Cognition 182 (2019) 84–94

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task?

Markus Ostarek[a,b,*], Dennis Joosen[a], Adil Ishag[c], Monique de Nijs[a], Falk Huettig[a]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] International Max Planck Research School for Language Sciences, The Netherlands
[c] International University of Africa, Khartoum, Sudan

ARTICLE INFO

Keywords:
Language comprehension
Conceptual processing
Perceptual simulation
Embodied cognition
Sentence-picture verification

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. Visual noise specifically targeting high-level visual processes (Experiment 4) only had a minimal effect on the match effect. We conclude that the shape match effect in the sentence-picture verification paradigm is unlikely to rely on perceptual simulation.

1. 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 & Caramazza, 2008; Mahon, 2015).

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 & Desai, 2011; Binder et al., 2016; Collins & Loftus, 1975; Cree & McRae, 2003; Fernandino et al., 2016; Fernandino, Humphries, Conant, Seidenberg, & Binder, 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, 2008; 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 behavioural 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, 2017b, 2017a; 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; 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 shapes (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007). As many objects can occur in multiple different shapes, listeners often need to incorporate contextual information in order to retrieve the appropriate

* Corresponding author at: MPI Psycholinguistics, Wundtlaan 1, Nijmegen, The Netherlands.
E-mail address: markus.ostarek@mpi.nl (M. Ostarek).

https://doi.org/10.1016/j.cognition.2018.08.017
Received 7 September 2017; Received in revised form 27 August 2018; Accepted 27 August 2018

shape representations. Using the sentence-picture verification task, a classic experiment by 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 or not 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) and about reproducibility more generally (Pashler & Wagenmakers, 2012; 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 et al., 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 behavioural pattern, as studies using the sentence-picture verification paradigm can only tell us something about the kind of information that 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. Hirschfeld, Zwitserlood, and Dobel (2011) conducted a magnetoencephalography study using the sentence-picture verification paradigm to assess changes in neural activity for shape matching vs. mismatching pictures. They observed a stronger positivity to pictures following shape matching vs. mismatching sentences in occipital cortex at ca. 120 ms after picture onset (M1), suggesting a top-down modulation of early visual processing as a function of shape match vs. mismatch. However, changes in the way the target picture was visually processed do not necessarily imply that visual processes were activated during comprehension. Indeed, that scenario would predict repetition suppression, not enhancement. Therefore, the data are consistent with with top-down input from higher-level cortical areas. Thus, 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.

2. 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). We predicted that visual interference would decrease the match advantage based on four considerations: (1) the match effect pertains to visual shape information, (2) processing of shape information in early visual cortex has been shown to be modulated in the sentence-picture verification task (Hirschfeld et al., 2011), (3) previous studies reported interference effects of visual noise on semantic processing of single words (Edmiston & Lupyan, 2017; Ostarek & Huettig, 2017a), and (4) the intuitive proposal that contextually embedded language tends to engage more specific representations and might thus be more likely to activate modality-specific processes than single words (Kurby & Zacks, 2013; Zwaan, 2014).

2.1. Method

2.1.1. 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, and one due to excessive error rates ($>$ 20%), resulting in 110 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.

2.1.2. Materials, set-up, and design

We used the materials from the original Zwaan et al. (2002) study that were provided by Rommers, Meyer, Praamstra, and Huettig (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 cm 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,