Running head: INDEPENDENCE OF EARLY CONCEPTUAL AND LINGUISTIC PROCESSES

# Early conceptual and linguistic processes operate in independent channels

Ansgar D. Endress & Mary C. Potter
Massachusetts Institute of Technology, Cambridge, MA

Ansgar D. Endress
Massachusetts Institute of Technology
Department of Brain & Cognitive Sciences, Room 46-4127
43 Vassar Street, Cambridge, MA 02139
E-mail: ansgar.endress@m4x.org

**Abstract**

Language and concepts are intimately linked to each other, but how do they interact? Here, we probe the relation between conceptual and linguistic processing at the earliest processing stages. We presented observers with sequences of visual scenes at 200ms per scene. Results show that observers understood and remembered the scenes' abstract gist and, therefore, their conceptual meaning. However, they remembered the scenes at least as well when simultaneously performing a linguistic secondary task (i.e., reading and retaining sentences); in contrast, a nonlinguistic secondary task (equated for secondary task difficulty) impaired performance on the scenes. Further, encoding scenes interfered with the nonlinguistic secondary task and vice-versa, while scene processing and the linguistic secondary task did not affect each other. At the earliest stages of conceptual processing, the extraction of meaning from visually presented linguistic stimuli and of conceptual information from the world take place in remarkably independent channels.

## Early conceptual and linguistic processes operate in independent channels

Language and concepts are intimately linked to each other. For example, conceptual real-world knowledge, or even just seeing visual arrays of objects, can affect how we initially interpret the grammatical structure of sentences that refer to those objects (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Tanenhaus, & Garnsey, 1994, but see Rayner, Carlson, & Frazier, 1983; Clifton et al., 2003). Further, the semantic or conceptual meaning of words can even affect low-level perception. For example, when listening to verbs describing upward motion, observers are impaired in detecting actual downward motion, and vice-versa (e.g., Meteyard, Bahrami, & Vigliocco, 2007; for effects of language on visual processes such as attention and search, see, among many others, Huettig & Altmann, 2007; Spivey, Tyler, Eberhard & Tanenhaus, 2001).

Such results raise the question of whether processes that derive meaning from sensory data, be they linguistic or nonlinguistic, rely on shared mechanisms that are interdependent at all levels, from the lowest levels of, say, motion perception to the highest level of actually representing meaning. Different research traditions offer a spectrum of positions on this venerable question. Traditions affirming such an interdependence include the “Whorfian” view that language constrains the concepts and percepts we can entertain (Whorf, 1956), and the “embodied”, “simulationist” view that understanding any concept involves mentally simulating its referent (e.g., we might understand the meaning of “upward” by mentally simulating what upward motion looks like; see e.g. Barsalou, 1999). Other authors hold that conceptual and linguistic information are processed by completely modular and encapsulated processors (e.g., Fodor, 1983; Pylyshyn, 1999). In between, still other authors propose that linguistic stimuli are analyzed by dedicated processors, but that these processors can also incorporate nonlinguistic information when available (Tanenhaus et al., 1995; Trueswell et al., 1994).

While these results suggest that conceptual and linguistic processing interact in important ways, they do not address the question of whether the underlying processors are shared. In fact, in previous experiments, both the linguistic and nonlinguistic information mapped onto related meanings. Consequently, conceptual information derived from linguistic or nonlinguistic sources provided a prior context, which might have exerted top-down effects on both linguistic and nonlinguistic processes, without these processes being shared or identical. As top-down effects have been observed even at the level of the thalamus (O’Connor, Fukui, Pinsk & Kastner, 2002), such effects could occur at the earliest processing stages. For example, in Meteyard et al.’s (2007) studies, participants continuously listened to verbs representing a direction of motion at a rate of 1/s; motion was displayed only during randomly spaced periods of 150ms. Hence, listening to upward or downward verbs might have placed participants in upward or downward "mind-sets," where thinking about upward and downward motion might have influenced their motion perception. To test the interdependence of linguistic and nonlinguistic processes, one needs, therefore, to use a situation where top-down effects are precluded, and where one kind of process cannot establish a prior context for the other one.

In the following experiments, we probe the relation between language and nonlinguistic concepts at the earliest processing stages. We preclude top-down effects by having both kinds of information processed simultaneously and under time pressure, and by having the two kinds of processes analyze information that is largely unrelated. In each trial, participants viewed a sequence of six unrelated scenes presented at a rate of 200ms or 250ms per scene (see Figure 1). Observers can encode visual scenes presented at this rate at a rather abstract conceptual level. Not only do they succeed on recognition tests of the scenes (Potter, Staub, Rado, & O'Connor, 2002), but they succeed even when tested on verbal labels for the scenes; for example, they can decide whether they have seen a scene corresponding to the description "people in street" (Potter, Staub, & O'Connor, 2004). In other words, people extract not only low-level visual information as in traditional visual short-term memory studies, but also the conceptual gist of the scenes (e.g., Intraub, 1981; Potter, 1976).

After having seen the six rapidly presented scenes, participants completed a yes-no recognition test. In a random half of the trials, participants were shown 10 scenes one at a time (half had appeared in the sequence, and half were new). In the remaining trials, participants were shown 10 verbal labels for scenes (half corresponding to the scenes they had seen, half new), and had to decide whether they had seen a scene corresponding to the labels.

In Experiment 1, we establish that observers can extract the gist of scenes at a rate of 200ms/picture, as in previous experiments investigating conceptual short-term memory (Potter et al., 2002, 2004). In Experiment 2, we test whether a linguistic secondary task interferes with scene memory. Specifically, a written word was presented in the center of each scene, the sequence of six words forming a sentence that was syntactically acceptable but made little sense, such as “miners duly locate truly tired ladies.” Such sentences are likely to trigger linguistic processing as shown in earlier studies using RSVP sentences, where words were presented one by one at rates of up to 12 words/s (Potter, Kroll, & Harris, 1980; Potter & Lombardi, 1990). Following each sequence, participants were either tested on their memory for the scenes or for the sentence.

In Experiment 3, we ask whether a nonlinguistic secondary task interferes with scene memory. The center of each scene contained a small box with grid lines. Participants were instructed to press a key when they detected a change in the density of the grid lines. The box was shown in Experiment 1 as well, but participants were simply told to look at it. Experiment 4 replicates Experiment 2, but using semantically more sensible sentences; Experiments 5a and 5b provide additional controls.

## General method

### Participants

Ninety-six native speakers of English (55 women, 41 men, mean age 23.3) from the MIT community were included in the analyses of the primary task of each of Experiments 1 to 5b (16 per experiment). No participant took part in more than one experiment reported here.

### Stimuli

As described in Potter et al. (2004), scenes were colored photographs collected from the World Wide Web and commercial sources, and the labels corresponding to the pictures were generated by two research assistants. Scenes (and the corresponding labels) were randomly organized into sets of 11 pictures (6 RSVP items and 5 “new” items), with the constraint that the items in a set had no obvious relation to each other.

