Early Conceptual and Linguistic Processes Operate in Independent Channels

Ansgar D. Endress and Mary C. Potter
Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology

Abstract
Language and concepts are intimately linked, but how do they interact? In the study reported here, we probed the relation between conceptual and linguistic processing at the earliest processing stages. We presented observers with sequences of visual scenes lasting 200 or 250 ms per picture. Results showed that observers understood and remembered the scenes' abstract gist and, therefore, their conceptual meaning. However, observers remembered the scenes at least as well when they simultaneously performed a linguistic secondary task (i.e., reading and retaining sentences); in contrast, a nonlinguistic secondary task (equated for difficulty with the linguistic task) impaired scene recognition. Further, encoding scenes interfered with performance on the nonlinguistic task and vice versa, but scene processing and performing the linguistic task did not affect each other. At the earliest stages of conceptual processing, the extraction of meaning from visually presented linguistic stimuli and the extraction of conceptual information from the world take place in remarkably independent channels.

Language and concepts are intimately linked. For example, conceptual real-world knowledge or even just seeing visual arrays of objects can affect how people 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 Clifton et al., 2003; Rayner, Carlson, & Frazier, 1983). Further, the semantic or conceptual meaning of words can affect even 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 Huettig & Altmann, 2007, and Spivey, Tyler, Eberhard, & Tanenhaus, 2001, among many others).

Such results raise the question of whether processes that derive meaning from sensory data, be these data 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 that people can entertain (Whorf, 1956) and the embodied, simulationist view that understanding any concept involves mentally simulating its referent (e.g.,

to understand the meaning of "upward," people mentally simulate what upward motion looks like; e.g., see Barsalou, 1999). Other researchers hold that conceptual information and linguistic information are processed by completely modular and encapsulated processors (e.g., Fodor, 1983; Pylyshyn, 1999). In between these two views, still other accounts suggest that linguistic stimuli are analyzed by dedicated processors, but that these processors can also incorporate nonlinguistic information when it is available (Tanenhaus et al., 1995; Trueswell et al., 1994).

Although 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

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

level of the thalamus (O'Connor, Fukui, Pinsk, & Kastner, 2002), such effects could occur at the earliest processing stages. For example, in the studies conducted by Meteyard et al. (2007), participants continuously listened to verbs representing a direction of motion at a rate of one verb per second; visual stimuli moved only during randomly spaced periods of 150 ms. Hence, listening to upward or downward verbs might have placed participants in upward or downward mind-sets, in which thinking about upward and downward motion might have influenced their motion perception. To test the interdependence of linguistic and nonlinguistic processes, researchers therefore need to create a situation in which top-down effects are precluded and in which one kind of process cannot establish a context for the subsequent one.

In the experiments reported here, we probed the relation between language and nonlinguistic concepts at the earliest processing stages. We precluded top-down effects by having both kinds of information processed simultaneously and under time pressure, and by feeding the two kinds of processes information that was largely unrelated. In each trial, participants viewed a sequence of six unrelated scenes presented at a rate of 200 ms or 250 ms per picture (see Fig. 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 such scenes (Potter, Staub, Rado, & O'Connor, 2002), but they also succeed even when tested on descriptions of the scenes; for example, they can decide whether they viewed a scene corresponding with the description "people in street" (Potter, Staub, & O'Connor, 2004). Such findings indicate that people extract not only low-level visual information, as in traditional studies of visual short-term memory, but also the conceptual gist of the scenes (e.g., Intraub, 1981; Potter, 1976).

After viewing the six rapidly presented scenes, participants completed a yes/no recognition test of scene memory consisting either of 10 scenes (all experiments) or 10 descriptions of scenes (first four experiments only) presented one at a time (Fig. 1). In the latter trials, participants had to decide whether they had viewed scenes corresponding with the descriptions. In both types of trial, half of the test items were old, and half were new.

In Experiment 1, we established that observers can extract the gist of scenes presented at a rate of 200 ms per picture when no secondary task is involved; this finding replicates the results of previous experiments investigating conceptual short-term memory (Potter et al., 2002; Potter et al., 2004). In Experiment 2, we tested whether a linguistic secondary task interferes with scene memory. Specifically, a written word was presented in the center of each scene; the resulting sequence of six words formed 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 rapid serial visual presentation (RSVP), in which words were presented one by one at rates of up to 12 words per second (Potter, Kroll, & Harris, 1980; Potter & Lombardi, 1990). Following each sequence, participants were tested on their memory either for the scenes or for the sentence.

In Experiment 3, we asked whether a nonlinguistic secondary task interferes with memory for scenes. The center of each scene contained a small box with grid lines. Participants were
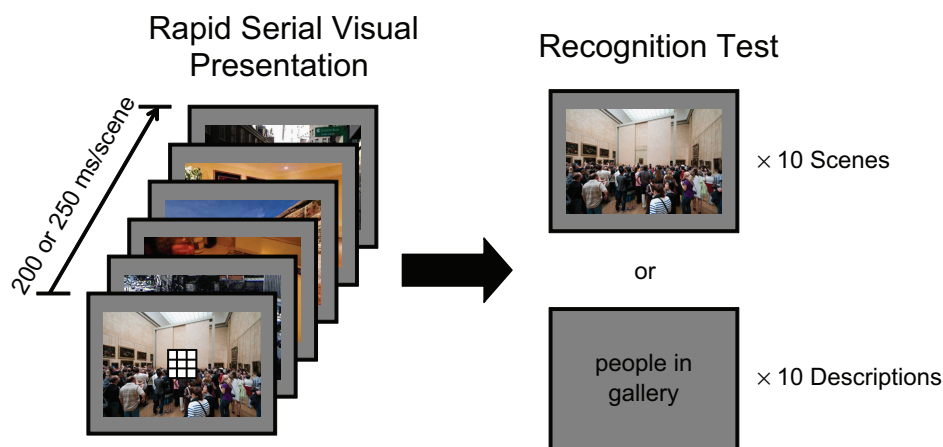

Fig. 1. Illustration of the display sequence and the scene-recognition test used in all experiments. In each trial, participants saw a rapidly presented sequence of six scenes (left) at a rate of either 200 ms per picture (Experiments 1–4) or 250 ms per picture (Experiments 5a and 5b); a small white box appeared in the center of each scene. In Experiments 1 and 3, the box contained grid lines (shown here). In Experiment 3, participants had to detect changes in the density of the grid lines. In Experiments 2 and 4, each of the six boxes contained a word, and the six words formed a sentence; participants were instructed to remember the sentence. The scene-recognition test consisted of 10 items (5 new, 5 old). Half of the test items used pictures, and half used descriptions of the scenes. Experiments 5a and 5b were similar to Experiments 2 and 3, respectively, except that memory was tested only on scenes and not on scene descriptions.

instructed to press a key when they detected a change in the density of the grid lines. Experiment 4 replicated Experiment 2, but using sentences that were semantically more sensible. Experiments 5a and 5b provided additional controls.

General Method
Participants

Ninety-six native speakers of English (55 women, 41 men; mean age = 23.3 years) from the Massachusetts Institute of Technology community participated in the study. Participants were distributed evenly among the six experiments (i.e., 16 participants in each experiment). No participant took part in more than one experiment.

Stimuli

Scenes used in the primary task of each experiment were created following the methodology of Potter et al. (2004). These scenes consisted of color photographs collected from the World Wide Web and commercial sources. Descriptions corresponding with these pictures were generated by two research assistants. Scenes (and the corresponding descriptions) were randomly organized into sets of 11 pictures (6 items for the study phase and recognition task of each trial, and 5 new items for the recognition task only), with the constraint that the items in a set had no obvious relation with each other. The center of each scene contained a white box with stimuli that varied across the six experiments according to the secondary task.

Procedure

Each experiment comprised 80 trials. Each trial began with an RSVP phase: Following a central fixation cross, a sequence of six scenes appeared on a computer screen (Fig. 1). Participants then completed a scene-recognition test by pressing premarked "Yes" and "No" keys on a keyboard. In all but the first experiment, participants also completed a secondary task, the nature of which varied according to the purpose of the experiment. Participants received four practice trials before starting the experiment.

Data analysis

Our primary dependent measure was the percentage of correct responses. The percentage of correct responses in the scene-recognition task was analyzed in a repeated measures analysis of variance (ANOVA) with relative test position (i.e., old test picture's position among the five old test pictures or new test picture's position among the five new test pictures) and test modality (scene vs. description) as within-subjects factors. The percentage of correct responses in the secondary task was compared across experiments using ANOVAs. Further, we ran *t* tests to compare performance in both the primary and the secondary tasks against the chance level of 50%.

Scene-recognition performance was compared across experiments using a logistic mixed-effects model (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000) with secondary task (i.e., experiment), test modality (scene vs. description), absolute test position (1–10), and all interactions between these factors entered into the model as predictors. The initial model included intercept adjustments for participants, trial number, and test item as random-effects predictors; slope adjustment for test items relative to the slope of the test-modality predictor was also included as a random-effects predictor. The final model included only those (fixed- and random-effects) predictors that contributed significantly to the likelihood of the model.

All participants were included in the analyses of the primary task of Experiment 1. Across Experiments 2 through 5, a total of 8 participants were excluded from the analyses of the primary task because their performance on the secondary task did not differ from chance according to a one-tailed binomial test; using this criterion guaranteed that the remaining participants paid attention to the secondary task. (The pattern of results was qualitatively unchanged when these participants were included.)

Experiment 1
Method

In Experiment 1, scenes were presented at a rate of 200 ms each. The central box (35 × 35 pixels) contained a regular grid pattern of 2 × 2, 3 × 3, 5 × 5, 11 × 11, or 17 × 17 equally spaced horizontal and vertical lines. The box appeared in synchrony with the scenes. On half of the trials, the density of the grid lines in one box changed relative to the previous box. This change occurred equally often on the second, third, fourth, and fifth scene. After a change, the box changed back to its original grid line density when the next scene appeared, and no further density changes occurred in that trial. Density changes always crossed two density steps (e.g., from 2 to 5 lines, from 3 to 11 lines).

Participants were instructed to look at the center box and also remember the scenes. Following the presentation of the six scenes, participants completed a recognition task consisting of 10 test items. Half of these items had appeared in the RSVP sequence, and half were new. In a random half of the trials, participants were tested on scenes; in the remaining trials, they were tested on verbal descriptions of the scenes. The sixth scene never appeared in the test phase of any trial because it was not masked by a subsequent scene and was therefore easily remembered (Potter et al., 2002). No picture appeared in more than one trial.

Results and discussion

To analyze an equal number of scene test trials in Experiments 1 through 4, we considered only those trials in Experiment 1 in which the density of the grid lines in the center of the scenes

did not change. Our analysis showed that participants successfully remembered scenes (Fig. 2 and Table 1; see Fig. S1 in the Supplemental Material available online for proportions of hits and false alarms). However, they performed better when tested on scenes than when tested on descriptions, presumably because pictures provide participants with visual and conceptual information in addition to the gist of the scenes (the only information carried by the descriptions). Replicating the findings of earlier work (Potter et al., 2002; Potter et al., 2004), our results showed that participants performed worse on items in later test positions than in earlier test positions, probably because of decay or interference. However, both when tested on scenes and tested on descriptions, participants performed significantly above chance on items in all test positions.[1]

Experiment 2
Method

Experiment 2 addressed the question of how linguistic and conceptual mechanisms interact at early processing stages. The procedure was the same as in Experiment 1, with two exceptions. First, the central box in the RSVP sequence (3.5 × 0.8 degrees of visual angle) showed a word rather than grid lines. One word was presented with each scene, and the six words formed a sentence that was syntactically acceptable but made little sense, such as "miners duly locate truly tired ladies."

Second, in half of the trials, participants were tested on their memory for the sentences; immediately after the RSVP sequence, they saw an entire sentence on the screen and had to
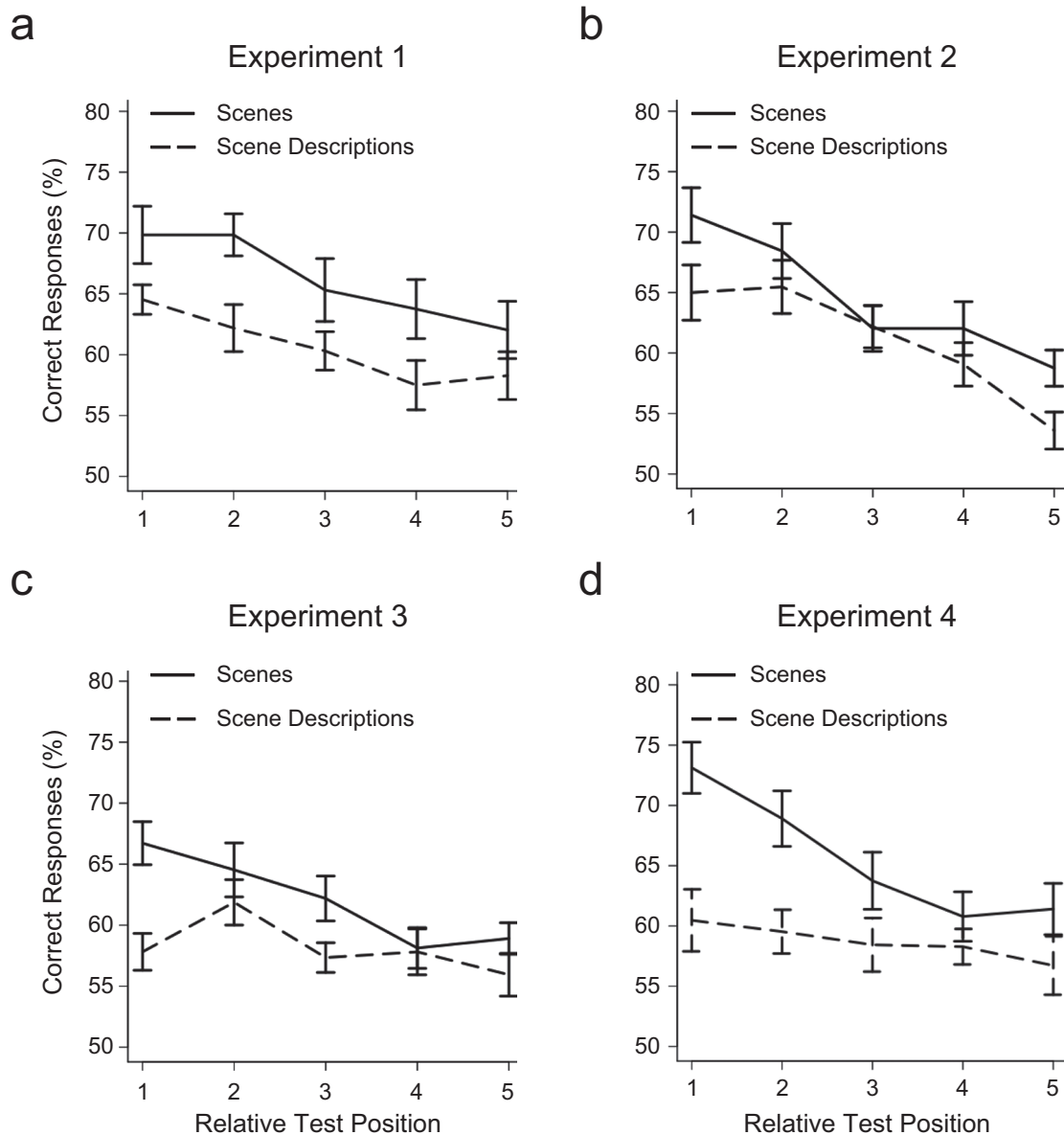

Fig. 2. Results from Experiments 1 through 4: mean percentage of correct responses on the recognition test as a function of the type of test item (scene or description) and the item's relative test position. In Experiment 1 (a), participants did not perform a secondary task. In the other experiments, the secondary task consisted of recalling nonsense sentences (Experiment 2; b), detecting changes in the density of grid lines (Experiment 3; c), and recalling sentences that made semantic sense (Experiment 4; d). Error bars represent standard errors of the mean.

Table 1. Results of Analyses of Scene-Recognition Performance in All Experiments

	Effect		
Experiment	Relative test position	Test modality	Relative Test Position × Test Modality
Experiment 1	$F(4, 60) = 11.1, p < .0001,$ $\eta_p^2 = .426$	$F(1, 15) = 19.3, p < .0005,$ $\eta_p^2 = .563$	n.s.
Experiment 2	$F(4, 60) = 15.7, p < .0001,$ $\eta_p^2 = .511$	$F(1, 15) = 14.8, p < .002,$ $\eta_p^2 = .496$	n.s.
Experiment 3	$F(4, 60) = 6.4, p < .0002,$ $\eta_p^2 = .299$	$F(1, 15) = 12.5, p < .003,$ $\eta_p^2 = .454$	$F(4, 60) = 2.55, p < .048,$ $\eta_p^2 = .145$
Experiment 4	$F(4, 60) = 5.2, p < .001,$ $\eta_p^2 = .257$	$F(1, 15) = 20.7, p < .0004,$ $\eta_p^2 = .58$	$F(4, 60) = 2.6, p < .045,$ $\eta_p^2 = .148$
Experiment 5a	$F(4, 60) = 22.1, p < .0001,$ $\eta_p^2 = .595$	—	—
Experiment 5b	$F(4, 60) = 4.2, p < .005,$ $\eta_p^2 = .219$	—	—

Note: In Experiments 1 through 4, participants were tested on scenes and descriptions of scenes. In Experiments 5a and 5b, participants were tested on scenes only.

decide whether or not a word had been changed. In half of these trials, one word had been replaced by a new word that preserved the same grammatical structure as in the original sentence. In the remaining trials, participants were tested on their recognition of scenes and scene descriptions as in Experiment 1.

The sentences were composed according to 10 different grammatical templates by drawing quasirandomly from lists of words in the relevant form classes (nouns, verbs, adverbs, adjectives, and prepositions). Words selected were reasonably frequent (COBUILD frequencies between 100 and 10,000 according to the CELEX corpus; Baayen, Piepenbrock, & Gulikers, 1995) and had two syllables and four to six letters.

If understanding images involves linguistic resources, we would expect a large decrement in scene-memory performance between Experiments 1 and 2. In contrast, if people grasp the conceptual meaning of scenes by virtue of nonlinguistic mechanisms, we would expect either no decrease in scene-memory performance or a limited decrease because of the attentional demands of performing two tasks simultaneously.

Results and discussion

In the primary (scene-memory) task in Experiment 2, participants performed better when tested on scenes than when tested on descriptions (as in Experiment 1). Further, they performed worse on items in later test positions than in earlier test positions (Fig. 2 and Table 1; see Fig. S1 for proportions of hits and false alarms). Comparing the results of the primary tasks of Experiments 1 and 2 using a mixed-effects model (Table 2)

Table 2. Results of the Logistic Mixed-Effects Model Comparing Scene-Recognition Performance Between Experiments

|	Predictor					
Experiments compared	Experiment	Absolute test position	Test modality	Experiment × Absolute Test Position	Experiment × Test Modality	Absolute Test Position × Test Modality
1 vs. 2	n.s.	$Z = -9.2,$ $p = .0001$	$Z = 6.4,$ $p < .0001$	$Z = -2.7,$ $p = .007$	n.s.	$Z = 2.2,$ $p = .026$
1 vs. 3	$Z = 2.0,$ $p = .044$	$Z = -9.6,$ $p < .0001$	$Z = 7.7,$ $p < .0001$	n.s.	n.s.	$Z = 3.4,$ $p = .0006$
2 vs. 3	$Z = 2.8,$ $p = .005$	$Z = -10.8,$ $p < .0001$	$Z = 6.7,$ $p < .0001$	$Z = -2.6,$ $p = .008$	n.s.	$Z = 3.7,$ $p = .0003$
2 vs. 4	n.s.	$Z = -13.1,$ $p < .0001$	$Z = 6.0,$ $p < .0001$	$Z = 2.0,$ $p = .048$	$Z = 2.4,$ $p = .014$	$Z = 3.8,$ $p = .0001$
3 vs. 4	$Z = 2.4,$ $p = .016$	$Z = -12.2,$ $p < .0001$	$Z = 7.1,$ $p < .0001$	n.s.	$Z = 2.5,$ $p = .011$	$Z = 4.7,$ $p = .0001$
5a vs. 5b	$Z = 2.9,$ $p = .003$	$Z = -4.0,$ $p < .0001$	—	n.s.	—	—
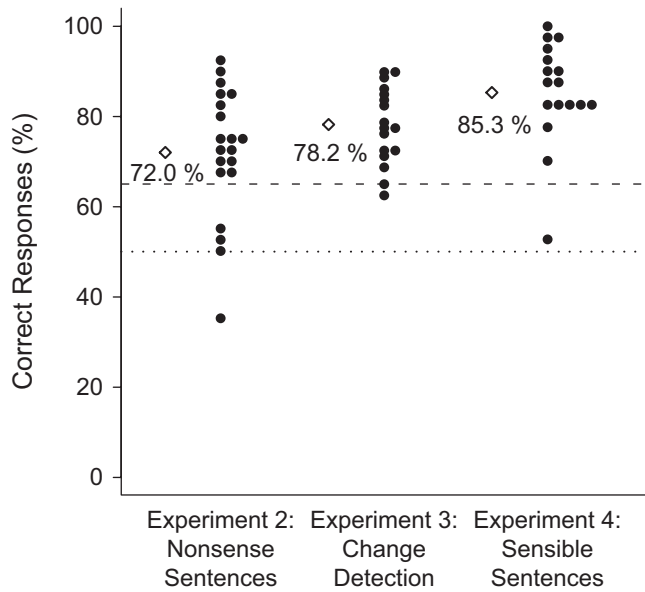
Fig. 3. Percentage of correct responses in the secondary tasks in Experiments 2 through 4. Dots indicate individual participants' means, and diamonds indicate averages across the sample. The dotted line denotes the chance level of performance across the sample (50%), and the dashed line shows the chance level of performance for individual participants (65%, as determined by a one-tailed binomial test).

showed no main effect of experiment, but participants performed numerically better in Experiment 2 despite having to attend to a secondary task (see Tables S1 and S2 in the Supplemental Material for similar comparisons using hits and correct rejections). Across both experiments, recognition was better for scenes than for descriptions and for earlier than for later test positions. In addition, the separation between performance for scenes and performance for descriptions diminished for later test positions.[2] In the secondary task in Experiment 2, participants successfully detected changed words (Fig. 3 and

Table 3; see Fig. S2 in the Supplemental Material for proportions of hits and false alarms).

Compared with Experiment 1, Experiment 2 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. Previous research has shown that, at least after massive training, some types of natural scene processing can occur with very limited attentional involvement (Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002; but see Yi, Woodman, Widders, Marois, & Chun, 2004); regardless, we 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 understanding of scenes and understanding of language rely on different 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 understanding of scenes was not impaired by the linguistic secondary task of Experiment 2. As in Experiment 2, participants completed a secondary task; in contrast with Experiment 2, however, this task was nonlinguistic. If understanding of scenes is simply unaffected by secondary tasks, we would expect to replicate the results of Experiment 2 and observe no impairment in scene recognition. In contrast, scene-recognition performance might be affected by nonlinguistic secondary tasks that tap into processes required for understanding scenes (e.g., visual processing), even if scene-recognition performance is not affected by linguistic secondary tasks.

Experiment 3
Method

The procedure for Experiment 3 was the same as for Experiment 1, except that participants were instructed to press a key

Table 3. Percentage of Correct Responses in the Secondary Tasks of All Experiments and Results of Analyses Comparing Performance Against Chance

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

Note: In Experiments 2 through 4, the secondary task was always performed as a secondary task. In Experiments 5a and 5b, the secondary tasks were performed both as a secondary task and as the sole task. The t tests compared the percentage of correct responses against the chance level of 50%.

when they detected a change in the density of the grid lines in the center box (the density was changed on half of the trials). Before starting the experiment, participants received four practice trials in which only the center box was presented, without any scenes. Scene-recognition was tested as in Experiment 1.

Results and discussion

To analyze an equal number of scene test trials in Experiments 1 through 4, we considered only those trials in Experiment 3 in which the density of the grid lines in the center of the scenes did not change. In the primary task of Experiment 3, participants performed better when tested on scenes than when tested on descriptions (Fig. 2 and Table 1; see Fig. S1 for proportions of hits and false alarms); further, they performed worse on items in later test positions than in earlier test positions, mirroring the results of Experiment 1.

Unsurprisingly, given that participants had to perform a secondary task in Experiment 3 but not in Experiment 1, results of the mixed-effects model comparing scene-recognition performance across these experiments showed that participants performed worse in Experiment 3 than in Experiment 1 (Table 2; see Tables S1 and S2 for similar comparisons using hits and correct rejections). These results contrast with the comparison of Experiments 1 and 2, in which 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 2).

Regarding the secondary task, our analysis showed that participants successfully detected density changes in the grid lines (Fig. 3 and Table 3; see Fig. S2 for proportions of hits and false alarms); performance did not differ from that on the secondary task in Experiment 2, $F(1, 35) = 2.3$, $p = .137$, $\eta^2 = .062$, although performance on the sentence task in Experiment 2 was numerically worse.

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 of remembering scenes), participants were significantly better at recognizing scenes in Experiment 2 than in Experiment 3. This finding suggests 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.

Method

In Experiment 4, we controlled for the possibility that participants did not fully process nonsense sentences by replicating Experiment 2 but using simple, semantically coherent six-word sentences (e.g., "Carol rants about the lousy food"). The only constraints imposed on the words were that they were reasonably frequent and had at most 8 letters. These more interpretable sentences would be more likely to trigger normal sentence processing than the less meaningful sentences in Experiment 2.

Results and discussion

In the primary task of Experiment 4, participants performed better when tested on scenes than when tested on descriptions; further, they performed worse on items in later test positions than in earlier test positions (Fig. 2 and Table 1; see Fig. S1 for proportions of hits and false alarms). In the secondary task of Experiment 4, participants successfully detected changed words (Fig. 3 and Table 3; see Fig. S2 for proportions of hits and false alarms). Secondary task performance was better than in Experiment 2, $F(1, 35) = 9.1$, $p = .005$, $\eta^2 = .207$, and in Experiment 3, $F(1, 32) = 4.2$, $p = .049$, $\eta^2 = .116$. Scene-recognition performance was significantly better than in Experiment 3, but not compared with performance in Experiment 2 (Table 2; see Tables S1 and S2 for similar comparisons using hits and correct rejections).[3] 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 use to remember scenes. Participants sometimes report trying to find descriptions for scenes, thereby occupying resources that would no longer be available to encode the scenes. A similar observation has been made in experiments in which 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 more impaired than in various control tasks that did not involve verbalization of the stimuli (Schooler & