).

Sophisticated MVPA studies recently explored contextual effects on object recognition processes using fMRI (Bracci, Daniels, & Op de Beeck, 2017; Harel, Kravitz, & Baker, 2014; Nastase et al., 2017; Peelen, Fei-Fei, & Kastner, 2009) and provided some clues about cortical regions that could be engaged in flexible conceptual processing. In particular, representational similarity analysis (RSA) has proven to be a useful tool to test how the representation of a dimension of interest changes according to task-demands. Bracci et al. (2017), for instance, demonstrated that fronto-parietal regions show dynamic changes in the content they represent as a function of task-relevance. By contrast, occipital and temporal regions tended to respond to object features regardless of whether or not they were task-relevant. Recent studies using videos of naturalistic events as stimuli also observed similar context-effects in ventral occipital-temporal cortex (Çukur, Nishimoto, Huth, & Gallant, 2013; Nastase et al., 2017). The emerging picture thus is that fronto-parietal areas, and to a lesser extent high-level visual areas, adapt to contextual demands by more explicitly representing task-relevant features, whereas early visual regions seem to process stimulus features regardless of context.

We reasoned that in visual object recognition basic visual processing areas are mainly bound to bottom-up analysis of object features. Our main hypothesis was that in the language domain, by contrast, visual processing areas are available as a pool of information that can be flexibly sampled from during semantic processing. Based on this, we expected to find effects of flexibility in the visual cortex, including early visual areas.

We used searchlight RSA to probe language processing with respect to the conceptual dimensions of size and animacy. These dimensions are particularly interesting because they are typically assumed to be very central aspects of our conceptual knowledge due to their behavioral relevance. Moreover, thanks to extensive neuroimaging work, it is relatively well-understood how they map onto the cortex (Martin, 2007; Martin, Wiggs, Ungerleider, & Haxby, 1996). In particular, animacy is strongly associated with the fusiform gyrus and the posterior middle and superior temporal gyri (Chao, Haxby, & Martin, 1999) which are implicated with biological (Bonda, Petrides, Ostry, & Evans, 1996) and non-biological motion perception (Martin et al., 1996). Real-world object size is associated with the lateral occipital cortex, inferior temporal lobe, and parahippocampal cortex (Konkle & Caramazza, 2013; Konkle & Oliva, 2012). Our study investigated how the task relevance of size and animacy information impacts processing in these regions.

A recent study by Borghesani et al. (2016) showed that it is possible to decode real world object size derived from written words from early visual cortex and animacy from anterior temporal areas. Based on this, these authors suggested that there is a posterior-to-anterior gradient of semantic information: Whereas posterior areas of the ventral stream may provide information about visual properties, more anterior areas may be sensitive to increasingly abstract information (such as semantic category membership), in line with the view that the anterior temporal lobe (ATL) is a high-level conceptual convergence zone (Lambon Ralph, Sage, Jones, & Mayberry, 2010). According to this view, the ATL is expected to represent abstract conceptual knowledge invariant to the details of a particular exemplar or situation (Patterson, Nestor, & Rogers, 2007; Peelen & Caramazza, 2012).

We predicted that the extent to which early visual regions carry information about a conceptual feature are not only determined by the nature of the feature but also largely by its relevance in a given context. Thus we expected that information about size and animacy would be retrieved from early visual areas when it is relevant for the task. The view that perceptual properties are retrieved from sensory cortex and then converge in the ATL (Coutanche & Thompson-Schill, 2014) further predicts that similar flexibility effects in the visual cortex should be evident there.

Method

Participants

27 healthy native Dutch speakers participated in the study, six of which were excluded due to a technical problem and six because they did not comply with the task in at least one run. All of the analyses were conducted on the remaining 15 participants (all right-handed, 10 female, mean age 23.6). We had ethical approval from the local CMO ethics board for Research involving Human Subjects in Arnhem-Nijmegen (dossier CMO 2014/288).

Stimuli, Design and Procedure

Participants were informed that they were going to see and hear Dutch words and that they had to perform one of two tasks across runs: Either they were asked to indicate with a left or right index finger button-press (counterbalanced across runs) whether the word referred to something that would fit in a normal shoe box or not (size task), or they were asked to decide whether the word referred to an animal or not (animal task). Participants performed two runs per task in alternation; the order of tasks was counterbalanced across participants. The stimuli (listed in the Appendix) were spoken and written versions of 24 concrete words with varying degrees of animacy (ranging from 1.0 for "key" to 8.7 for "monkey") and real world object size (ranging from 1.33 for "button" to 8.6 for whale; henceforth simply "size"). Each trial lasted 4.5 s. In spoken word trials, the central fixation cross was displayed throughout. In the written word trials, a word was presented centrally for 600ms and was then replaced by the fixation cross for the remaining 3900ms. An independent group of 20 participants had rated the words on animacy and size on a 9-point Likert scale. They were presented with written versions of all the words twice and were asked to evaluate 1) how sentient and alive the word referents are and 2) how big they are. The instructions encouraged them to use the whole spectrum from one to nine. The results showed that the two dimensions were orthogonal (Pearson and

Spearman rank correlation: p>0.05). The stimuli classes (animal vs. no animal; big vs. small) were matched on frequency (Keuleers, Brysbaert, & New, 2010), age of acquisition (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014), concreteness (Brysbaert et al., 2014), number of letters, number of phonemes, and number of syllables (all p-values > 0.4). Each word was presented four times per run (twice in written form and twice in spoken form), resulting in 16 repetitions per word and 384 trials in total (96 trials per run). We decided for mixed auditory and visual presentation of the words in order to focus on the semantic level and avoid low-level effects related to word-form processing, phonological encoding, etc. This was particularly important for the word decoding (split half) analyses described below as words are likely discriminable due to word-form features (besides differences in meaning).

	Frequency	Age of Acquisition	Concrete-ness	Letters	Phonemes	Syllables
animate	23.99 (48.06)	5.80 (1.52)	4.90 (0.09)	5.42 (1.56)	5.0 (1.71)	1.58 (0.51)
inanimate	25.53 (29.52)	5.46 (1.23)	4.79 (0.19)	5.25 (1.22)	4.50 (1.00)	1.67 (0.49)
big	31.68 (49.90)	5.77 (1.49)	4.86 (0.11)	5.33 (1.23)	4.75 (1.36)	1.67 (0.49)
small	16.64 (24.45)	5.47 (1.26)	4.84 (0.11)	5.50 (1.38)	4.83 (1.40)	1.58 (0.51)

Table 1: Stimulus characteristics. Mean and standard deviations are shown for words denoting animate, inanimate, big, and small referents.

Data collection and preprocessing

Data were collected at a 3T Siemens Prisma scanner with a 32-channel head coil at the Donders Centre for Cognitive Neuroimaging. Four functional runs were acquired in which participants performed the language task. Two additional runs were acquired where participants saw morphed pictures of cats and dogs, which were not used for the current analyses. Functional images were acquired using a multi-band sequence (68 slices, acceleration factor = 4, FOV = 210mmx210mm, spatial resolution = 2x2x2 mm, TR = 1.5 s, TE = 39.60 ms, flip angle = 75°). A structural scan was acquired to obtain high-resolution anatomical images (MPRAGE sequence, 192 volumes, spatial resolution = 1mm).

Preprocessing was carried out in SPM12. Functional images were slice-time corrected and spatially realigned to the first volume of the first run, co-registered to the structural scan, and then normalized to MNI space.

Multi-voxel pattern analysis

All of the analyses below were conducted using spherical searchlight analysis (Kriegeskorte, Goebel, & Bandettini, 2006). GLMs were used to compute t-maps that served as the input for MVPA. For the RSA analyses, one t-map was computed per word per task. For the word-decoding split-half analyses, one t-map was computed per split. The whole brain was searched for the information in question using searchlights with a radius of 3 voxels (ca. 125 voxels per searchlight) such that each voxel was the center of the searchlight once, following the implementation of Kriegeskorte et al. (2006) in PyMVPA (Hanke et al., 2009).

RSA

One hypothesis was that task-related changes in conceptual processing arise because task-relevant semantic features are more explicitly represented in regions involved in processing these features. Our stimuli varied orthogonally in terms of size and animacy, allowing us to probe changes in the way these two conceptual dimensions are reflected in multi-voxel patterns depending on task demands. Model dissimilarity matrices (RDMs) were constructed for these two dimensions based on behavioral ratings collected from an independent group of 20 participants on a 9-point Likert scale. In particular, the model RDMs reflected the distance in mean ratings for all word pairs. Searchlight RSA was used to identify areas that carry information about size and/or animacy by spearman correlating the model RDMs with neural RDMs per searchlight (the Spearman rank correlation coefficient per searchlight is mapped onto the central voxel). This was done separately for the data obtained in the size-task and the animal-task.

Permutation tests were used for statistical inference. Specifically, the relevant analysis was repeated 100 times per participant with randomly shuffled t-map labels. A group-level empirical null-distribution was generated for each voxel by randomly picking one of 100 correlation maps per participant and averaging these randomly chosen maps 100 000 times to create an empirical group-level null distribution. The voxel-wise threshold was set to 0.005 (as recommended by Stelzer, Chen, and Turner (2013), meaning that a voxel only surpasses the threshold if its correlation coefficient is higher than 99.5% of the coefficients in the null-distribution in that location. The alpha-level at the cluster level was set to 0.05 and false discovery rate was controlled with the Benjamini and Hochberg method (Benjamini & Hochberg, 1995). We used a cluster-size cutoff of 20 voxels.

To assess flexibility statistically, per-voxel correlation coefficients reflecting the correlations of RDMs based on patterns in searchlight spheres with the size RDM in the animal-task were subtracted from those in the size-task. Equivalently for flexible processing of animacy, per-voxel correlation coefficients reflecting the correlations of neural RDMs with the animacy RDM in the size-task were subtracted from those in the animal-task.

Invariance in terms of size and animacy was operationalized as the reverse of flexibility. Specifically, the correlation maps containing difference scores (correlation values of a model RDM with neural RDMs in the corresponding task minus correlation values of the model RDM with neural RDMs in the other task) were squared and their sign was flipped, such that values close to zero indicated invariance.

Word decoding

We sought to map out those brain regions whose multi-voxel patterns differentiate between the 24 concepts we used and test the effect of the task context manipulation on word-specific activation patterns in two independent splits of the data. To that end, we used a correlation-based split-half analysis which can be thought of as a linear classifier (Harel et al., 2014). This analysis is based on the rationale that in regions involved in semantic processing of the words patterns evoked by the same word at two independent points in time should be more similar than patterns evoked by different words.

For the within-task word decoding analysis, per participant GLMs were used to produce one t-map per word for the first half and the second half of the experiment, such that in both halves participants had performed one size-task run and one animal-task run. The patterns corresponding to all items in the first half were then correlated with all items in the second half, resulting in a matrix of Pearson correlation coefficients in which the diagonal cells represent the correlation of a word and itself (in two independent splits) and the off-diagonal cells represent the correlation of each word with all the other words. The mean value of all off-diagonal cells was subtracted from the mean diagonal value to obtain an index of word decoding. As activity patterns in both splits are informed by data from the same tasks, this analysis provides an index of within-task word decoding.

Conversely, for the between-task word decoding analysis, one t-map was created per word per task, such that one split contained activity patterns of all words in the size-task and the other split contained activity patterns of the same words in the animal-task.

To quantify flexibility in the multi-voxel patterns evoked by the same words in different contexts, within-task pattern consistency was contrasted with between-task pattern consistency. This was achieved by subtracting the voxel-wise correlation coefficients obtained in the between-task decoding analysis from those obtained in the within-task decoding analysis per participant to produce a difference score reflecting to what extent patterns activated by the same words changed as a function of task demands. The resulting indices of flexibility (within minus between-task word decoding indices) were tested for statistical significance with a permutation test based on permuted maps obtained by similarly subtracting all between-task decoding permutation maps from all within-task decoding maps per participant. Conversely, invariance to task-demands was assessed by testing for clusters whose difference scores were closer to zero than is to be expected by chance.

Results

Behavioral results

Participants performed the animal task highly accurately (M=99.3%, SD=0.08). According to our classification of items as 'big' or 'small', accuracies were lower in the size task (M=93.8, SD=0.24, z=3.05, p=0.002). Note that the size judgement is somewhat subjective and individuals do not always agree about the size of word referents. Reaction times were shorter in the animal task (M=832 ms, SD=215 ms) compared to the size task (M=940 ms, SD=260 ms; estimate=55.65, SE=8.77, t=6.35).

RSA analyses

Size decoding

To investigate how the processing of size information changes when it is task-relevant vs. irrelevant, a size RDM based on behavioral ratings of real world object size was correlated with neural RDMs across the brain in response to the same set of words encountered in the size task and the animal task. When we directly tested for stronger representation of the size dimension in the size task relative to the animal task, we observed significant clusters in the left lingual gyrus, posterior fusiform gyrus, middle occipital gyrus, posterior, middle temporal gyrus, anterior inferior temporal gyrus, bilateral precuneus, superior parietal lobule, bilateral premotor, motor, supplementary motor area, right inferior frontal gyrus, and the cerebellum (see Figure 1).

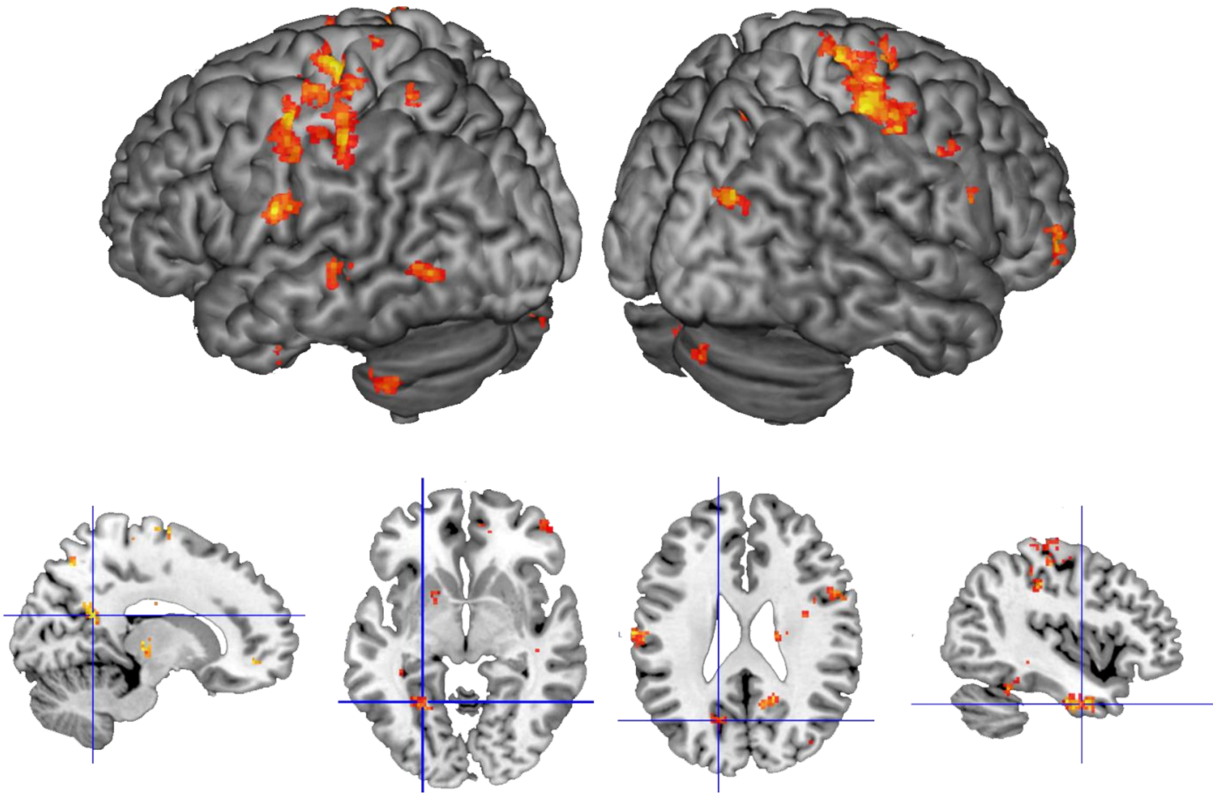


Figure 1: Flexible size processing. Top: Strong effects of flexibility in bilateral precentral and postcentral cortices, cerebellum, posterior temporal lobe, and middle occipital cortex. Bottom (from left to right): Crosshairs on right precuneus, left lingual gyrus, left cuneus, left anterior temporal lobe. Brighter colors reflect higher correlations.

This suggests that size information was flexibly retrieved from posterior brain regions involved in visual perception that interacted with higher-level areas, such as the anterior temporal lobe. Moreover, the effects in motor and premotor areas indicate that participants engaged in item-specific motor imagery, likely related to the attempts to fit the word referents in a mental shoe box that differ depending on the typical size of the respective word referents.

In line with this, in the size task, the size RDM showed significant correlations with neural RDMs in a distributed set of areas including early and higher-level visual areas, bilateral temporal areas, motor and premotor cortex, parietal areas, and the cerebellum (see Figure 2).

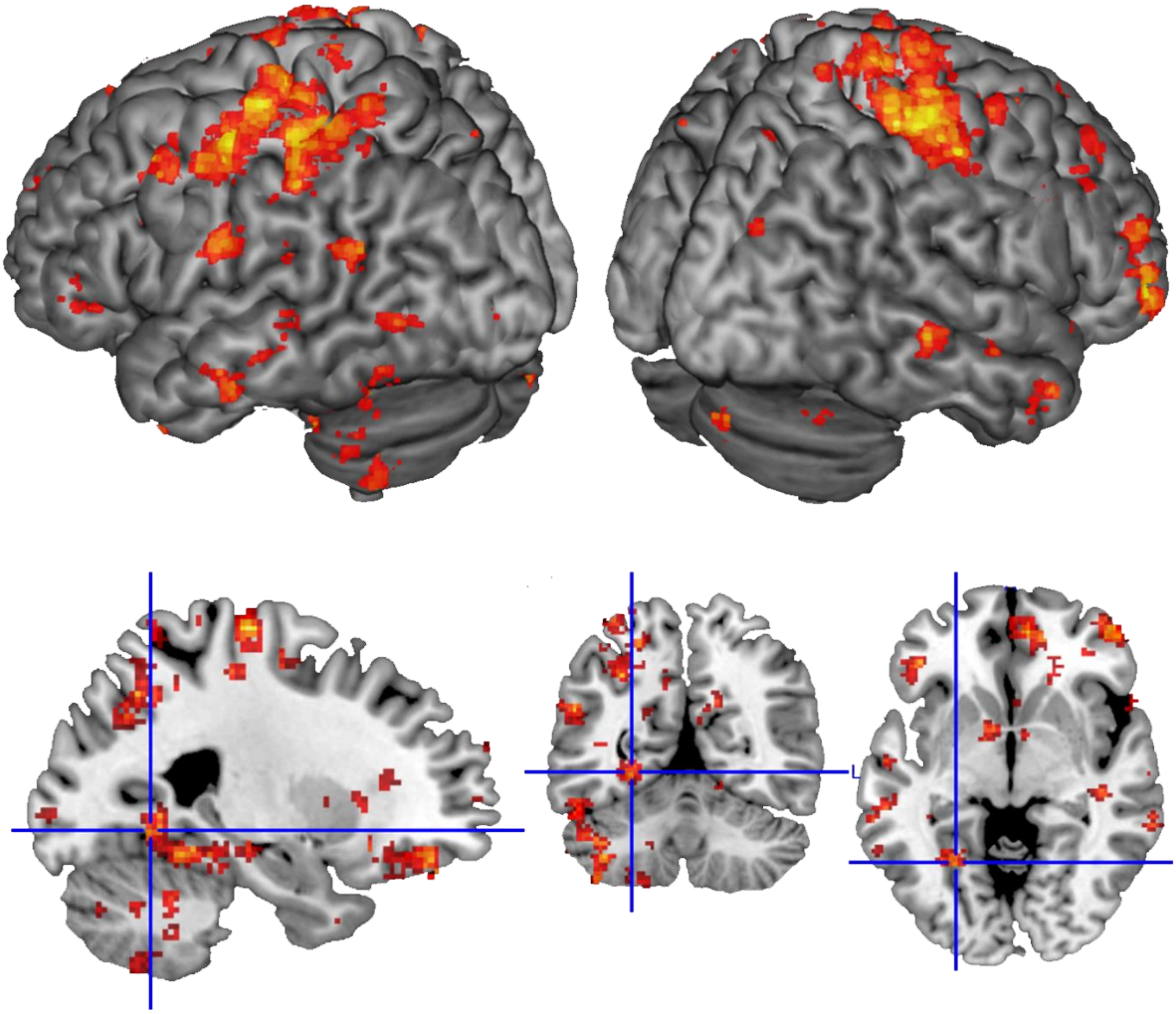


Figure 2: Size decoding in the size task. Top: A distributed network of temporal, parietal and frontal areas. Bottom (from left to right): Sagittal, coronal, and axial slice with crosshairs on the lingual gyrus. Brighter colors reflect higher correlations.

In the animal task on the other hand, none of these areas showed significant correlations with the size RDM. Surprisingly, some prefrontal areas and the right inferior parietal lobule which were not observed in the size task showed significant correlations (see Figure 3).

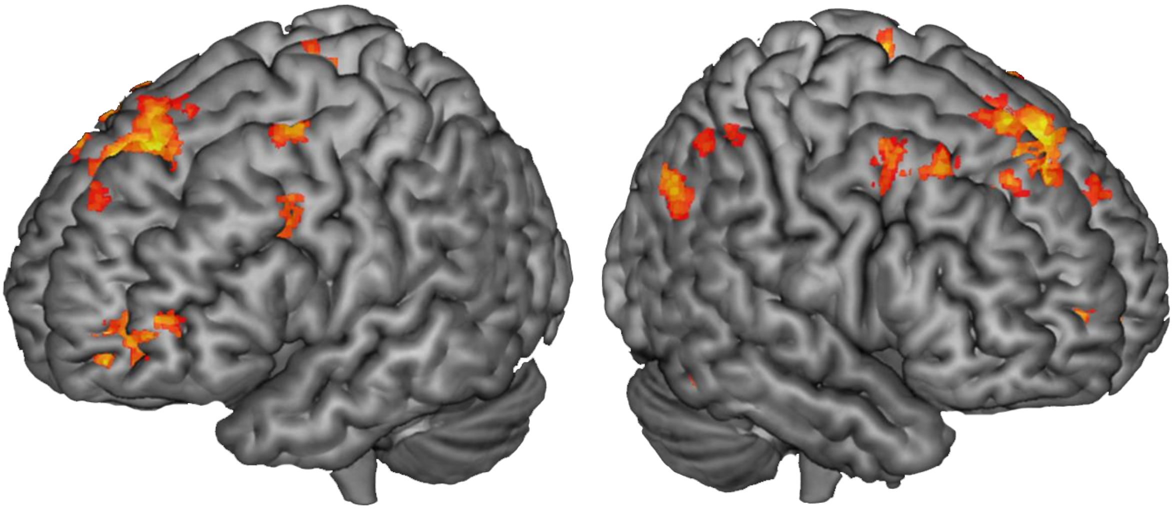


Figure 3: Size decoding in the animal task. Fewer areas carried size information, positive results were found in prefrontal cortex and right inferior parietal lobule. Brighter colors reflect higher correlations.