The box at the center of the pictures in Experiments 1 and 3 had a size of 35×35 pixels, and contained equally spaced horizontal and vertical gridlines. The box appeared in synchrony with the scenes. The line density changed equally often on the 2nd, 3rd, 4th and 5th scene, changing back to the original density on the following scene.

Sentences in Experiment 2 were composed according to 10 different grammatical templates by drawing quasi-randomly from lists of words in the relevant form classes (nouns, verbs, adverbs, adjectives, and some prepositions) and made little sense. Words were reasonably frequent and had 2 syllables and 4 to 6 letters. Words in change trials were selected using the same criteria. In Experiment 4 the sentences were more meaningful; the only constraints imposed on the words were to be reasonably frequent and to have at most 8 letters.

## Procedure

Each experiment comprised 80 trials. Trials started with a central fixation cross, followed by 6 visual scenes (200ms per scene). In Experiments 1 and 3, the center of the scenes contained the small box with grid lines. In Experiment 1, participants were instructed to look at the box, and to remember the scenes. In Experiment 3, participants were instructed to press a key when they detected a change of the density of the lines inside the box. Before starting the experiment, participants received four practice trials where only the small box was presented, without any scenes.

Following the 6 scenes, participants were tested for their recognition memory. In a random half of the trials, participants were tested on visual scenes; in the remaining trials, they were tested on labels for the scenes. The sixth scene was never tested because it was unmasked and therefore easily remembered (Potter et al., 2002). Participants were tested on 10 items (scenes or labels), half of which had appeared in the scene sequence, and half of which were new. No picture occurred in more than one trial. Responses were collected from pre-marked “Yes” and “No” keys on the keyboard. Participants received four practice trials before starting the experiment.

In Experiment 2 and 4, the squares in the center of the scenes were replaced by a box in which each word of a six word sentence was presented, word by word, synchronized with the onset of the scenes.

In half of the trials of Experiments 2 and 4, participants were tested on their memory for the sentences; they saw an entire sentence on the screen, and had to decide whether or not a word had been changed; on half of these trials, one word was replaced by a new word, preserving grammaticality. In the remaining trials, participants were tested on their recognition of scenes and scene labels as in Experiments 1 and 3.

To analyze an equal number of scene test trials in Experiments 1--4, we considered only those trials in Experiment 1 and 3 in which the line density of the square in the center of the scenes did not change. Data were analyzed in terms of five relative test positions of the five old pictures, matched with the five distractor pictures (rather than the 10 absolute test positions), using a repeated-measures ANOVA with the within-subject predictors relative test position and test modality (scene vs. scene label).

## Experiment 1

Results and discussion

The results of Experiment 1 are shown in Figure 2a and Table 1. Participants successfully remembered scenes. In this and all other experiments, they performed better when tested on scenes than when tested on labels, presumably because scenes provide participants with visual and conceptual information in addition to the gist of the scenes (the only information carried by the labels). Replicating earlier work (Potter et al., 2002, 2004), participants performed worse in later test positions, probably due to decay or interference; this was the case in all other experiments as well. However, both when tested on scenes and on scene labels, participants performed significantly above chance in all positions, again in this and all other experiments. As Figures 2 and 4a and Tables 1 show, analogous effects of the test modality and position were observed in each of the other experiments, but will not be reported in the text.

## Experiment 2

Experiment 2 addresses the question of how linguistic and conceptual mechanisms interact at early processing stages by presenting a word in the center of each scene. The six words formed a sentence that was syntactically acceptable but made little sense, such as “miners duly locate truly tired ladies.” For a random half of the trials, participants were tested on scenes or scene labels. In the other trials, participants were tested on their memory for the sentences.

If scene understanding involves linguistic resources, we would expect a large decrement in scene memory between Experiments 1 and 2. In contrast, if we grasp the conceptual meaning of scenes by virtue of nonlinguistic mechanisms, we would expect only a limited decrease in scene memory due to the attentional demands of performing two tasks simultaneously, or even no decrease at all.

## Results and discussion

In the secondary task in Experiment 2, participants successfully detected changed words (Figure 3 and Table 2). Scene recognition performance is shown in Figure 2b, and was compared between Experiments 1 and 2 using a logistic mixed-effects model (Table 3). While there was no main effect of Experiment, participants performed numerically better in Experiment 2 despite their secondary task. Scene recognition performance was better for scenes and for earlier test positions, and the separation between the two test modalities diminished for later test positions. As shown in Table 3, analogous effects of the test modality and position were observed in each of the other between-experiment comparisons, but will not be reported in the text.[1]

Compared to Experiment 1, Experiment 2 not only revealed no decrease in performance: participants performed numerically (if not significantly) better than in Experiment 1, even though they had to read sentences in addition to monitoring the scenes. While previous research has shown that, at least after massive training, some types of natural scene processing can occur with very limited attentional involvement (Li, VanRullen, Koch, & Perona, 2002; Fei-Fei, VanRullen, Koch, & Perona, 2005; but see Yi, Woodman, Widders, Marois, & Chun, 2004), one would expect performance in Experiment 2 to be worse than in Experiment 1, simply because participants had to complete two tasks rather than one. However, if scene understanding and language rely on disjoint sets of processes, participants might complete both tasks, without any detrimental effect of one task on the other.

In Experiment 3, we further explored the question of why scene understanding was not impaired by the linguistic secondary task of Experiment 2. As in Experiment 2, participants completed a secondary task; in contrast to Experiment 2, however, this task was nonlinguistic in nature. If scene understanding is simply unaffected by secondary tasks, we would expect to replicate the results of Experiment 2, observing no impairment in scene recognition. In contrast, scene recognition performance might be affected by <u>nonlinguistic</u> secondary tasks that tap into processes required for scene understanding (e.g., visual processing), even if scene recognition performance is not affected by linguistic secondary tasks.

## Experiment 3

Participants were presented with rapid sequences of 6 scenes. As in Experiment 1, the center of each scene contained a small box with grid lines. Participants were instructed to press a key when they detected a change in the density of the grid lines. Recognition of the scenes was tested as before.

## Results and discussion

In the secondary task in Experiment 3, participants detected density changes in the small box (Figure 3 and Table 2); this performance did not differ from that on the secondary task in Experiment 2, $F(1,35)=2.3$, $p=.137$, $\eta^2=.062$, although the sentence task in Experiment 2 was numerically harder.

Scene recognition performance is shown in Figure 2c, and was compared between Experiments 1 and 3 using a logistic mixed-effects model (Table 3). Unsurprisingly, given that participants had to perform a secondary task in Experiment 3 but not in Experiment 1, participants performed worse in Experiment 3 than in Experiment 1. These results contrast with the comparison of Experiments 1 and 2, where participants performed numerically (if not significantly) better in Experiment 2 although they had to complete a secondary task. We surmise that the crucial difference between Experiments 2 and 3 is that participants completed a linguistic secondary task in Experiment 2 and a visual-attention secondary task in Experiment 3, and that some mechanisms involved in the visual task, but not language, are needed to understand scenes. Accordingly, participants performed significantly better in Experiment 2 than in Experiment 3 (Table 3).

