BRAIN RESEARCH 1377 (2011) 78–83

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

How vision is shaped by language comprehension — Top-down feedback based on low-spatial frequencies

Gerrit Hirschfeld[*], *Pienie Zwitserlood*

Institute for Psychology, University of Münster, Germany
Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, Germany

ARTICLE INFO

Article history:
Accepted 17 December 2010
Available online 28 December 2010

Keywords:
Language comprehension
Vision
Embodied cognition
Spatial frequencies
Top-down feedback

ABSTRACT

Effects of language comprehension on visual processing have been extensively studied within the embodied-language framework. However, it is unknown whether these effects are caused by passive repetition suppression in visual processing areas, or depend on active feedback, based on partial input, from prefrontal regions. Based on a model of top-down feedback during visual recognition, we predicted diminished effects when low-spatial frequencies were removed from targets. We compared low-pass and high-pass filtered pictures in a sentence–picture-verification task. Target pictures matched or mismatched the implied shape of an object mentioned in a preceding sentence, or were unrelated to the sentences. As predicted, there was a large match advantage when the targets contained low-spatial frequencies, but no effect of linguistic context when these frequencies were filtered out. The proposed top-down feedback model is superior to repetition suppression in explaining the current results, as well as earlier results about the lateralization of this effect, and peculiar color match effects. We discuss these findings in the context of recent general proposals of prediction and top-down feedback.

1. Introduction

With language, "we can shape events in each others brains with exquisite precision" (Pinker, 1994). How this is achieved has received much interest since the advance of the embodied cognition perspective (Glenberg and Kaschak, 2002). According to the embodied- or grounded-cognition framework, conceptual knowledge is grounded in perceptual and motor states (Barsalou, 2008). For language, this means that listeners represent mentioned scenes by activating perceptual and motor simulations of these scenes. These simulations rely on similar neural resources that would be activated during perception of the described scene, and hence act as perceptual symbols (Barsalou, 1999). While perceptual symbol systems were introduced as an alternative to abstract representations, recent proposals take a more balanced view and treat them as complementary representational systems (Dove, 2009; Mahon and Caramazza, 2008). Such a dual view renders the question of what brings about these effects even more important. Our goal in this study was to investigate the mechanisms that give rise to very specific context effects. Can these effects be explained by passive repetition suppression, or are more complex processes involved?

---

[*] *Corresponding author*. Psychological Institute II, Westfälische Wilhelms-Universität Münster, Fliednerstraße 21, 48149 Münster, Germany. Fax: +49 251 8334104.
E-mail address: gerrit.hirschfeld@uni-muenster.de (G. Hirschfeld).
URL: http://www.psycholinguistics.uni-muenster.de (G. Hirschfeld).

0006-8993/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2010.12.063

A plethora of behavioral and neurocognitive studies has demonstrated that language comprehenders activate contextually relevant perceptual or motor features (for a review see: Barsalou, 2008). Neurocognitive studies have explored the activation of sensory and motor areas in language comprehension at a neural level. For example, fMRI studies have investigated how verbs that describe actions executed with different effectors are processed (Buccino et al., 2005; Hauk et al., 2004). Similarly, specific talked-about sensory modalities and areas for motor-planning are activated during story comprehension (Speer et al., 2009). Generally, these neurocognitive studies have focused on very coarse differentiations between broad semantic categories, while behavioral work has focused on fine-grained differences between stimuli.

The sentence–picture-verification paradigm has shown how very detailed grained perceptual features are activated during language comprehension, and this has been widely replicated (Brunye, 2009; Dijkstra et al., 2004; Madden and Therriault, 2009; Noordzij et al., 2005; Stanfield and Zwaan, 2001; Zwaan et al., 2002). In this paradigm, participants have to verify whether pictured targets were mentioned in a preceding sentence or not. Critically, the relation between sentences and pictures is manipulated. Objects mentioned in the sentence can either match or mismatch task-irrelevant perceptual features that are implied in the sentence (e.g. their orientation, shape or direction of motion). Reaction times are typically facilitated when these implied features match the visually presented targets (i.e. the match-effect: Stanfield and Zwaan, 2001).

Neurocognitive studies on the neural basis of mental imagery provide converging evidence that sensory cortices are activated during mental-imagery tasks (Kosslyn et al., 2001), but also by tasks that do not require conscious imagery, such as the sentence–picture-verification task (Carpenter et al., 1999). While these studies suggest an involvement of perceptual-processing regions, the precise mechanisms are far from clear. The most parsimonious mechanism for the match effect assumes that simulating an object has the same effect as actually seeing it. Specifically, simulations can evoke effects comparable to passive repetition suppression, which is frequently observed when visually presented objects are repeatedly accessed (Grill-Spector et al., 2006; Henson, 2003). According to this explanation, targets that match the orientation of the description have a processing advantage, because language comprehension engages the same neurons that code the orientation as in picture perception. If the context implies the same orientation as shown in the subsequent picture the overlap between simulation and the actual object is larger than in a mismatching condition (Stanfield and Zwaan, 2001).

A recent MEG-study with the sentence–picture-verification paradigm found the opposite. Pictures that matched the implicitly mentioned shape evoked larger M1 amplitudes than mismatching targets, when reduced activations in perceptual areas would have been expected (Hirschfeld et al., 2011). Furthermore, unrelated pictures evoked a large N400-like response in the left hemisphere that was similar for matching and mismatching targets. A recent EEG-study revealed systematic differences between the effects of visual adaptation and mental imagery on visual processing (Ganis and Schendan, 2008). While adaptation decreased the amplitude of the N170 component, imagery enhanced it. The authors surmise that amplifying effects of linguistic contexts or imagery result from top-down feedback rather than passive repetition suppression (Ganis and Schendan, 2008; Hirschfeld et al., 2011). The crucial difference between this view and previous explanations, which also entail top-down feedback, is that only former but latter assumes reentrant top-down feedback connections.

A highly detailed model for such a reentrant feedback mechanism has been proposed by Moshe Bar (2004). According to this model visual–object recognition involves a very fast initial analysis of low-spatial-frequency information in a visual scene. This initial analysis relies on the dorsal magnocellular pathway to the orbitofrontal cortex (Kveraga et al., 2007) that generates initial guesses about possible objects. These are integrated with the bottom-up stream of the slower ventral visual pathway via top-down projections from the orbitofrontal cortex to fusiform gyrus. Thus, context might not affect primary sensory processing regions alone, but also induce changes in reentrant orbitofrontal connections that make use of low spatial frequency information, which is rapidly extracted from the target.

This top-down feedback mechanism can also account for hemispheric differences in the sentence–picture-verification task (Lincoln et al., 2008). Patients with damage to the left hemisphere exhibit similar context effects than controls, but patients with damage to the right hemisphere do not show any effect. Assuming that the right hemisphere is better tuned to low-spatial frequencies (Grabowska and Nowicka, 1996), a feedback mechanism based on low-spatial frequencies would predict such differences (Lincoln et al., 2008).

1.1. *The present study*

The aim of the present study was to test if the context effect in the sentence–picture-verification task depends on a feedback mechanism that is driven by low-spatial frequencies. We used a sentence–picture-verification task and manipulated the targets by high-pass or low-pass filtering, a manipulation that is known to differentially activate top-down feedback (Bar et al., 2006). The rationale was as follows. If context effects are triggered by feedback connections involving the magnocellular pathway (Kveraga et al., 2007), removing low-spatial-frequency information with a high-pass filter should diminish the match effect.

Participants read sentences that implied a specific shape of a mentioned object, followed by a pictured object. They had to decide if the target object was mentioned in the sentence or not. Two picture variants A and B that showed the object in different shapes were used (e.g. a sitting or a flying duck). Each picture was paired with three sentences, sentence A implied the same shape as depicted in the picture A (e.g. “the ranger saw a duck in the lake”), sentence B implied the same shape as depicted in picture B (e.g. “the ranger saw a duck in the air”), or an unrelated sentence not mentioning the depicted object at all (e.g. “the ranger buttered his bread”). Previous research used the interaction between picture-variant and sentence as an index of match effect (Stanfield and Zwaan, 2001; Zwaan et al., 2002). The unrelated condition was added to test whether responses are modulated in match, mismatch or both conditions.