There was only one overlapping cluster (23 voxels) located in the left medial superior frontal gyrus that was significant both in the size and animal task, suggesting that size was processed highly flexibly according to task demands. Even though there were virtually no searchlights whose neural RDM significantly correlated with the size RDM in both tasks, in principle it is still possible that some of them are statistically invariant to task demands. To test this, we ran a permutation analysis that tests for clusters of searchlights in which the difference of size decoding in the size task vs. the animal task is closer to zero than expected by chance based on the empirical null distribution. No such cluster was found.

Animacy decoding

Equivalent analyses were carried out to probe flexible processing of information about animacy. Clusters showing stronger correlations with the animacy RDM in the animal task compared to the size task were found in the left calcarine sulcus, left posterior fusiform gyrus, bilateral precuneus (extending into the cuneus on the right), bilateral middle temporal lobe, right superior temporal lobe, bilateral premotor cortex, left cerebellum, and the left middle/superior temporal pole (see Figure 4).

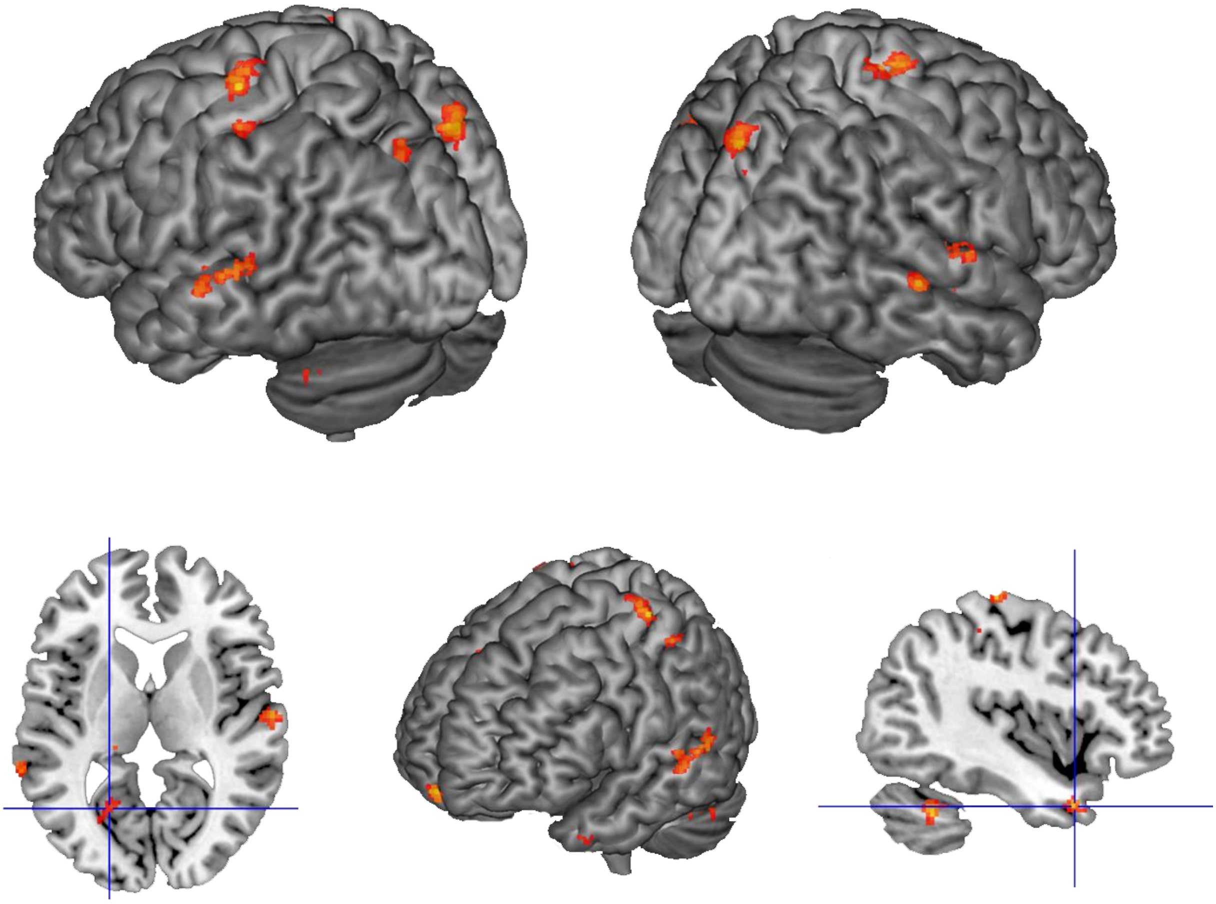


Figure 4: Flexible animacy processing. Top: Bilateral superior and middle temporal areas associated with animacy, precuneus, and premotor cortex. Bottom (from left to right): Axial slice with crosshairs on primary visual cortex; More frontal view showing an additional cluster in orbitofrontal cortex and ATL; Sagittal slice with crosshairs on ventral ATL. Brighter colors reflect higher correlations.

Despite these strong effects of flexibility, decoding of information about animacy was relatively weak both in the animal task and the size task. At the initial voxel-wise threshold (0.005), in the animal task significant clusters were observed in the left precentral gyrus, postcentral gyrus, superior parietal lobule, and cerebellum, whereas in the size task small clusters were only found in the anterior cingulate and superior frontal gyrus. At a slightly more lenient voxel-wise threshold (0.01; displayed below in Figure 5), in the animal task additional clusters emerged in the right calcarine sulcus, left precuneus/superior occipital gyrus, left middle temporal pole, and the medial prefrontal cortex.

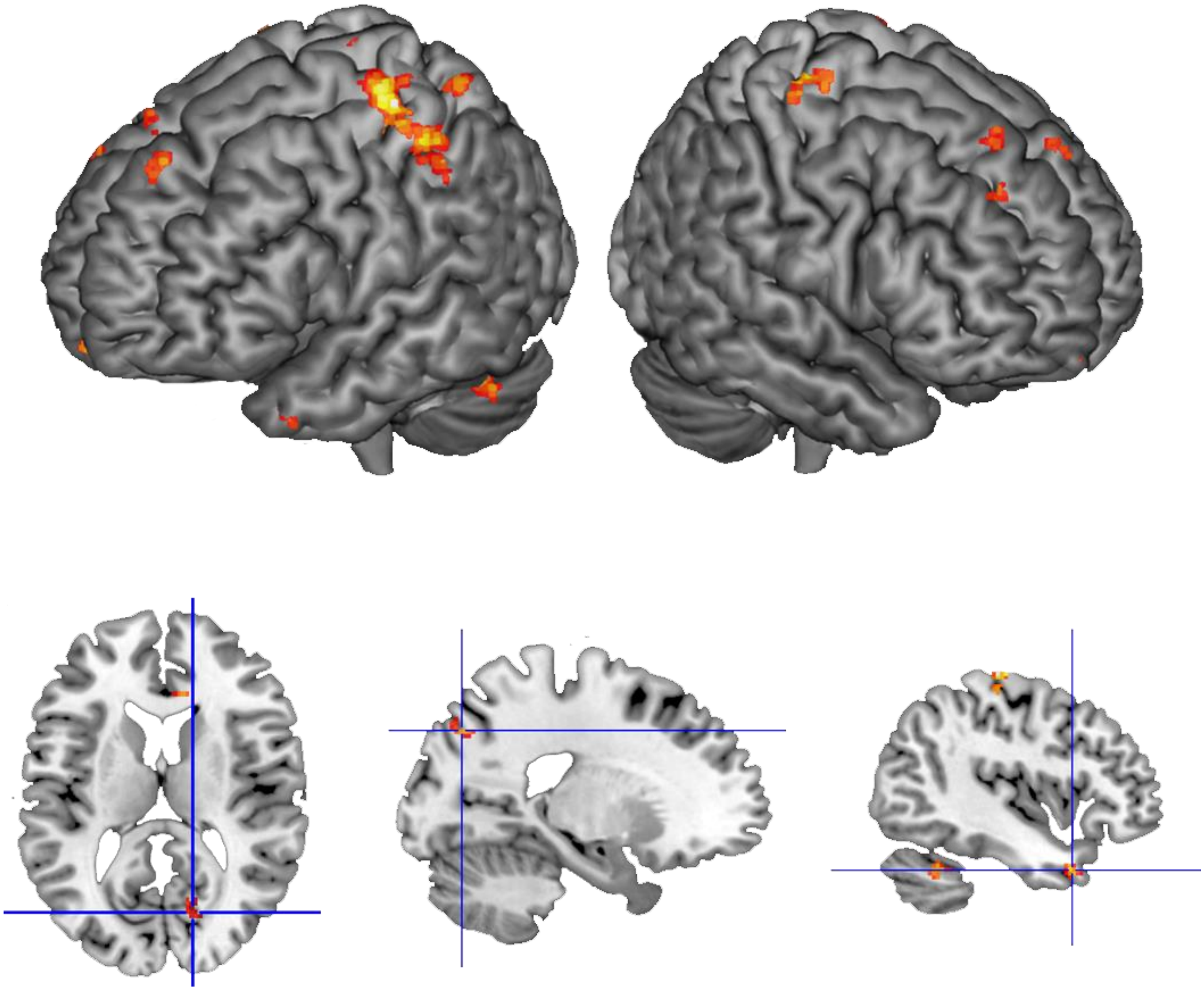


Figure 5: Animacy decoding in the animal task. Top: Animacy information on the cortical surface. Bottom (from left to right): Axial slice with crosshairs on the right primary visual cortex; Sagittal slice with crosshairs on the left precuneus; Sagittal slice with crosshairs on the left ATL. Brighter colors reflect higher correlations.

In the size task, one additional cluster was observed in the left angular gyrus (see Figure 6).