## Experiment 4

Although the difficulty of the secondary tasks in Experiments 2 and 3 was matched in terms of task performance (at least when each secondary task was presented with the same primary task, remembering scenes), participants were significantly better at recognizing scenes in Experiment 2 compared to Experiment 3, suggesting that language processing is largely independent of scene comprehension. It is possible, however, that participants did not fully process the nonsense sentences used in Experiment 2. In Experiment 4, we control for this possibility by replicating Experiment 2, but using simple, semantically interpretable six word sentences (e.g., “Carol rants about the lousy food”). These more interpretable sentences should be more likely to trigger normal sentence processing than the less meaningful sentences in Experiment 2.

### Results and discussion

In the secondary task of Experiment 4, participants successfully detected changed words (Table 2). The secondary task performance was better than in Experiment 2, $F(1,35)=9.1$, $p=.005$, $\eta^2=.207$, and than in Experiment 3, $F(1,32)=4.2$, $p=.049$, $\eta^2=.116$. Scene recognition performance in Experiment 4 is shown in Figure 2d, and was significantly better than in Experiment 3, but not compared to Experiment 2 (Table 3).[2] Thus, making the sentences more normal and meaningful did not increase interference with picture processing.

A plausible conclusion from these data is that linguistic tasks involve processes that are independent from those involved in scene understanding. Alternatively, such tasks might prevent counterproductive verbal strategies that participants might use to remember the scenes. Participants sometimes report trying to find labels for the scenes, thereby occupying resources that would no longer be available to encode the scenes. A similar observation has been made in experiments where participants had to keep faces or colors in long-term memory; when instructed to verbally describe the face or the color during a retention period of several minutes, their recognition performance was substantially impaired compared to various control tasks that did not involve verbalization of the stimuli (Schooler & Engstler-Schooler, 1990). Similarly, a secondary language task of the sort used in Experiments 2 and 4 might inhibit counterproductive verbal strategies, whereas a nonlinguistic secondary task would show the usual negative effect of having a secondary task.

Preventing verbal strategies might, therefore, offset the attentional costs associated with performing a secondary task. However, Experiments 1 to 4 might have encouraged such strategies, because participants were tested not only on actual visual scenes, but also on verbal labels for the scenes. Possibly, a linguistic secondary task might reveal interference with scene understanding if participants' performance on scenes had been tested only with actual scenes, but not with verbal labels. We tested this possibility in Experiments 5a and 5b.

## Experiment 5

Experiments 5a and 5b are replications of Experiments 2 and 3, respectively, with three crucial changes. First, and most importantly, participants were never tested on verbal labels for the scenes, but only on actual visual scenes. As a result, the test items should no longer encourage a verbal memory strategy for the scenes. Second, we made the two secondary tasks more similar. In Experiment 5a, participants read the same sentences as in Experiment 2, again presented word by word in the center of a scene. In a random half of the trials, participants were then tested on single words; that is, they had to decide whether or not a test word had occurred in the sentence (on half of the trials it had). On the other half of the trials, they were tested on scene memory. As in Experiment 3, participants in Experiment 5b had to detect changes of the density of lines in a small square; however, rather than pressing a key as soon as they saw a density change, on half the trials they had to report after the trial whether or not a density change had occurred. In the remaining trials they were tested on their memory for the scenes. Third, after they had completed the experiments, participants were tested on the secondary task in isolation, with no primary task and no scenes being shown. In addition to these changes, in Experiment 5 we increased the presentation duration to 250ms per picture.[3]

Results and discussion

In the secondary task in Experiment 5a, participants successfully discriminated words that had occurred in the sentence from words that had not (Table 2 and Figure 4b). (Below, we will return to the performance on the secondary task when it was presented as the only and primary task.) In the secondary task in Experiment 5b, participants detected density changes in the small box (Figure 4b and Table 2); this performance did not differ from that on the secondary task in Experiment 5a, $\underline{F}(1,32)=.6$, $\underline{p}=.447$, $\underline{\eta}^2=.018$, although the sentence task in Experiment 5a was numerically harder.

Scene recognition performance in Experiments 5a and 5b is shown in Figure 4a; the results were compared using a logistic mixed-effects model. Crucially, and replicating the results of Experiments 2 and 3, participants performed better in Experiment 5a than in Experiment 5b (Table 3), suggesting that the nonlinguistic secondary task of Experiment 5b interfered more with scene understanding than the linguistic secondary task of Experiment 5a, even though the two secondary tasks were equally difficult.

While the two secondary tasks were matched for difficulty when used as secondary tasks, one task might be easier than the other when tested in isolation, without a primary task. To address this question, participants in Experiments 5a and 5b completed their respective secondary tasks without any interfering primary tasks after having finished the main experiment. As shown in Figure 4b and

Table 2, performance on the linguistic secondary task was similar when used as a secondary task and when it was presented in isolation. In contrast, performance on the change detection task was almost perfect in the absence of interfering scenes. An ANOVA (excluding two participants, one in each experiment, who did not complete the second presentation of the secondary tasks) with the between-subjects predictor task type (nonsense sentences vs. change detection) and the within-subject predictor task status (secondary task vs. sole task) revealed main effects of both the task type, $F(1,30)=4.4$, $p=.044$, $\eta^2_p=.129$, and the task status, $F(1,30)=29.2$, $p<.0001$, $\eta^2_p=.438$. Importantly, we observed an interaction between these factors, $F(1,30)=7.5$, $p=.01$, $\eta^2_p=.112$. While the performance on the linguistic task differed only marginally depending on whether participants had to complete a concomitant primary task, $F(1,16)=3.9$, $p=.065$, $\eta^2_p=.197$, performance on the nonlinguistic secondary task was markedly improved when the task was presented in isolation, $F(1,16)=36.0$, $p<.0001$, $\eta^2_p=.720$. In other words, not only did the nonlinguistic secondary task interfere more with scene understanding than the linguistic secondary task, but scene understanding also interfered more with the nonlinguistic secondary task than with the linguistic secondary task. Remarkably, the performance on the linguistic secondary task was almost unaffected by concomitant scene understanding, while the comparison of Experiments 1 and 2 reveals that scene understanding was unaffected by the presence of a linguistic secondary task. Hence, linguistic stimuli seem to be processed by mechanisms that are separate from those involved in visual scene understanding, even if both the scenes and the linguistic stimuli are presented visually.