Data were analyzed using mixed-effects models, with crossed-random effects for subjects and items (Baayen et al., 2008; Jaeger, 2008). These analyses replace the traditional separate subject and item analysis, as they allow the simultaneous modeling of both. Binomial mixed-effects models are more accurate for error-data than analysis of transformed error rates (Jaeger, 2008). Our analysis proceeded in two steps. First, only trials affording yes-responses were compared, leaving out the unrelated sentences. A model was fitted with *sentence* (A, B, vs. unrelated), *picture* (variant A vs. variant B), *filter-condition* (high-pass, low-pass), and their interactions to the latency and error data. A *repetition* (first to sixth block) effect was included to capture longitudinal effects of familiarization with the material. Next, the latency data were modeled collapsing the factors *sentence* and *picture*, into the factor *condition* (match, mismatch, and unrelated) that was entered using treatment coding.

2. Results

Fillers and warm-up items were removed from the data. For the reaction-time analysis incorrect responses (7.12%) and extreme outliers (trials slower than mean reaction time plus 4× sd) were excluded from the analysis (1.11%). Reaction times were log-transformed to achieve normality (Baayen et al., 2008).

Analysis of the errors showed no speed-for-accuracy-trade-off. Instead, reaction times were faster in correct than to incorrect responses (638 ms vs. 818 ms; beta=0.02; t=8.93; $p_{MCMC}<0.001$). The error rates differed between conditions (Table 1). Reaction times and errors were thus analyzed separately.

A linear mixed model with *repetition*, *sentence*, *picture*, *filter*, and the interaction between *sentence*, *picture* and *filter* was fitted to the latency- and error-data (Table 2). There were no main effects for filtering, indicating, that the two filtered pictures were processed similarly. Reaction times were faster for later repetitions (beta=−0.1; t=−53.75; $p_{MCMC}<0.001$). Importantly there was a significant triple interaction between *sentence*, *picture* and *filter* (beta=−0.07; t=−2.49; $p_{MCMC}=0.01$). The match effect was modulated by the frequency information contained in the picture. Analyzing the two filter conditions separately showed a strong interaction between *sentence* and *picture* for low-pass filtered targets (beta=−0.09; t=−4.89; $p_{MCMC}<0.001$) indicating shorter response latencies when sentence and picture matched than when they mismatched. This interaction was not significant for high-pass filtered targets (beta=−0.02; t=−1.31; $p_{MCMC}>0.1$). The match effect was only apparent when targets contained low-spatial frequencies (Fig. 1). Error rates showed fewer errors to later *repetitions* (beta=−0.14; z=−6.9; p<0.001), and a similar processing advantage when picture and sentence matched (beta=−0.62; z=−3.11; p<0.01). Furthermore, we found more errors to B-sentences (beta=0.31; z=−2.12; p<0.001) and B-pictures (beta=0.43; z=1.83; p=0.07). Error rates were not affected by *filter* conditions, nor were there any interactions (all p>0.4).

The next analysis investigated whether the modulations of the latency effect were due to changes in match, mismatch or both conditions (Table 3). There were no main effects for *filter*, and responses were faster to later *repetitions* (beta=−0.10; t=−65.12), $p_{MCMC}<0.001$). Both *conditions* requiring a yes response (match and mismatch) were faster than the unrelated baseline (all betas<−0.1; t<−16; $p_{MCMC}<0.001$). The interaction between *filter* and *condition* was significant (beta=−0.04; t=−3.07; $p_{MCMC}<0.01$), showing larger differences between the match and unrelated condition for low-pass pictures (about 130 ms) than for high-pass-filtered pictures (about 100 ms). The corresponding differences between the mismatch and unrelated conditions were not significantly different for low- vs. high-pass-filtered targets (p=0.65). This argues against a disadvantage for mismatching targets that is stronger for low-pass compared to high-pass-filtered pictures. Thus, in addition to generally faster yes-responses, a specific advantage for matching contexts could be shown only when targets contained low-spatial frequencies.