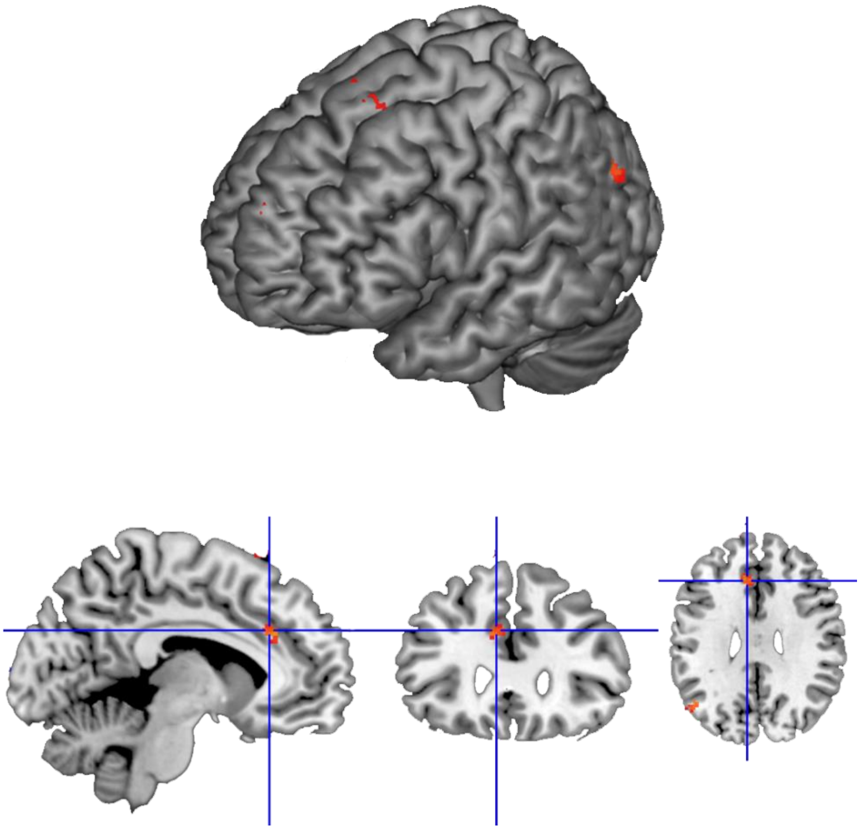


Figure 6: Animacy decoding in the size task. Top: The left cortical surface shows one cluster in the left angular gyrus and one in the superior frontal gyrus. Bottom (from left to right): Sagittal, coronal, and axial slice with crosshairs on the anterior cingulate cortex. Brighter colors reflect higher correlations.

There was no overlap of significant clusters in the animal task and the size task. Furthermore, the analysis of invariance revealed only a small cluster in the left lingual gyrus and right anterior cingulate gyrus that did not overlap with any of the clusters that were found to provide information about animacy in the animal or size task.

These results suggest that multi-voxel patterns reflect information about animacy highly flexibly depending on task requirements. Strikingly, we did not find evidence for task-invariant representation of animacy in any brain region, consistent with the results obtained in the size decoding analyses above.

Within- and between-task word decoding

The previous pattern similarity analyses point to strikingly flexible processing of size and animacy information. Given that both are likely to be relatively dominant and behaviorally important features this raises the question of whether the way word meanings are processed changes entirely from one context to the other. Indeed, some authors have proposed that there are no conceptual cores and that word meanings are constructed in an inherently context-dependent manner (Evans, 2009; Lebois, Wilson-Mendenhall, & Barsalou, 2015; McCloskey & Glucksberg, 1978). On the opposite end of the possibility spectrum, an alternative is that core word meanings are invariant and additional conceptual information is provided optionally when it is task-relevant. To address this question, we sought to investigate how flexible vs. stable patterns are in response to individual words within tasks and between tasks. Previous studies demonstrated that it is possible to decode a limited number of individual concepts (Correia et al., 2014; Zinszer, Anderson, Kang, Wheatley, & Raizada, 2016). However, to our knowledge no study has looked into the effect of task demands on the multi-voxel patterns related to individual words.

To obtain an index of within-task word decoding, we used a split-half analysis where each split contains data from one size-task run and one animal-task run. For each searchlight, an index of word decoding was calculated by subtracting correlations between different words from the correlations between the same words in the two splits. Permutation tests revealed significant clusters in the left intraparietal sulcus, left superior and inferior parietal cortex (angular gyrus and supramarginal gyrus), bilateral middle frontal gyrus, right superior orbito-frontal cortex, left middle temporal lobe, right precuneus, bilateral middle occipital gyrus, and primary visual cortex (see Figure 7).

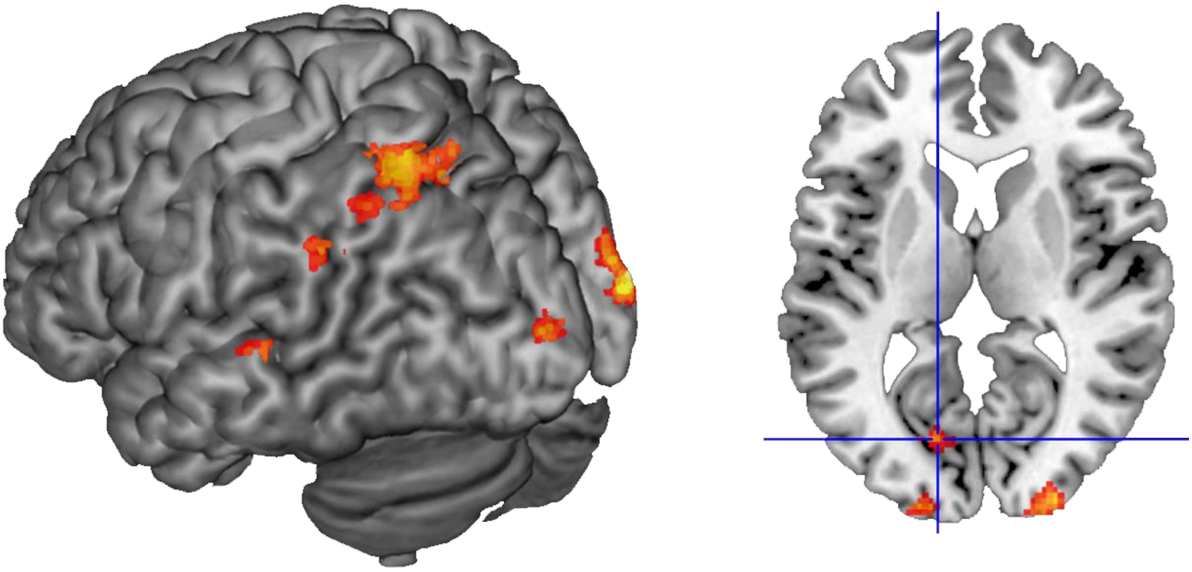


Figure 7: Within-task word decoding. Left: Clusters on the cortical surface. Right: Axial slice with crosshairs on the additional cluster in left primary visual cortex. Brighter colors reflect higher word decoding indices.

The equivalent analysis for between-task word decoding used both size-task runs as one split and the animal-task runs as the other. As such, it can only yield significant results if word-specific patterns are similar across tasks. It revealed a distributed set of significant clusters that only partly overlapped with the ones obtained in the within-task word decoding analysis (see Figure 8).

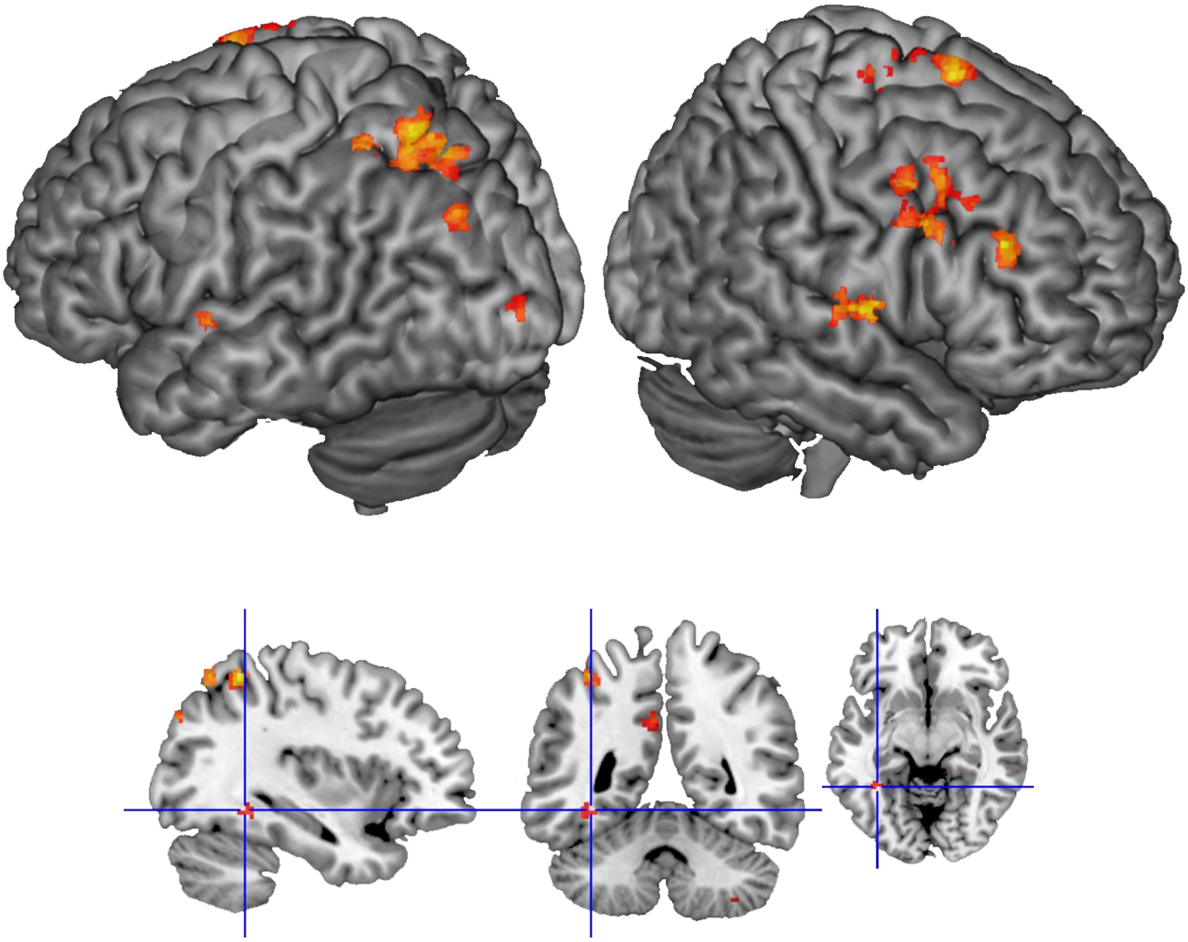


Figure 8: Between-tasks word decoding. Top: Clusters on the cortical surface. Bottom (from left to right): Sagittal, coronal, and axial slices with crosshairs on the left fusiform gyrus. Brighter colors reflect higher word decoding indices.

Additional clusters were located in the bilateral superior temporal gyri, supplementary motor area, precentral gyrus, middle frontal gyrus, left fusiform gyrus, and precuneus. Overlap was observed in the left intraparietal sulcus, inferior and superior parietal cortex, bilateral middle frontal gyrus, right precuneus, and left middle occipital gyrus (see Figure 9).