## General discussion

In the experiments presented here, we probe the relation between language and conceptual processing at the earliest processing stages when stimuli are presented for durations of a single fixation. Previous research using similar presentation rates has revealed that observers extract and retain abstract conceptual information on top of visual information (Potter et al., 2004). Using this assay, we show that participants' grasp of the conceptual meaning of scenes is almost unaffected by a linguistic secondary task and vice versa, while scene understanding and a nonlinguistic secondary task mutually interfere. These results are not simply due to the nonlinguistic secondary tasks using more visual processing and memory resources, for three reasons. First, stimuli for the linguistic secondary task occluded at least as much surface area in the scenes as those for the nonlinguistic secondary task, and both needed to be processed visually. Second, the nonlinguistic secondary task in Experiment 3 did not require any visual memory at all, as participants had to react to a stimulus change immediately. Third, the processing advantage for scene recognition with a linguistic secondary task was maintained even when participants were tested on scene labels, which (presumably) rely more on conceptual information than on visual information. Taken together, our results thus suggest that the nonlinguistic secondary task interferes with processes that are crucial to scene understanding, while the linguistic secondary task appears to be essentially irrelevant to scene understanding.

Further, previous results suggest that linguistic and nonlinguistic processes can remain independent not only initially, but even in complex behaviors such as communication. For example, in languages such as English, the canonical word order is subject-verb-object (e.g., Mary sees John), while languages such as Turkish have the word order subject-object-verb (e.g., Mary John sees). However, when people have to gesture events (rather than to encode them verbally), they use the subject-object-verb order --- irrespective of the word order of their native language (Goldin-Meadow, So, Ozyürek, & Mylander, 2008; Langus & Nespor, 2010)---suggesting that the linguistic use of concepts and roles such as agents and patients does not affect how other processes use the same concepts and roles.

Despite the intimate link between language and conceptual structure, initial linguistic and nonlinguistic processes that derive meaning from sensory data thus appear to operate in remarkably independent channels. Interactions between linguistic and nonlinguistic conceptual processes might reflect top-down effects, occurring only if one set of processes establishes a prior context that is relevant to the other set of processes. For example, when listening to verbs describing upward motion, observers might be impaired in detecting actual downward motion (e.g., Meteyard, Bahrami, & Vigliocco, 2007) because listening to upward motion verbs might activate conceptual representations that exert top-down influences on motion perception, even though the processes used to understand verbs and to perceive motion are distinct and independent from one another. In the absence of such top-down effects, linguistic stimuli appear to be analyzed by dedicated linguistic processors at the earliest processing stages, providing further evidence for the remarkable modularity of processes that analyze different aspects of our environment.

## References

Baayen, R., Davidson, D., & Bates, D. (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). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577--660.

Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. Journal of Memory and Language, 49(3), 317--334.

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. Visual Cognition, 12(6), 893--924.

Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. Proceedings of the National Academy of Sciences of the United States of America, 105(27), 9163--8.

Huettig, F., & Altmann, G. T. M. (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.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. Journal of Experimental Psychology. Human Perception and Performance, 7(3), 604--610.

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. Cognitive Psychology, 60(4), 291--318.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. Proceedings of the National Academy of Sciences of the United States of America, 99(14), 9596--9601.

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

O'Connor, D. H.; Fukui, M. M.; Pinsk, M. A. & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. Nature Neuroscience, 5(11), 1203--9.

Pinheiro, J., & Bates, D. (2000). Mixed-effects models in S and S-PLUS. Berlin, Germany: Springer.

Potter, M. C. (1976). Short-term conceptual memory for pictures. Journal of Experimental Psychology: Human Learning and Memory, 2(5), 509--22.

Potter, M. C., Kroll, J., & Harris, C. (1980). Comprehension and memory in rapid sequential reading. In R. Nickerson (Ed.), Attention and performance VIII (pp. 395--418). Hillsdale, NJ: Erlbaum.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. Journal of Memory and Language, 29(6), 633--654.

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. Journal of Experimental Psychology. Human Perception and Performance, 30(3), 478--89.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: the time course of rapid forgetting. Journal of Experimental Psychology. Human Perception and Performance, 28(5), 1163--75.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. Behavioral and Brain Sciences, 22(3), 341--365.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. Journal of Verbal Learning and Verbal Behavior, 22(3), 358--374.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. Cognitive Psychology, 22(1), 36--71.

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

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--1634.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. Journal of Memory and Language, 33(3), 285--318.

Yi, D.-J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load. Nature Neuroscience, 7(9), 992--996.

**Author Notes**

This research was supported by NIH grant MH47432. We thank Á. Kovács for comments and Emily McCourt for assistance.

# Footnotes

---

[1] The effects of test position and modality were more pronounced in some experiments than others (see Table 3), but these differences are not relevant to the main questions addressed here.

[2] Compared to Experiment 2, the effect of test modality was somewhat more pronounced in Experiment 4, and that of test position somewhat less pronounced. While multiple factors might have contributed to these differences (e.g., the difficulty of the secondary tasks, the use of everyday meanings in Experiment 4 and so on), we note that Experiment 2 and 4 both replicate our crucial result that scene recognition is better with a linguistic secondary task compared to a non-linguistic secondary task.

[3] In Experiment 5a we used the nonsense sentences from Experiment 2 rather than the more sensible sentences from Experiment 4 because this allowed us to equate the task difficulty between the linguistic and the non-linguistic secondary task. Further, we increased the presentation duration to 250ms to vary the stimulus parameters.

Table 1

Results of repeated-measures ANOVAs on scene recognition performance with the within-subject predictors relative test position and test modality (scene vs. scene label). Data were analyzed in terms of five relative test positions of the five old pictures, matched with the five distractor pictures (rather than the 10 absolute test positions). In Experiments 5a and 5b, participants were tested on scenes but not on scene labels.

| | | Experiment[a] | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5a | 5b |
| **Test position** | $F(4,60)$ | 11.1 | 15.7 | 6.4 | 5.2 | 22.1 | 4.2 |
| | $p$ | <.0001 | <.0001 | < .0002 | .001 | <.0001 | .005 |
| | $\eta^2_p$ | .426 | .511 | .299 | .257 | .595 | .219 |
| **Test modality** | $F(1,15)$ | 19.3 | 14.8 | 12.5 | 20.7 | NA | NA |
| | $p$ | .0005 | .002 | .003 | .0004 | | |
| | $\eta^2_p$ | .563 | .496 | .454 | .58 | | |
| **Interaction** | $F(4,60)$ | *ns* | *ns* | 2.55 | 2.6 | NA | NA |
| | $p$ | | | .048 | .045 | | |
| | $\eta^2_p$ | | | .145 | .148 | | |

[a] In Experiments 2 to 5, a total of eight additional participants were excluded from the analyses of the primary task because their performance on the secondary task did not differ from chance by a one-tailed binomial test; as a result, the remaining participants were guaranteed to have paid attention to the secondary task. (The pattern of results is qualitatively unchanged if these participants are included.)

Table 2

Percentage of correct responses in the secondary tasks and associated t-tests against the chance level of 50% in the different experiments. In Experiments 5a and 5b, the secondary tasks where performed once as secondary task, and once as sole task without any additional task.

| | Experiment | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | As secondary task | | As sole task | |
| | 2 | 3 | 4 | 5a | 5b | 5a | 5b |
| ***M*** | 72.0% | 78.2% | 85.3% | 83.3% | 86.4% | 86.8% | 98.5% |
| **SD** | 14.7% | 8.6% | 11.5% | 13.6% | 8.8% | 14.2% | 1.8% |
| **t** | 6.7 | 13.5 | 12.6 | 10.4 | 16.5 | 10.7 | 102.0 |
| **df** | 19 | 16 | 16 | 17 | 15 | 16 | 14 |
| ***p*** | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 |
| **Cohen's *d*** | 1.5 | 3.3 | 3.1 | 2.4 | 4.1 | 2.6 | 26.0 |

Table 3

Scene recognition performance in the different experiments was compared using a logistic mixed-effects model (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000). We started with the following fixed effect predictors: secondary task (i.e., experiment), test modality (scene vs. scene label), absolute test position, all interactions between these predictors. The initial model included the following random effect predictors: intercept adjustments for participants, trial number, test item; slope adjustment for test items relative to the slope of the test modality predictor. The final model included only those (fixed and random effect) predictors that contributed significantly to the likelihood of the model.

| | | **Comparison between experiments**[a] | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 vs. 2 | 1 vs. 3 | 2 vs. 3 | 2 vs. 4 | 3 vs. 4 | 5a vs. 5b |
| **Experiment** | *Z* | *ns* | 2.0 | 2.8 | *ns* | 2.4 | 2.9 |
| | *p* | | .044 | .005 | | .016 | .003 |
| **Test position** | *Z* | -9.2 | -9.6 | -10.8 | -13.1 | -12.2 | -4.0 |
| | *P* | .0001 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 |
| **Test modality** | *Z* | 6.4 | 7.7 | 6.7 | 6.0 | 7.1 | NA |
| | *p* | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 | NA |
| **Experiment ×** | *Z* | -2.7 | *ns* | -2.6 | 2.0 | *ns* | *ns* |
| **test position** | *p* | .007 | | .008 | .048 | | |
| **Experiment ×** | *Z* | *ns* | *ns* | *ns* | 2.4 | 2.5 | NA |
| **test modality** | *p* | | | | .014 | .011 | NA |
| **Test position ×** | *Z* | 2.2 | 3.4 | 3.7 | 3.8 | 4.7 | NA |
| **test modality** | *p* | .026 | .0006 | .0003 | .0001 | .0001 | NA |

[a] The exclusion criteria in Experiments 2 to 5 are given in Table 1.

## Figure Captions

Figure 1: Paradigm used in all experiments. (left) In each trial, participants saw a sequence of scenes at a rate of 200ms per scene, following a fixation cross. In Experiments 1 and 3, the center of each scene presented a small box with gridlines. In Experiment 1, participants were instructed to look at the box; in Experiment 3, they had to detect changes in the gridline density. (right) Following the rapidly presented scenes, participants completed a recognition test with 10 items (5 new, 5 old). In half of the scene recognition trials, they were tested on the scenes; in the other scene recognition trials, they were tested on labels corresponding to the scenes. In Experiments 2 and 4, the boxes were replaced by 6 words forming a sentence, presented word by word, synchronized with the onset of the scenes; participants were instructed to remember the sentence. On half the trials they were tested on the pictures, and on the other half on the sentences. Experiments 5a and 5b were similar to Experiments 2 and 3, except that participants were never tested on scene labels.

Figure 2: Percentage of correct responses when participants were tested on scenes (solid lines) or verbal scene labels (dashed lines) as a function of the relative test position. Error bars represent SEM.

Figure 3: Performance in the secondary tasks in Experiment 2 to 4. Dots represent participants, the diamonds the sample averages, the dotted line the sample chance level of 50%, and the dashed line the

chance level of individual participants of 65% (as determined by a one-tailed binomial test).

Figure 4: Results of Experiments 5a and 5b. (a) Percentage of correct responses as a function of the relative test position when the secondary task was to read sentences (dotted line) or to detect a density change (dot-dash line). In contrast to Experiments 1 to 4, participants were tested only on scenes, and not on scene labels. Error bars represent SEM. (b) Performance in the secondary tasks, when tested with a primary task and in the absence of a primary task, respectively. Dots represent participants, the diamonds the sample averages, the dotted line the sample chance level of 50%, and the dashed line the chance level of individual participants of 65% (as determined by a one-tailed binomial test).

Figure 1

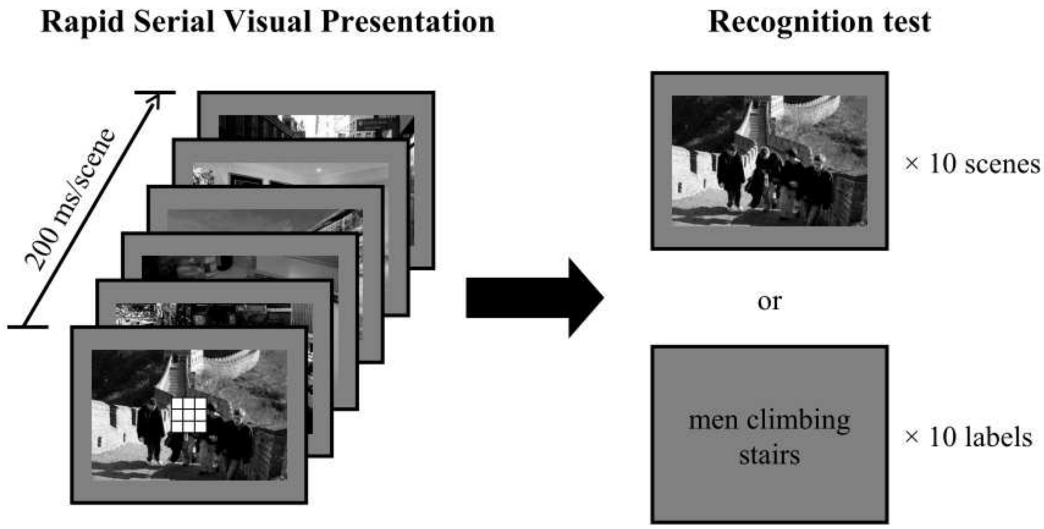

Rapid Serial Visual Presentation
200 ms/scene
Recognition test
× 10 scenes
or
men climbing stairs
× 10 labels

Figure 2

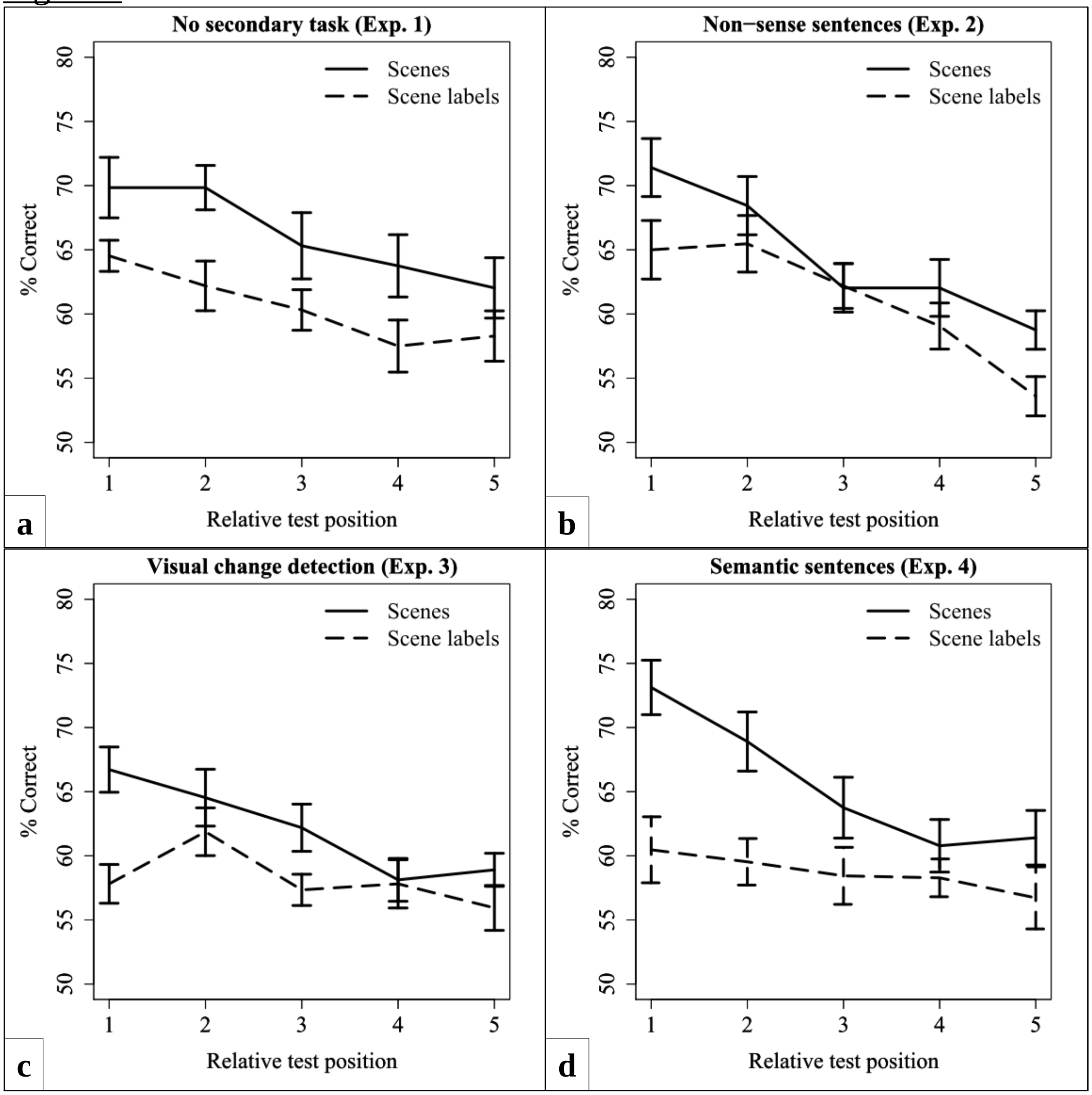

No secondary task (Exp. 1)
Non−sense sentences (Exp. 2)
Visual change detection (Exp. 3)
Semantic sentences (Exp. 4)
Scenes
Scene labels
% Correct
Relative test position
a
b
c
d

Figure 3

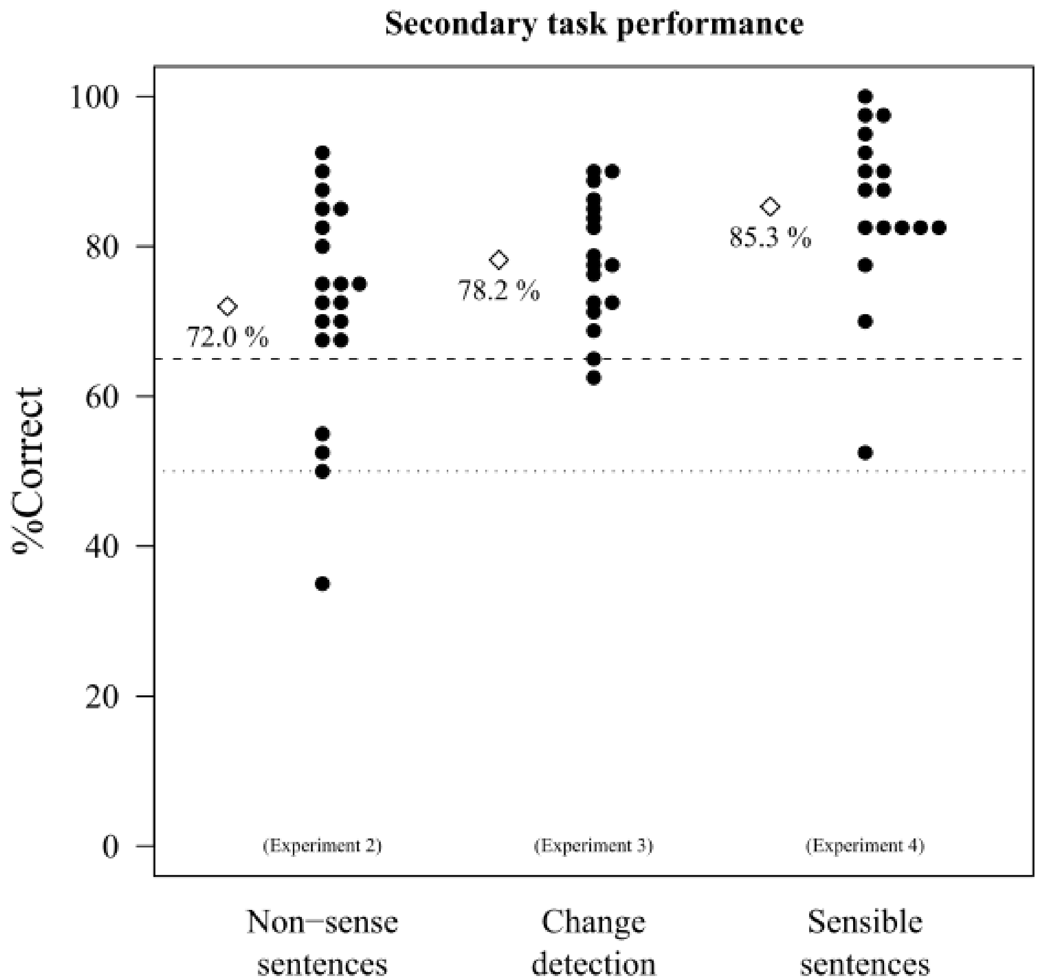
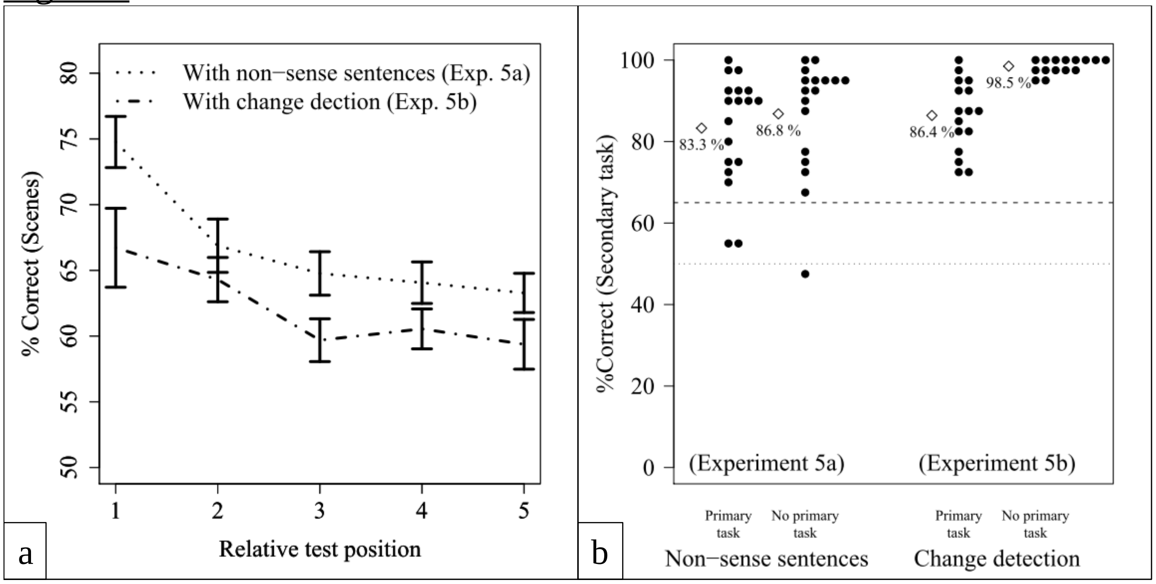

Secondary task performance
%Correct
100
80
60
40
20
0
72.0 %
78.2 %
85.3 %
(Experiment 2)
(Experiment 3)
(Experiment 4)
Non−sense sentences
Change detection
Sensible sentences

Figure 4


a
% Correct (Scenes)
50
55
60
65
70
75
80
With non–sense sentences (Exp. 5a)
With change dection (Exp. 5b)
1
2
3
4
5
Relative test position
b
%Correct (Secondary task)
0
20
40
60
80
100
83.3 %
86.8 %
86.4 %
98.5 %
(Experiment 5a)
(Experiment 5b)
Primary task
No primary task
Primary task
No primary task
Non–sense sentences
Change detection

## Supplementary materials

Table S1

The "hit rate" for scene recognition performance in the different experiments was compared using a logistic mixed-effects model (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000). We started with the following fixed effect predictors: secondary task (i.e., experiment), test modality (scene vs. scene label), absolute test position, all interactions between these predictors. The initial model included the following random effect predictors: intercept adjustments for participants, trial number, test item; slope adjustment for test items relative to the slope of the test modality predictor. The final model included only those (fixed and random effect) predictors that contributed significantly to the likelihood of the model.

| | | **Comparison between experiments (hit rate)**[a] | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 vs. 2 | 1 vs. 3 | 2 vs. 3 | 2 vs. 4 | 3 vs. 4 | 5a vs. 5b |
| **Experiment** | *Z* | 3.3 | *ns* | 4.0 | *ns* | 3.6 | *ns* |
| | *p* | .001 | | <.0001 | | .0004 | |
| **Test position** | *Z* | -8.2 | -8.3 | -10.9 | -15.2 | -14.5 | -12.5 |
| | *P* | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 |
| **Test modality** | *Z* | 4.0 | *ns* | 4.4 | 5.1 | 4.3 | NA |
| | *p* | <.0001 | | <.0001 | <.0001 | <.0001 | |
| **Experiment ×** | *Z* | 6.1 | *ns* | 4.1 | 3.9 | *ns* | *ns* |
| **test position** | *p* | <.0001 | | <.0001 | <.0001 | | |
| **Experiment ×** | *Z* | *ns* | *ns* | *ns* | *ns* | 2.2 | NA |
| **test modality** | *p* | | | | | .025 | |
| **Test position ×** | *Z* | 3.1 | *ns* | 5.1 | 4.8 | 6.2 | NA |
| **test modality** | *p* | .002 | | <.0001 | <.0001 | <.0001 | |

[a] The exclusion criteria in Experiments 2 to 5 are given in Table 1.

Table S2

The “correct rejection” rate for scene recognition performance in the different experiments was compared using a logistic mixed-effects model (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000). We started with the following fixed effect predictors: secondary task (i.e., experiment), test modality (scene vs. scene label), absolute test position, all interactions between these predictors. The initial model included the following random effect predictors: intercept adjustments for participants, trial number, test item; slope adjustment for test items relative to the slope of the test modality predictor. The final model included only those (fixed and random effect) predictors that contributed significantly to the likelihood of the model.

| | | **Comparison between experiments (correct rejections)**[a] | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 vs. 2 | 1 vs. 3 | 2 vs. 3 | 2 vs. 4 | 3 vs. 4 | 5a vs. 5b |
| **Experiment** | *Z* | -3.4 | *ns* | -2.9 | ns | -3.1 | 2.7 |
| | *p* | .0006 | | .004 | | .002 | .007 |
| **Test position** | *Z* | 2.6 | 2.8 | 6.8 | 5.7 | 5.5 | 11.8 |
| | *p* | .008 | .005 | <.0001 | <.0001 | <.0001 | <.0001 |
| **Test modality** | *Z* | *ns* | 2.5 | 2.3 | ns | 2.5 | NA |
| | *p* | | .011 | .022 | | .013 | |
| **Experiment ×** | *Z* | 3.0 | 2.3 | *ns* | ns | *ns* | *ns* |
| **test position** | *p* | .003 | .021 | | | | |
| **Experiment ×** | *Z* | *ns* | *ns* | 3.5 | 2.1 | *ns* | NA |
| **test modality** | *p* | | | .0004 | .032 | | |
| **Test position ×** | *Z* | 2.9 | 2.9 | 4.5 | 3.0 | 3.1 | NA |
| **test modality** | *p* | .004 | .004 | <.0001 | .003 | .002 | |

[a] The exclusion criteria in Experiments 2 to 5 are given in Table 1.

Figure S3

Results of the primary tasks of Experiments 1 to 4 in terms of hit rats and false alarm rates. Error bars represent SEM.

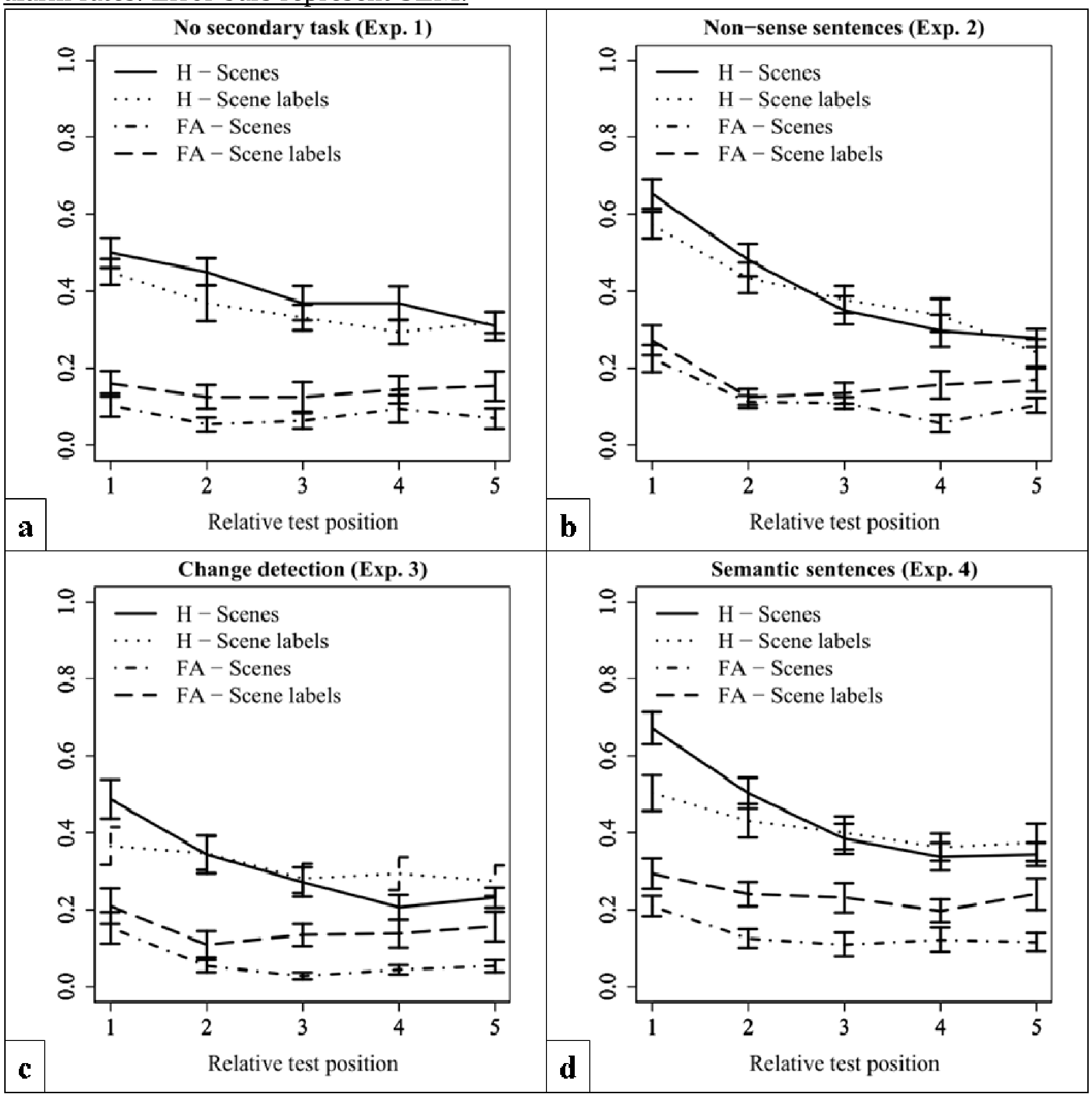


Figure S4

Hit rates and false alarm rates in the secondary tasks in Experiment 2 to 4. Bars represent the sample averages; error bars represent SEM.

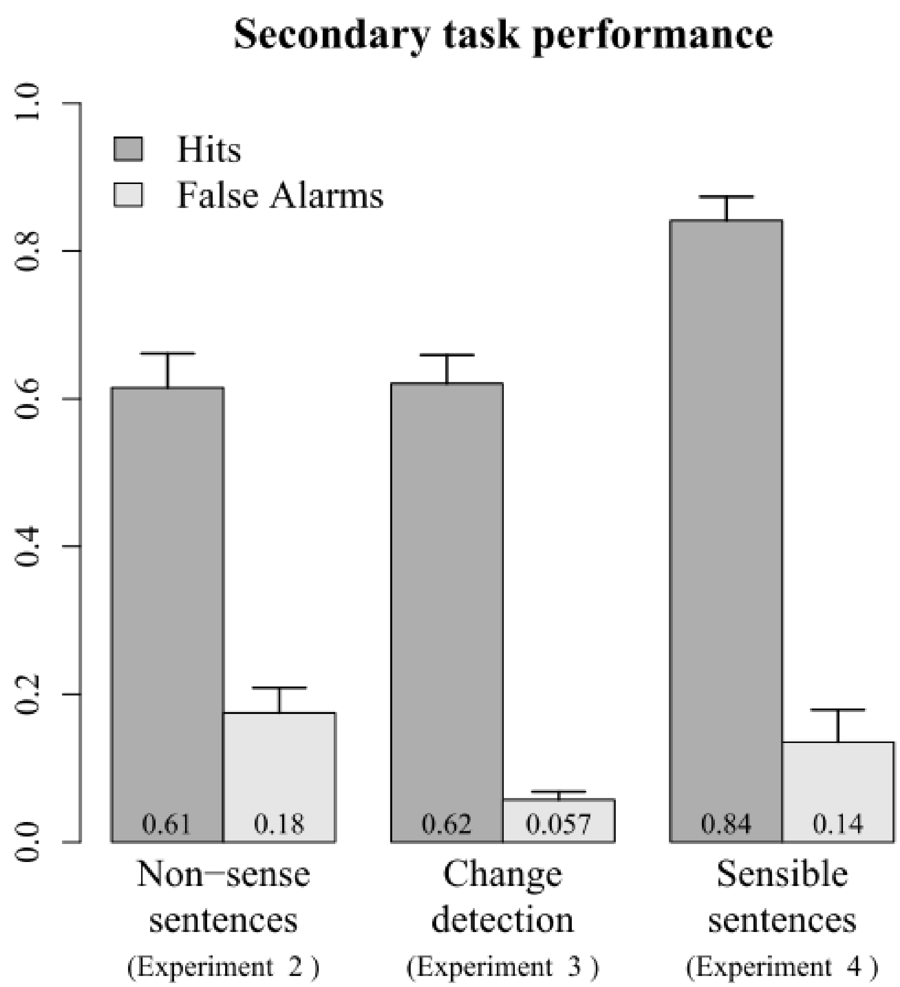


Figure S5

Results of Experiments 5a and 5b. (a) Hit rates and false alarm rates in the primary, scene recognition task. Error bars represent SEM. (b) Hit rates and false alarm rates in the secondary tasks, when tested with a primary task and in the absence of a primary task, respectively. Bars represent the sample averages; error bars represent SEM.