3. Discussion

The aim of this study was to test the neural mechanism that underlies the match effect in sentence–picture verification (Stanfield and Zwaan, 2001; Zwaan et al., 2002). Specifically, we investigated the impact of high- and low-spatial-frequency information contained in the target on the context effects. Our results show a match effect in latencies and errors that was modulated by the spatial-frequency information in the targets. While there was a strong match effect for low-pass-filtered targets, this effect was not significant for high-pass filtered targets. Comparing the match and mismatch conditions to an unrelated baseline, we found no baseline differences between the filter conditions, but a specific facilitatory effect for low-pass-filtered targets in the match condition.

Table 1 – Reaction times and error rates.

Sentence — variant		High-pass filtered		Low-pass filtered	
		Picture A	Picture B	Picture A	Picture B
A	RT ms/sd	664/355	671/349	624/344	660/373
	Errors %	6.75	8.72	7.89	10.10
B	RT ms/sd	674/364	656/328	650/366	620/314
	Errors %	7.89	6.75	10.40	8.48
Unrelated	RT ms/sd	750/352	780/382	732/342	771/364
	Errors %	4.96	4.90	4.42	4.90

Note. The match condition consists of sentence A–picture A, and sentence B–picture B trials; the mismatch condition consists of sentence A–picture B and sentence B–picture A trials.

Table 2 – Summary of mixed models predicting yes-responses.

Parameter	Latency				Errors			
	Estimate	SE	t	p_{MCMC}	Estimate	SE	z	P
Intercept	6.76	0.04	151.34	<0.001	−2.86	0.23	−12.52	<0.001
Sentence (b)	0.01	0.01	0.85	0.41	0.31	0.15	2.12	<0.001
Picture (b)	0.01	0.03	0.46	0.63	0.43	0.24	1.83	0.07
Filter (low)	−0.06	0.06	−0.94	0.29	0.25	0.31	0.82	0.41
Repetition	−0.10	0.00	−53.75	<0.001	−0.14	0.02	−6.90	<0.001
Sentence (b) : picture (b)	−0.02	0.02	−1.31	0.20	−0.62	0.20	−3.11	<0.01
Sentence (b) : filter (low)	0.03	0.02	1.61	0.11	0.06	0.20	0.30	0.76
Picture (b) : filter (low)	0.04	0.04	0.90	0.32	−0.03	0.33	−0.10	0.92
Sentence (b): picture (b): filter (low)	−0.07	0.03	−2.49	0.01	0.04	0.27	0.13	0.90

Note. The match-effect is reflected by the interaction between sentence and picture variant.

Overall, these data clearly confirm our hypothesis about the role of reentrant top-down-feedback mechanisms for the match effect in the sentence–picture-verification task. The experiment also revealed two unexpected findings. First, the match effect was at least numerically larger in the low-pass condition than in a pretest using unfiltered targets. This fits with ERP-research showing that effects of prior knowledge on visual processing can be enhanced when low-pass filtering the targets (Abdel Rahman and Sommer, 2008). Second, some of our analyses revealed differences between sentence versions. We did not expect any systematic differences between the sentence or picture versions, as they were arbitrarily chosen when designing the experiment. We believe that these differences reflect the tendency on behalf of the experimenters to assign the more prevalent sentence variant with A. Note that the design controls for any confounding main effects of sentence variant and that the critical index for a context effect, the interaction between sentence and picture, and its modulation by frequency was reliable above and beyond such differences.

The mechanism we propose for the linguistic-context effects parallels the model of visual object recognition in context (Bar, 2004). In this view, linguistic contexts can amplify subsequent visual processing (Hirschfeld et al., 2011) because of reentrant top-down feedback connections from frontal cortex to visual processing in the fusiform gyrus (Ganis and Schendan, 2008). This is triggered by low spatial frequencies processed via the magnocellular pathway. The critical difference between the proposed model and repetition-suppression explanations that also involves top-down feedback is that the latter only assume top-down feedback taking place during language comprehension proper (Stanfield and Zwaan, 2001). As the previous behavioral and electrophysiological research used priming procedures that only provide evidence as to the end-product of the sentence comprehension process, much work is needed to investigate what is happening during listening. Neuroimaging studies have demonstrated activations of visual processing regions during conscious imagery (Kosslyn et al., 2001) and sentence–picture matching (Carpenter et al., 1999), In addition, the inferior frontal gyrus has been implicated in conceptual-verification tasks (Kan et al., 2003). Due to the low temporal resolution of these methods, it is unclear whether these regions are important during listening or matching. For example, while left-inferior-frontal-gyrus activations are in line with localizations of MEG effects, no anterior cingulate activation has been observed. Investigating connectivity patterns during listening may be one way to elucidate this in more detail. The reentrant feedback model explains what happens during the processing of the target picture, and this neurocognitive mechanism is an important alternative to the implicit passive repetition explanation for the match effect, both empirically as well as theoretically.

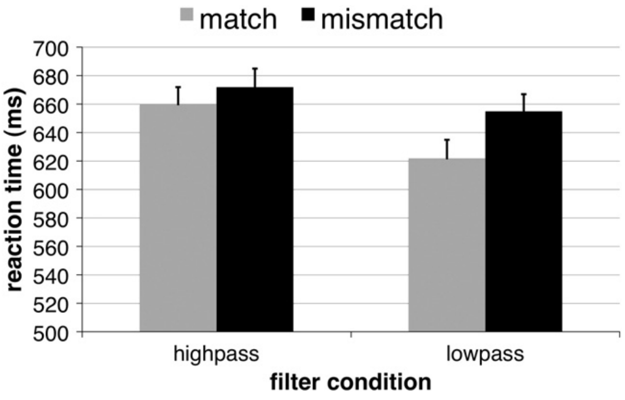

Fig. 1 – Reaction times in the two filter conditions for matching (sentence A–picture A and sentence B–picture B) and mismatching (sentence A–picture B and sentence B–picture A) trials. Error bars indicate 95% CI.

Table 3 – Summary of mixed models comparing the unrelated baseline to the match and mismatch condition.

Parameter	Estimate	SE	t	p_{MCMC}
Intercept	6.91	0.04	165.72	<0.001
Condition (match)	−0.16	0.01	−17.54	<0.001
Condition (mismatch)	−30.15	0.01	−16.28	<0.001
Filter (low)	−0.02	0.06	−0.35	0.72
Rep	−0.10	0.00	−65.12	<0.001
Filter (low): condition (match)	−0.04	0.01	−3.07	<0.01
Filter (low): condition (mismatch)	−0.01	0.01	−0.45	0.65

Note. The sentence and picture conditions were combined to form the match (sentence A–picture A and sentence B–picture B) and mismatch (sentence A–picture B and sentence B–picture A) condition.

The proposed model can integrate behavioral and electrophysiological findings into the neural basis of the context effects in sentence–picture verification. The effect is not apparent in patients with right-hemisphere damage (Lincoln et al., 2008), because they process low-spatial frequencies differently (Grabowska and Nowicka, 1996). Furthermore, this mechanism predicts that only features processed via the magnocellular pathway facilitate recognition in the sentence–picture paradigm. It thus can account for the unexpected finding that providing matching color-information impairs visual recognition (Connell, 2007). While previous explanations for this reversal of effects have focused on general differences in color and shape processing (Aginsky and Tarr, 2000), the top-down feedback mechanism proposed here directly predicts the effects should be smaller.

At the theoretical level, the importance of feedback connections for the context effects fits nicely with recent advances in theorizing about embodied cognition. Barsalou (2009) has suggested that the main function of the perceptual-symbol system is to generate predictions. Behavioral and electrophysiological experiments have shown that embodied simulations are used to predict upcoming words during sentence comprehension (Chwilla et al., 2007; Glenberg and Robertson, 2000). Our data can be seen as first evidence that the effects of language comprehension on visual processing invoke the same neural networks that are considered critical for the generation of predictions in visual perception (Bar, 2007). Such top-down connections are also central in computational models that aim to explain attention effects in visual-search tasks (Hamker, 2005). Rather than implying that embodiment effects are merely due to attention, we wish to emphasize that effects of language on visual processing as well as of attention reflect emergent properties of the brain (Gazzaniga, 2010).

To sum up, the current study demonstrates that low-spatial-frequency information is rapidly processed and important for effects of sentential context on picture recognition. Context effects in the sentence–picture task were diminished when these spatial frequencies were removed. The model we propose is in accordance with empirical data on sentence–picture verification, and fits with an emerging understanding of brain function. To conclude, the embodiment effects in the sentence–picture task engage more complex processing steps than previously acknowledged.

4. Experimental procedures

4.1. *Participants*

62 students from the University of Münster participated in this experiment, 31 in each filter condition.

4.2. *Materials*

108 pictures of 54 objects served as targets. Images were taken from a large collection of object photographs (Hemera Photo Objects). Objects were photographed without background. All images had a resolution of 75 pixels/in., scaled to fill a 400×400 pixel square that measured approx. 9°×9° visual angle in the experiment. Target pictures were filtered using either a high-pass filter to remove low-spatial frequencies, or using a low-pass filter to remove high-spatial frequencies. Filtering was performed with Adobe Photoshop 7 (Adobe Systems, San Jose, CA). In the high-pass condition, we set the high-pass filter to a radius of 0.1 to 0.3 pixels resulting in images filtered at 46 cycles per image. For low-pass filtering we used a Gaussian blur filter with a 6.1 pixel kernel (approx. 9 cycles per image (Bar et al., 2006). Filtering was followed by an adjustment of luminance and contrast to make them equally easy to perceive. The fact, that we found no main-effect *filtering* on response latencies and error rates shows that this was successful.

Each object was pictured in two states that differed in global shape of the object, e.g. variant A showed a flying duck, variant B a sitting duck. For each picture four context sentences were created; sentence A implied the object in the same state as depicted in picture A (e.g. “The ranger saw a duck in the air”), sentence B implied the object in the alternative state that was depicted in picture A (e.g. “The ranger saw a duck in the lake”), and two unrelated sentences that did not mention the object at all. We included the unrelated sentences as a baseline for the two picture versions to assess whether the match effects are due to changes in both or only one sentence variant. An additional 210 pictures and 102 sentences served as fillers. Sentences were presented in upper-case letters, black against a light grey background.

4.2.1. *Pretest without filtering*

Sentences and pictures were tested for the occurrence of the match effect. An additional 32 participants completed the experiment with unfiltered target pictures. Analysis of the errors showed no speed-for-accuracy-tradeoff, but faster reaction times to correct compared to incorrect responses (723 ms vs. 1080 ms; beta=0.25; t=19.59; $p_{MCMC}<0.001$). A linear mixed model predicting the yes-responses with repetition, sentence, picture, and the interaction between sentence and picture was fitted to the latency data. Reaction times were significantly faster in later repetitions (beta=−0.095, t=36.73; $p_{MCMC}=0.001$). The interaction between sentence and picture (beta=−0.06; t=−3.43; $p_{MCMC}=0.001$) indicated faster responses when sentence and picture matched. Responses to trials involving sentence-variant B were slower than to variant A (beta=0.041; t=3.28; $p_{MCMC}=0.001$). Participants' errors were not predicted by repetition, but showed a strong interaction between sentence and picture (beta=-1.35; z=-6.57; $p_z<0.001$). Participants made more errors to sentence-variant B (beta=0.79; z=5.37; $p_z<0.001$), and picture-variant B (beta=0.74; z=3.65; $p_z<0.001$). These results clearly replicate previous behavioral effects with line-drawings (Stanfield and Zwaan, 2001; Zwaan et al., 2002) and demonstrate that our sentences and photorealistic pictures can be used as targets.

4.3. *Design and procedure*

Each trial began with the presentation of a sentence in the middle of the screen for 1200 ms. Next a fixation cross appeared for 350 ms, followed by the target that disappeared as soon as the participants pressed a button. Inter-trial interval was 2250 ms.

The six occurrences of one object (two picture variants, each presented in three contexts) were distributed across six blocks. Each block contained 54 experimental items of which

36 required a “yes” response and 18 requiring a “no” response. To balance the “yes” and “no” responses 18 filler items were added, that also afforded a “no” response. Each block contained 72 items. Each participant saw all six blocks, with self-timed breaks after the second, and fourth blocks. Before the first block and after the breaks, three practice trials were included. The order of blocks was balanced across participants using a Latin-Square. The order of trials within each block was randomized for each individual participant. Participants were tested in a quiet computer pool in groups of maximal 16 persons.

Acknowledgment

We thank Thomas Hösker for help in conducting the experiment.

REFERENCES

Abdel Rahman, R., Sommer, W., 2008. Seeing what we know and understand: how knowledge shapes perception. Psychological Bulletin and Review. 15, 1055–1063.

Aginsky, V., Tarr, M.J., 2000. How are different properties of a scene encoded in visual memory? Visual Cognition. 7, 147–162.

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, 390–412.

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

Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. Trends in Cognitive Sciences 11, 280–289.

Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmidt, A.M., Dale, A.M., Hamalainen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., Halgren, E., 2006. Top-down facilitation of visual recognition. Proceedings of the National Academy of Sciences of the United States of America 103, 449–454.

Barsalou, L.W., 1999. Perceptual symbol systems. Behavior Brain Sciences 22, 577–660.

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

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

Brunye, T.T., 2009. When you and I share perspectives. Psychological Science 20, 27–32.

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

Carpenter, E., Just, M.A., Keller, T.A., Eddy, W.F., Thulborn, K.R., 1999. Time course of fMRI-activation in language and spatial networks during sentence comprehension. Neuroimage 10, 216–224.

Chwilla, D.J., Kolk, H.H., Vissers, C.T.W.M., 2007. Immediate integration of novel meanings: N400 support for an embodied view of language comprehension. Brain Research 1183, 109–123.

Connell, L., 2007. Representing object colour in language comprehension. Cognition 102, 476–485.

Dijkstra, K., Yaxley, R.H., Madden, C.J., Zwaan, R.A., 2004. The role of age and perceptual symbols in language comprehension. Psychology and Aging 19, 352–356.

Dove, G., 2009. Beyond perceptual symbols: a call for representational pluralism. Cognition 110, 412–431.

Ganis, G., Schendan, H.E., 2008. Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. Neuroimage 42, 1714–1727.

Gazzaniga, M.S., 2010. Neuroscience and the correct level of explanation for understanding mind: an extraterrestrial roams through some neuroscience laboratories and concludes earthlings are not grasping how best to understand the mind–brain interface. Trends in Cognitive Sciences 14, 291–292.

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

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

Grabowska, A., Nowicka, A., 1996. Visual–spatial-frequency model of cerebral asymmetry: a critical survey of behavioral and electrophysiological studies. Psychological Bulletin 120, 434–449.

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends in Cognitive Sciences 10, 14–23.

Hamker, F.H., 2005. The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. Cerebral Cortex 15, 431–447.

Hauk, O., Johnsrude, I., Pulvermuller, F., 2004. Somatotopic representation of action words in human motor and premotor cortex. Neuron 41, 301–307.

Henson, R., 2003. Neuroimaging studies of priming. Progress in Neurobiology 70, 53–81.

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, 91–96.

Jaeger, T.F., 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language 59, 434–446.

Kan, I., Barsalou, L., Solomon, K., Minor, J., Thompson-Schill, S., 2003. Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. Cognitive Neuropsychology 20, 525–540.

Kosslyn, S.M., Ganis, G., Thompson, W.L., 2001. Neural foundations of imagery. Nature Reviews. Neuroscience 2, 635–642.

Kveraga, K., Boshyan, J., Bar, M., 2007. Magnocellular projections as the trigger of top-down facilitation in recognition. The Journal of Neuroscience 27, 13232–13240.

Lincoln, A.E., Long, D.E., Swick, D., Larsen, J., Baynes, K., 2008. Hemispheric asymmetries in the perceptual representation of words. Brain Research 1188, 112–121.

Madden, C., Therriault, D., 2009. Verb aspect and perceptual simulations. The Quarterly Journal of Experimental Psychology 62, 1294–1302.

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

Noordzij, M.L., van der Lubbe, R.H.J., Postma, A., 2005. Strategic and automatic components in the processing of linguistic spatial relations. Acta Psychologica 119, 1–20.

Pinker, S., 1994. The language instinct: how the mind creates language. HarperCollins, NY.

Speer, N., Reynolds, J., Swallow, K., Zacks, J., 2009. Reading stories activates neural representations of visual and motor experiences. Psychological Science 20, 989–999.

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

Zwaan, R., Stanfield, R., Yaxley, R., 2002. Do language comprehenders mentally represent the shapes of objects? Psychological Science 13, 168–171.