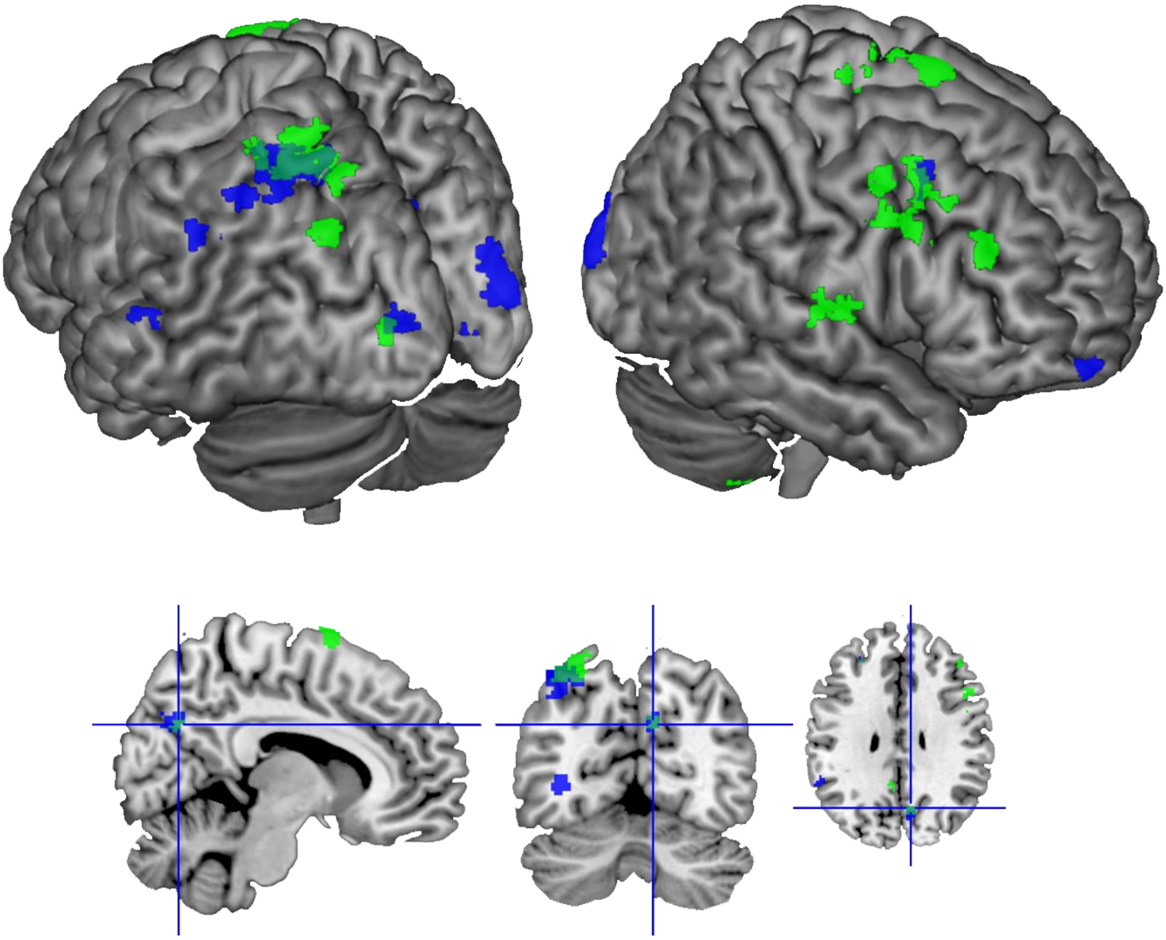


Figure 9: Within-task (blue) and between-task (green) decoding. Top: Overlap on the cortical surface was found around the left intraparietal sulcus, in bilateral middle frontal gyrus, right precuneus, and left middle occipital gyrus. Bottom (from left to right): Sagittal, coronal, and axial slice with crosshairs on the right precuneus.

Overlap in parietal and prefrontal regions suggests that in these regions word-specific patterns were similar in both tasks. To assess statistically whether patterns were indeed comparable within- and between-tasks, an index of flexibility vs. invariance was computed by subtracting the between-task word decoding index from the within-task word decoding index per searchlight. Invariance was operationalized as regions that show difference scores closer to zero than expected by chance and that overlap with a region with significant within or between-task word decoding[8]. The only cluster

---

[8] Additional regions that showed no evidence of word-level decoding could simply be unrelated to word processing.

displaying these characteristics was located in and around the left intraparietal sulcus (see Figure 10). We observed the same result in an exploratory follow-up analysis with a slightly more lenient voxel-wise threshold of 0.01, displayed below.

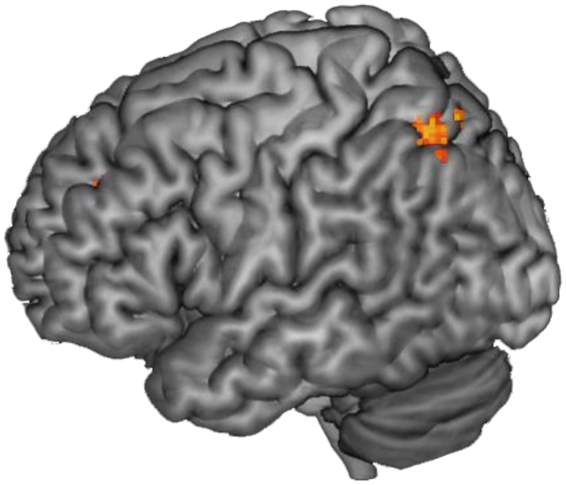


Figure 10: Invariant word processing in the left intraparietal sulcus and surrounding cortex. Brighter colors reflect higher task invariance indices.

This suggests at least some task-related changes in word-level activity patterns in the remaining areas that showed significant within-task decoding. To directly probe flexible word processing, we performed a second permutation test on the flexibility indices we obtained by computing within-task word decoding indices minus between-task word decoding indices, now testing for significantly positive values. As such it addresses directly whether there are any searchlights showing significantly higher within than between-task word decoding. No clusters were significant at a voxel-wise threshold of 0.005, but when it was reduced to 0.01 (as above), it revealed clusters in the right middle occipital and right superior occipital gyri overlapping with the clusters found in the within-task (but not between-task) word decoding analysis, one cluster on the inferior surface of the left posterior fusiform gyrus extending slightly into the cerebellum, and one cluster in the right ventral anterior inferior temporal lobe (see Figure 11).

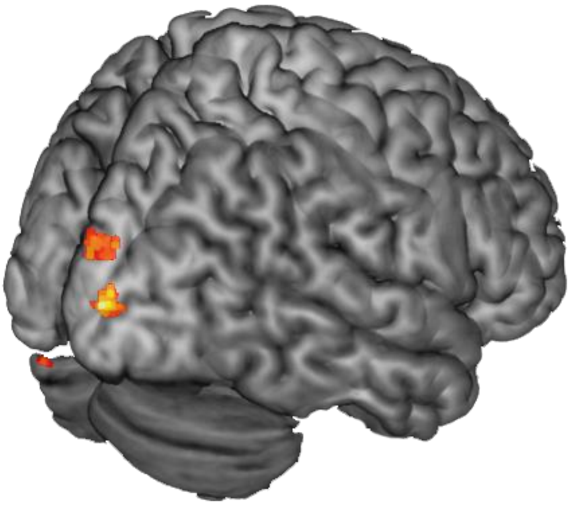


Figure 11: Flexible word processing in the occipital lobe. Brighter colors reflect higher flexibility indices.

Discussion

In this study, we used multi-voxel pattern analysis in combination with representational similarity analysis to investigate how task demands affect semantic processing. By orthogonally manipulating size and animacy, we were able to study how these conceptual dimensions are processed when they are relevant vs. irrelevant for the task at hand. Our results suggest that sensory processing systems (including early visual cortex) and the anterior temporal lobe are the neural substrate for flexible conceptual processing. A direct comparison of word specific within-task and between-task activity patterns provided evidence for context-invariant semantic processing in the intraparietal sulcus and surrounding cortex.

Effects of flexibility in visual and motor cortex

Size information was strongly present in visual areas including the lingual gyrus, posterior fusiform gyrus, middle occipital gyrus, and precuneus, but only when it was task-relevant. The finding that even early visual areas were involved in processing conceptual size information suggests that in the size task participants recruited these areas in a task-driven manner to determine the word referents' typical size. Borghesani et al. (2016) reported significant decoding of real-world object size from primary visual cortex in the context of passive reading with occasional semantic similarity judgement trials. In contrast to these results, we were unable to decode objects size from any occipital areas when the dimension was not task-relevant. This is somewhat surprising given that our approaches were largely similar in terms of number of items, number of word repetitions, and analysis. One explanation could be that our searchlights were too small and size information is more broadly distributed in early visual areas. However, an exploratory follow-up analysis using the whole of V1 as region of interest did not yield a statistically significant result either. Therefore, based on our data we suggest that size information can be flexibly retrieved from early visual cortex, but is not activated automatically.

Similar to the dimension of size, animacy was more strongly reflected in neural patterns in the left primary visual cortex, posterior fusiform gyrus, and bilateral precuneus when it was task-relevant, suggesting that information about referents' visual properties was flexibly retrieved from visual processing areas.

Moreover, especially regarding size information, strong effects of flexibility were observed in premotor and motor regions. Recall that the size task consisted in judging whether the word referent fits in a shoe box or not. As this action differs consistently with the size of the object at hand, the finding that patterns in motor areas are predicted by the size model suggest that participants simulated the action of fitting the object in the box, in line with previous evidence of motor and premotor cortex involvement in motor imagery (Porro et al., 1996) and action language comprehension (Desai, Binder, Conant, & Seidenberg, 2009; Hauk, Johnsrude, & Pulvermüller, 2004; Shtyrov, Butorina, Nikolaeva, & Stroganova, 2014; Vukovic, Feurra, Shpektor, Myachykov, & Shtyrov, 2017).

Thus, we propose that the activation of sensory-motor processes is highly task-dependent and can provide conceptual information 'tailor-made' for the current contextual demands. This resonates with recent behavioral evidence that low-level visual processes are causally involved in the comprehension of concrete nouns when the task requires retrieval of visual information, but not categorical semantic information

(Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017). In a related vein, it has become clear that motor cortex activation in response to action language is not automatic (Postle, McMahon, Ashton, Meredith, & de Zubicaray, 2008; Raposo, Moss, Stamatakis, & Tyler, 2009; van Dam et al., 2012; Watson, Cardillo, Ianni, & Chatterjee, 2013). In the light of such results it has been suggested that activation in sensory-motor areas is epiphenomenal (Caramazza, Anzellotti, Strnad, & Lingnau, 2014; Mahon, 2015; Mahon & Caramazza, 2008). Based on our present results and recent evidence that low-level visual and motor processes can play a functional role for comprehension (Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017; Vukovic et al., 2017), we suggest that sensory-motor areas are a crucial but highly context-sensitive component of the human conceptual system.

Effects of flexibility in the ATL

In addition to visual and motor areas, flexibility effects were consistently observed in the left ventral ATL. Both in terms of size and animacy information, portions of the left ATL mirrored the task-dependent activation profile of the distributed sensory regions detailed above. This is consistent with the view that the ATL functions as a convergence zone that integrates information from distributed lower-level systems tuned to perceptual features (Coutanche & Thompson-Schill, 2014; Lambon Ralph et al., 2010). Our results suggest that convergence in the ATL is an online phenomenon and that ATL processes are determined by momentary processing in lower-level areas, in line with a recent report of strong functional connectivity between sensory areas and the ATL during semantic processing (van Ackeren & Rueschemeyer, 2014).

This contrasts with the common view derived from research on object perception that the posterior-to-anterior axis in the ventral stream reflects increasing invariance to variable low-level features in order to build abstract conceptual representations (DiCarlo, Zoccolan, & Rust, 2012; Logothetis & Sheinberg, 1996; Riesenhuber & Poggio, 1999). For this to succeed, ATL neurons need to ignore sensory processes related to variable contextual factors and selectively respond to invariant features of a concept. One possibility is that different ATL subregions fulfill these different functions, in line with variability in terms of neuroanatomical (Ding, Van Hoesen, Cassell, & Poremba, 2009) and functional connectivity (Pascual et al., 2013) in subfields of the temporal polar cortex.

However, we found no evidence for task-invariant neural codes of conceptual information in any parts of the ATL. Based on the experience-dependent way in which invariance develops (Li & DiCarlo, 2008) and the highly flexible use of concepts in human language, it is possible that it is not an effective solution to rely on invariant codes.

Individual word decoding

Besides the pattern similarity analyses looking specifically at size and animacy, we investigated to what extent multi-voxel patterns related to individual word meanings change when they are processed in different contexts. In line with the RSA results, the strongest evidence for flexibility was observed in occipital areas. Interestingly, one region located in the left intraparietal sulcus and surrounding inferior and superior parietal cortex was found to show highly consistent word-specific patterns of activity regardless of task context. The inferior parietal cluster included the angular gyrus that has been proposed to be a conceptual hub region (Binder & Desai, 2011). Thus, whereas the ATL seems to be part of a highly context-sensitive distributed network, portions of the inferior and superior parietal lobules appear to respond invariantly to individual words. It is thus possible that this region processes abstract semantic knowledge that interacts with a distributed network to adapt to contextual demands. However, even though our analyses used spoken and written words, we cannot fully rule out that the patterns reflect word-level phonological information with which this region is also implicated (Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002).

Note also that due to general limitations of MRI we might have missed additional areas sensitive to context-dependent and context-invariant word meanings: those which are more variable between participants, which are based on distributed patterns in larger patches of cortex than our relatively small searchlights can capture, or which rely on very fine-grained neural activity patterns that MRI cannot detect. Therefore, more work is needed to further delineate the computations carried out in the flexible and invariant areas we observed.

Summary

Language use is extremely flexible suggesting that context can drastically change how words are processed. We tested how task demands modulate patterns of brain activity in response to the same set of concrete nouns. Our results indicate that the left ATL in concert with distributed primary sensory and association areas specifically enhances processing of those conceptual dimensions that are relevant in a given situation. This is in line with the proposal that the ATL functions as a convergence zone that integrates multimodal features computed in modality-specific systems (Coutanche & Thompson-Schill, 2014; Lambon Ralph et al., 2010). Our results suggest that this integration process is at least partly determined by task-related processes in sensory cortices. Thus, it is conceivable the ATL computes momentary coherent multimodal information that enables flexible language comprehension. At the same time, word-specific patterns in the left inferior and superior parietal lobules appeared to be invariant to the task context suggesting that, possibly in parallel, the brain computes more stable semantic representations. These findings fit with the view that conceptual flexibility and abstraction rely on at least partly segregated cortical structures comprising sensory regions on the one hand and high-level multimodal regions on the other (Binder, 2016; Binder, Desai, Graves, & Conant, 2009; Fernandino et al., 2016).

Acknowledgments

We would like to thank Paul Gaalman and José Marques for their help with the preparation of the MRI sequences and data collection.

References

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. *Theories of Memory*, *1*, 29–31.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

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.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, *16*(11), 3737–3744.

Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage*, *143*, 128–140.

Bracci, S., Daniels, N., & Op de Beeck, H. (2017). Task Context Overrules Object-and Category-Related Representational Content in the Human Parietal Cortex. *Cerebral Cortex*, *27*(1), 310–321.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84.

Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied cognition and mirror neurons: a critical assessment. *Annual Review of Neuroscience*, *37*, 1–15.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2(10), 913–919.

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *Journal of Neuroscience*, *34*(1), 332–338.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. *Cerebral Cortex*, 25(9), 2584–2593.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770.

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M. S. (2009). Activation of sensory–motor areas in sentence comprehension. *Cerebral Cortex*, *20*(2), 468–478.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434.

Ding, S.-L., Van Hoesen, G. W., Cassell, M. D., & Poremba, A. (2009). Parcellation of human temporal polar cortex: a combined analysis of multiple cytoarchitectonic, chemoarchitectonic, and pathological markers. *Journal of Comparative Neurology*, *514*(6), 595–623.

Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, *92*, 281–292.

Evans, V. (2009). *How words mean: Lexical concepts, cognitive models, and meaning construction*. Oxford University Press on Demand.

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., … Seidenberg, M. S. (2016). Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. *Cerebral Cortex*, 26(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

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

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics*, *7*(1), 37–53. https://doi.org/10.1007/s12021-008-9041-y

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences*, *111*(10), E962–E971.

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

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

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, *33*(25), 10235–10242.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412.

Lambon Ralph, M. A. L., 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.

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*(8), 1764–1801.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, *321*(5895), 1502–1507.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*(1), 577–621.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 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*(1), 59–70.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*(6566), 649–652.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, *6*(4), 462–472.

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., … Haxby, J. V. (2017). Attention selectively reshapes the geometry of distributed semantic representation. *Cerebral Cortex*, 1–15.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Ostarek, M., & Huettig, F. (2017). A Task-Dependent Causal Role for Low-Level Visual Processes in Spoken Word Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1215.https://doi.org/http://dx.doi.org/10.1037/xlm0000375

Pascual, B., Masdeu, J. C., Hollenbeck, M., Makris, N., Insausti, R., Ding, S.-L., & Dickerson, B. C. (2013). Large-scale brain networks of the human left temporal pole: a functional connectivity MRI study. *Cerebral Cortex*, *25*(3), 680–702.

Patterson, K., Nestor, P. J., & 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. https://doi.org/10.1038/nrn2277

Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience*, *32*(45), 15728–15736.

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97.

Porro, C. A., Francescato, M. P., Cettolo, V., Diamond, M. E., Baraldi, P., Zuiani, C., … Di Prampero, P. E. (1996). Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*(23), 7688–7698.

Postle, N., McMahon, K. L., Ashton, R., Meredith, M., & de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *Neuroimage*, *43*(3), 634–644.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025.

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, *24*(11), 2218–2225.

Shtyrov, Y., Butorina, A., Nikolaeva, A., & Stroganova, T. (2014). Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proceedings of the National Academy of Sciences*, *111*(18), E1918–E1923.

Simon, O., Mangin, J.-F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, *33*(3), 475–487.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage*, *65*, 69–82.

van Ackeren, M. J., & Rueschemeyer, S.-A. (2014). Cross-modal integration of lexical-semantic features during word processing: evidence from oscillatory dynamics during EEG. *PloS One*, *9*(7), e101042.

Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147–182.

van Dam, W. O., van Dijk, M., Bekkering, H., & Rueschemeyer, S.-A. (2012). Flexibility in embodied lexical-semantic representations. *Human Brain Mapping*, *33*(10), 2322–2333.

Van Dam, W. O., Van Dongen, E. V., Bekkering, H., & Rueschemeyer, S.-A. (2012). Context-dependent changes in functional connectivity of auditory cortices during the perception of object words. *Journal of Cognitive Neuroscience*, *24*(10), 2108–2119.

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, *96*, 222–229.

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action concepts in the brain: an activation likelihood estimation meta-analysis. *Journal of Cognitive Neuroscience*, *25*(8), 1191–1205.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (can) prime furiously. *Psychological Science*, *23*(4), 364–369.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027.

Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T., & Raizada, R. D. (2016). Semantic structural alignment of neural representational spaces enables translation between English and Chinese words. *Journal of Cognitive Neuroscience*.

Appendix

List of stimuli

Animacy	Size	Words (Dutch)	Words (English)
animate	big	giraf	giraffe
animate	big	hond	dog
animate	big	inktvis	octopus
animate	big	walvis	whale
animate	big	walrus	hippo
animate	big	aap	monkey
animate	small	kat	cat
animate	small	hamster	hamster
animate	small	muis	mouse
animate	small	vlinder	butterfly
animate	small	kreeft	lobster
animate	small	goudvis	goldfish
inanimate	big	heuvel	hill
inanimate	big	boom	tree
inanimate	big	hut	hut
inanimate	big	pompoen	pumpkin
inanimate	big	gitaar	guitar
inanimate	big	tafel	table
inanimate	small	tomaat	tomato
inanimate	small	appel	apple
inanimate	small	bloem	flower
inanimate	small	sleutel	key
inanimate	small	veer	feather
inanimate	small	knoop	button

Chapter 8

SUMMARY AND DISCUSSION

Seemingly seamlessly, in less than half a second, words are transformed from a highly variable acoustic signal into meaningful units of communication allowing us to share feeling and thoughts, make agreements and plans, and ultimately get a glimpse into each other's minds. Language comprehension is an astonishing cognitive capacity. An understanding of the underlying cognitive and neural processes is a central endeavor in the cognitive sciences. Due to empirical developments (Barsalou, 2008; Martin, 2007a; Martin, Wiggs, Ungerleider, & Haxby, 1996; Pulvermüller, 1999, 2005) and theoretical considerations (Barsalou, 1999; Harnad, 1990), an influential framework has gained ground according to which sensory-motor processes, otherwise involved in perception and action, are instrumental for high-level functions including conceptual processing. This thesis particularly focused on visual processes and aimed to establish what role they play in language comprehension by combining behavioral, eye-tracking, and neuroimaging methods. Below, I summarize the main findings, embed them in the broader context of the literature, and give an outlook of what I believe are the most pressing next steps to move the field ahead.

Spoken words activate low-level visual processes

Chapter 2 employed continuous flash suppression to test whether spoken words activate basic visual processes related to the objects denoted by the words. Previous demonstrations of language-vision interactions only allowed limited conclusions about the processes involved because pictures were fully processed and it remained unclear at what stage the often-observed congruency effects arose (e.g., Zwaan, Stanfield, & Yaxley, 2002). The strength of CFS in that context is that it can be used in a way that allows tapping of visual processing in the absence of semantic processing. It had been shown previously that word cues can facilitate the detection of congruent objects in CFS suggesting that high-level cognition can penetrate vision (Lupyan & Ward, 2013) and that visual expectations constrained by word cues can be very specific (Pinto, van Gaal, de Lange, Lamme, & Seth, 2015). Chapter 2 adapted the previously used CFS paradigm to tap into semantic processing by reducing the temporal lag between words and targets to a minimum, by ensuring that the words were not predictive of the subsequent targets, and by informing participants about that state of affairs. The result was that passive listening of words enhanced the detection of congruent (but previously unseen) objects, suggesting that words activate very basic visual processes involved in the earliest stages of conscious perception. This effect was found to arise very quickly during word processing (earlier than 400 ms after word onset) and to be quite short-lived, a time-course highly consistent with that typically associated with semantic processing. As the target pictures were unknown to the participants and were typically not recognized during the experiment, this suggests that the visual processes activated during spoken word comprehension satisfy two prime requirements of semantic processing at once, namely specificity and abstraction.

In this study, the definition of the activated processes as 'low-level' was a functional one: The idea was that it would be reasonable to characterize those processes as low-level that drive the most basic form of detection. In the context of visual processing, 'low-level' is often mapped onto primary visual areas. It would be fascinating to investigate how early these effects are reflected in the visual cortex (anatomically and temporally) and, more generally, to explore the neural correlates of the word-induced detection boost to better understand the flow of information across and within systems. Recent EEG and MEG experiments showed that word cues can influence visual processing in the first hundred ms after picture onset (Boutonnet & Lupyan, 2015; Hirschfeld, Zwitserlood, & Dobel, 2011). Future studies could investigate how the transient effect of words on detection in CFS is reflected in neural activation in the visual cortex and beyond.

The neural correlates of conceptual flexibility

A more direct way of probing the neural processes involved in semantic processing is to measure neural activity in tasks that do not include a picture component. I capitalized on variants of multi-voxel pattern analysis to study bare word comprehension in Chapter 7. In particular, I investigated how contextual demands shape the way concrete words are processed. Representational similarity analysis was used to test the hypothesis that information about size and animacy is more explicitly reflected in word-specific activation patterns when it is task relevant vs. irrelevant. The results showed that information about the relevant dimensions was strongly enhanced in portions of the primary visual cortex and association areas. Interestingly, information about size and animacy was very weak when it was not relevant for the task, and virtually entirely absent in occipital areas and inferior temporal areas strongly associated with these dimensions (Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Konkle & Oliva, 2012; Martin, 2007b; Martin et al., 1996; Sha et al., 2015; Tranel, Damasio, & Damasio, 1997). This is somewhat surprising given that they are behaviorally and cognitively very important features. Some theories even hold that animacy is to a non-trivial extent innate (Capitani, Laiacona, Mahon, & Caramazza, 2003; Farah & Rabinowitz, 2003; Mahon & Caramazza, 2009). While it is of course possible that a genetically constrained neural architecture exists that is conductive to a categorization of objects as animate vs. inanimate (and perhaps big vs. small), my results suggest that the way these dimensions are processed is far from stable.

Highly flexible coding of size and animacy was also observed in the left ATL which is often considered a multimodal (or amodal) hub (Chen, Lambon Ralph, & Rogers, 2017; Patterson, Nestor, & Rogers, 2007; Pobric, Jefferies, & Lambon Ralph, 2010; Pobric, Lambon Ralph, & Jefferies, 2009; Lambon Ralph, 2014; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017; Lambon Ralph, Sage, Jones, & Mayberry, 2010). This finding may help constrain the computational principles of this area. In particular, it fits with the idea that featural integration in the ATL occurs online and is constrained by momentary processing in modality-specific cortices. In a sophisticated study combining representational similarity and functional connectivity analysis, Coutanche and Thompson-Schill (2014) provided evidence for the retrieval of color and shape information from separate visual areas and their convergence in the left ATL. MEG evidence supports the view that the ATL engages in dynamic bidirectional links with sensory regions during semantic processing (van Ackeren & Rueschemeyer, 2014; van Ackeren, Schneider, Müsch, & Rueschemeyer, 2014). This may allow the ATL to build context-sensitive temporary conceptual representations.

Task-invariant patterns in response to individual words were observed in portions of the superior and inferior parietal lobules, including the angular gyrus. The latter, based on its connectivity profile and consistent involvement in semantic tasks (Binder, Desai, Graves, & Conant, 2009), has also been suggested as a semantic hub (Binder & Desai, 2011). Recent fMRI studies examining conceptual combination provide some clues regarding the different role that the angular gyrus and the ATL may play. An emerging picture seems to be that attribute-based combination relies on the ATL (Chen et al., 2016; Correia et al., 2014; Coutanche & Thompson-Schill, 2014; Lambon Ralph et al., 2010), whereas functional combination required for assigning thematic roles and building event structure engage the angular gyrus (Binder & Desai, 2011; Boylan, Trueswell, & Thompson-Schill, 2015, 2017). My finding that the angular gyrus is not sensitive to animacy and size but responds task-invariantly to word meanings points to a role at the interface of semantics and syntax.

These proposals are speculative as they rest on a relatively little thin literature. However, since the introduction of machine learning techniques, it has become more feasible to match the questions psycholinguists have with the tools that cognitive neuroscience offers.

The challenge of item-independent concepts

A very interesting line of research has looked into words whose referents are typically perceived up vs. down in space. These up/down words are a great opportunity to explore how spatial aspects of word meanings are processed. Spatial conceptual features, such as 'up' and 'down', are particularly interesting because they do not consistently co-vary with any of the sensory-motor attributes that make up the concrete concepts that are typically studied. Previous experimental work revealed that processing up/down words induces changes in subsequent visuo-spatial behavior (Dudschig, Lachmair, de la Vega, De Filippis, & Kaup, 2012; Dudschig, Souman, Lachmair, de la Vega, & Kaup, 2013; Dunn, 2016; Estes, Verges, & Adelman, 2015; Estes, Verges, & Barsalou, 2008; Gozli, Chasteen, & Pratt, 2013; Ostarek & Vigliocco, 2017). Chapter 5 built on this work by showing that processing up/down words initially biases the trajectories of concurrently executed saccades in the congruent direction. As the saccade unfolds, the semantic relation between words and targets gets crucial: The congruency bias increases for semantically related targets and reverses for unrelated objects, suggesting that processing costs are incurred when there is a match in one domain and a mismatch in another.

The initial congruency bias exerted by up/down words on eye movements is consistent with the idea that spatial information is retrieved from the cortical oculomotor system or from regions that directly interact with it. It would also be possible, however, that the language-space interactions are driven by the task setting and/or implicit spatial expectations. To tease these alternatives apart, I conducted an fMRI experiment in which participants performed a semantic task on up/down words (and filler words) in the absence of a spatial task (Chapter 6). Multi-voxel pattern analysis was used to map out the brain regions that are sensitive to the vertical dimension implicit in the words. A cross-decoding analysis was also performed in which a classifier was trained to distinguish up vs. down saccades and then its ability to distinguish up vs. down words was assessed without further training. Both analyses in conjunction yielded strong evidence that up/down words trigger direction-specific processes in regions associated with overt and covert spatial attention (Corbetta et al., 1998). Information about the words' spatial associations was also present in temporal areas associated with high-level categorization, suggesting that a network of regions with qualitatively different roles is the basis for spatial semantics. Functional connectivity analyses and MEG experiments could further reveal the underlying processing dynamics and delineate the functions subserved by the subcomponents of this network.

The functional role of visual processes

A substantial amount of data has accumulated supporting the view that conceptual processing involves the sensory-motor systems (Kiefer & Pulvermüller, 2012; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012). One reason why it nevertheless remains controversial to what extent cognition is embodied is that there is very little causal evidence (Mahon, 2015). Instead of merely demonstrating an involvement, a much stronger test would be to provide evidence that sensory-motor processes are necessary. Lesion studies, despite being numerous (Bak & Chandran, 2012; Bak & Hodges, 2004; Bak, O'donovan, Xuereb, Boniface, & Hodges, 2001; Boulenger et al., 2008; Damasio et al., 1996; Desai, Herter, Riccardi, Rorden, & Fridriksson, 2015; Fernandino et al., 2013; Kemmerer, Rudrauf, Manzel, & Tranel, 2012; Tranel et al., 1997), have not resolved the issue because lesions and the corresponding impairments (partly due to plasticity and compensation strategies) are rarely very specific (Binder & Desai, 2011; Meteyard et al., 2012).

A promising but so far scarcely employed technique is transcranial magnetic stimulation (TMS). TMS can induce cortical activity and influence online cognition. Stimulation of the motor and premotor cortices has been reported to have a selective effect on action language comprehension in several studies (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005; Repetto, Colombo, Cipresso, & Riva, 2013; Vukovic, Feurra, Shpektor, Myachykov, & Shtyrov, 2017; Willems, Labruna, D'Esposito, Ivry, & Casasanto, 2011). However, as both interference (Repetto et al., 2013; Vukovic et al., 2017) and facilitation (Pulvermüller et al., 2005; Willems et al., 2011) were observed across studies, more work is needed to clarify the link between motor cortex and action language comprehension.

Chapter 3 used a similar approach optimized for the visual domain. The assumption was that highly imageable concepts largely derived from visual experience rely more strongly on the recruitment of visual processes than abstract concepts. To test whether low-level visual processes are causally involved in semantic processing of concrete words, we interfered with them while participants engaged in different tasks. The main result was that visual interference hindered the processing of imageable concrete words more than that of abstract words, but only when the task required the retrieval of visual information: The effect was observed in a concreteness judgement task, but not in a lexical decision and a word class judgement task. A very similar result was nearly simultaneously obtained by Edmiston and Lupyan (2017) who found that visual interference impaired access to visual, but not encyclopedic, information in a property verification task. Moreover, the task-dependence was mirrored in the TMS study by Vukovic et al. (2017) which reported an interference effect of primary motor cortex stimulation on action word comprehension in a concreteness task, but not in a lexical decision task. This convergence of evidence from different paradigms and modalities makes a solid case for the view that sensory-motor processes are flexibly, but causally, involved in semantic processing.

Having established the visual noise technique as a tool to effectively interfere with basic visual processes that can be recruited for access to visual information, Chapter 4 applied it to the sentence-picture verification paradigm (Stanfield & Zwaan, 2001). It is uncontroversial that the shape-match effect first reported by Zwaan et al. (2002) demonstrates the access to shape information during sentence comprehension. Moreover, it appears to be very robust (Zwaan & Pecher, 2012), at least in sentence-picture verification (see Rommers, Meyer, & Huettig, 2013). But does it rely on perceptual simulation? Surprisingly, despite the fact that it is a hallmark finding in the field, this has never been explicitly tested. To tackle this question, Chapter 4 used the visual noise technique that Chapter 3 established to be capable of interfering with basic visual processes involved in the retrieval of visual information. It followed the rationale that visual noise should make the shape-match effect disappear if it relies on low-level processes. This was neither the case when the same low-level visual noise was used as in Chapter 3 and by Edmiston and Lupyan (2017), nor when a type of higher-level visual noise based on superimposed distorted objects was used. The shape-match effect only vanished once the visual noise was composed of intact objects. This suggests that the shape-match effect occurs on a level that is independent of basic visual processes.

One potential criticism could be that the gap between the two types of visual noise and the semantic (object) noise was too big and visual processes at a higher level than we have tested drive the effect, which are nevertheless modality-specific. While it is difficult to rule out that possibility, it seems clear that low-level visual processes are not causal for the match advantage. This finding contrasts with the single word paradigms reviewed above and suggests that visual information derived from event-level representations may rely on abstraction away from the sensory systems. This proposal appears to be consistent with previous reports of conceptual combination in high-level cortical regions, such as the ATL, but not in the occipital cortex (Bemis & Pylkkänen, 2012; Coutanche & Thompson-Schill, 2014). At the same time, others have suggested that the increased availability of contextual specificity in sentences and discourse makes it more likely that sensory-motor processes are activated (Kurby & Zacks, 2013; Zwaan, 2014). More work will be needed to establish how the granularity of linguistic units relates to the type of mechanism involved.

Conclusion

This thesis set out to clarify the role that perceptual processes play in language comprehension. Using continuous flash suppression, visual interference, eye-tracking, and fMRI, I provided evidence for the view that the visual system is an important component of a highly flexible cognitive machinery. The overall evidence suggests that surprisingly low-level visual processes are recruited to provide information about visual features, especially when access to these features is task-relevant. Sensory-motor processes are activated in the context of a large-scale network including high-level multimodal association areas. Researchers have only begun to understand the functional division across and the complex interactions among these regions. A full mechanistic understanding of the processes underlying language comprehension should of course be the ultimate aim, but is probably one of the most challenging endeavors in the cognitive sciences. I hope that my thesis is a small step in the right direction, however small it may be.

References

Bak, T. H., & Chandran, S. (2012). What wires together dies together: verbs, actions and neurodegeneration in motor neuron disease. *Cortex*, *48*(7), 936–944.

Bak, T. H., & Hodges, J. R. (2004). The effects of motor neurone disease on language: further evidence. *Brain and Language*, *89*(2), 354–361.

Bak, T. H., O'donovan, D. G., Xuereb, J. H., Boniface, S., & Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease–dementia–aphasia syndrome. *Brain*, *124*(1), 103–120.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bemis, D. K., & Pylkkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, *23*(8), 1859–1873.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: a label advantage in object recognition. *Journal of Neuroscience*, *35*(25), 9329–9335.

Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2015). Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia*, *78*, 130–141.

Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2017). Relational vs. attributive interpretation of nominal compounds differentially engages angular gyrus and anterior temporal lobe. *Brain and Language*, *169*, 8–21.

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*(3–6), 213–261.

Chen, L., Lambon Ralph, M. A. L., & Rogers, T. T. (2017). A unified model of human semantic knowledge and its disorders. *Nature Human Behaviour*, *1*, 0039.

Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., … Lambon Ralph, M. L. (2016). The "when"and "where"of semantic coding in the anterior temporal lobe: temporal representational similarity analysis of electrocorticogram data. *Cortex*, *79*, 1–13.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., … Shulman, G. L. (1998). A Common Network of Functional Areas for Attention and Eye Movements. *Neuron*, *21*(4), 761–773. https://doi.org/10.1016/S0896-6273(00)80593-0

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *Journal of Neuroscience*, *34*(1), 332–338.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. *Cerebral Cortex*, 25(9), 2584–2593.

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

Desai, R. H., Herter, T., Riccardi, N., Rorden, C., & Fridriksson, J. (2015). Concepts within reach: Action performance predicts action language processing in stroke. *Neuropsychologia*, *71*, 217–224.

Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012). From top to bottom: spatial shifts of attention caused by linguistic stimuli. *Cognitive Processing*, *13*(1), 151–154.

Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. *PloS One*, *8*(2), e56872.

Dunn, B. M. (2016). *Which way is up? Grounded mental representations of space*. University of Glasgow. Retrieved from http://theses.gla.ac.uk/id/eprint/7460

Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, 92, 281–292.

Estes, Z., Verges, M., & Adelman, J. S. (2015). Words, objects, and locations: Perceptual matching explains spatial interference and facilitation. *Journal of Memory and Language*, *84*, 167–189.

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down object words orient attention to the objects' typical location. *Psychological Science*, *19*(2), 93–97.

Farah, M. J., & Rabinowitz, C. (2003). Genetic and environmental influences on the organisation of semantic memory in the brain: Is "living things" an innate category? *Cognitive Neuropsychology*, *20*(3–6), 401–408.

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia*, *51*(8), 1510–1517.

Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit spatial cues for visual attention. *Journal of Experimental Psychology: General*, *142*(4), 1028.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346.

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing–MEG dissociates early perceptual and late N400 effects. *Brain and Language*, *116*(2), 91–96.

Kemmerer, D., Rudrauf, D., Manzel, K., & Tranel, D. (2012). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex*, *48*(7), 826–848.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124.

Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, *126*(3), 338–349.

Lambon Ralph, M. A. L. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1634), 20120392.

Lambon Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Lambon Ralph, M. A. L., 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.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 420–429.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51.

Martin, A. (2007a). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Martin, A. (2007b). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*(6566), 649–652.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804.

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 579.

Patterson, K., Nestor, P. J., & 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. https://doi.org/10.1038/nrn2277

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13–13.

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

Pobric, G., Lambon Ralph, M. A. L., & Jefferies, E. (2009). The role of the anterior temporal lobes in the comprehension of concrete and abstract words: rTMS evidence. *Cortex*, *45*(9), 1104–1110.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*(2), 253–279.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797.

Repetto, C., Colombo, B., Cipresso, P., & Riva, G. (2013). The effects of rTMS over the primary motor cortex: The link between action and language. *Neuropsychologia*, *51*(1), 8–13.

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological Science*, *24*(11), 2218–2225.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, *35*(10), 1319–1327.

van Ackeren, M. J., & Rueschemeyer, S.-A. (2014). Cross-modal integration of lexical-semantic features during word processing: evidence from oscillatory dynamics during EEG. *PloS One*, *9*(7), e101042.

van Ackeren, M. J., Schneider, T. R., Müsch, K., & Rueschemeyer, S.-A. (2014). Oscillatory neuronal activity reflects lexical-semantic feature integration within and across sensory modalities in distributed cortical networks. *Journal of Neuroscience*, *34*(43), 14318–14323.

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, *96*, 222–229.

Willems, R. M., Labruna, L., D’Esposito, M., Ivry, R., & Casasanto, D. (2011). A functional role for the motor system in language understanding: evidence from theta-burst transcranial magnetic stimulation. *Psychological Science*, *22*(7), 849–854.

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PloS One*, *7*(12), e51382.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Acknowledgements

At the beginning of most PhD projects, things can go all kinds of ways, and that was certainly the case in mine. One major reason that these three years were so much fun was my supervisor/advisor Falk. If there is such a thing as a perfect supervisor, for me you came very close. I will always be grateful for everything you taught me, and for giving me the space and freedom I needed to develop as a researcher. Thanks for the faith you put in me.

To Antje, thank you very much for giving me the opportunity to work on a project that sometimes must have looked adventurous to you (at best) and for giving me all the support I needed.

Amie, you were the first colleague I got to know (before we even were colleagues). The idea of becoming your labmate was great and I think I really knew that I wanted to go for a PhD in Nijmegen when we were walking through the park of 150 trees where we took a picture on the colorful bench and somebody thought we were a couple and we just went with it for the sake of the photo. We spent an exciting first year in 381 and things surely got spiced up when you joined the party, Sara! Thanks for the nice walks and chats in the forest (few people seem to appreciate that the forest is one of the best things about the institute). Saoradh, and Miguel, I feel very lucky to have shared an office with you and to be able to count you among my friends. Jeroen, I really enjoyed collaborating with you on the MRI projects and am thankful that you shared your amazing programming and problem solving skills with me.

It was a true privilege being in such a stimulating place, surrounded by an extraordinary set of people who always (or let's say most of the time) have something interesting to say about science, life, the universe, or anything really. Thanks to the members of PoL for making this such a fun ride.

One reason why the MPI is so special is that the operations (administration) and technical group do a fantastic job at making sure everything works as smoothly as possible. Thanks for saving me a lot of nerves along the way.

These three years would have gone quite differently if I did not have the most wonderful partner by my side who not only is a constant inspiration but also saved me from going down into stress spirals countless times. Thank you, Nadine!

Finally, I want to thank my parents for always being maximally supportive, no matter what path I decide to follow.

Nederlandse samenvatting

Wat zijn de basismechanismes achter het begrijpen van taal? Theoretici zijn het er in het algemeen over eens dat we de betekenis van woorden vooral leren via onze ervaringen in de wereld. Dit betekent dat alles wat we doen, zien, horen of aanraken, bijdraagt tot een steeds groter wordende verzameling van informatie die we kunnen gebruiken om te achterhalen waar woorden naar verwijzen. Hoe de cognitieve en neurale processen die nodig zijn om deze bron van informatie te gebruiken werken, is nog steeds niet helemaal duidelijk. Volgens een recente theorie heeft het menselijke brein een manier gevonden om de evolutionair gezien oude sensorische systemen een nieuwe toepassing te geven. Het centrale idee is dat we bij het begrijpen van een woord karakteristieke sensorische eigenschappen van het object waar dat woord naar verwijst, ophalen (zoals de rondheid van een voetbal). Dit doen we door processen in het sensorisch systeem te reactiveren (alsof we echt een bal zien). Met behulp van gedragsexperimenten en neuroimaging, onderzocht ik in welke mate we bij het begrijpen van concrete woorden zoals *voetbal* een beroep doen op processen die ook gebruikt worden voor visuele perceptie. De resultaten tonen aan dat woorden simpele visuele processen kunnen activeren, zoals de processen die nodig zijn bij het herkennen van vormen en het controleren van oogbewegingen. De processen die geactiveerd worden tijdens het begrijpen van taal zijn niet automatisch, maar hangen af van de context waarin woorden voorkomen. Het visuele systeem lijkt deel uit te maken van een uiterst flexibel cognitief apparaat dat ontzettend goed is in het aanreiken van de aspecten van onze kennis die het meest relevant zijn in een bepaalde situatie.

CV

Markus Ostarek was born on January 15th 1990 in Rosenheim (Germany). He did a bachelor's degree in Romance linguistics (Ludwig-Maximilians-Universität Munich), spent six months in Paris teaching German, and then did a master's degree in neuroscience and linguistics at the University College London, where he started working on the neural and cognitive processes related to language perception and comprehension. He did a PhD at the Max Planck Institute for Psycholinguistics where he now works as a postdoctoral researcher.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*

29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*
63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*
64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*
66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*

67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*
69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*
70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*
71. Gaze behavior in face-to-face interaction. *Federico Rossano*
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*
73. Who is talking? Behavioral and neural evidence for norm-based coding in voice identity learning. *Attila Andics*
74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*
75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*
76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*
77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*
78. Contributions of executive control to individual differences in word production. *Zeshu Shao*
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*
80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*
81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*
82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*
83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*
84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*
85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*
86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*

87. The acquisition of morphophonological alternations across languages. *Helen Buckler*
88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*
89. Rediscovering a forgotten language*. Jiyoun Choi*
90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*
91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*
92. Information structure in Avatime. *Saskia van Putten*
93. Switch reference in Whitesands. *Jeremy Hammond*
94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*
95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sümer*
96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*
97. lncrementality and Flexibility in Sentence Production. *Maartje van de Velde*
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*
99. The request system in Italian interaction. *Giovanni Rossi*
100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*
101. Assessing birth language memory in young adoptees**.** Wencui Zhou
102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*
103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*
104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition*. Rósa Signý Gisladottir*
105. Modelling Multimodal Language Processing. *Alastair Smith*
106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*
107. Situational variation in non-native communication. *Huib Kouwenhoven*
108. Sustained attention in language production. *Suzanne Jongman*
109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects. *Malte Viebahn*

110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*

111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*

112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*

113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*

114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*

115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*

118. Language processing in a conversation context. *Lotte Schoot*

119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*

120. Talking Sense: the behavioral and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*

122. Sensorimotor experience in speech perception. *Will Schuerman*

123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*

124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*

125. Lapse organization in interaction. *Elliott Hoey*

126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*

127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*

128. The role of voice and word order in incremental sentence processing. *Sebastian Sauppe*

129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*

130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*

131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*
132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*
133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*
134. Indirect request comprehension in different contexts. *Johanne Tromp*
135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*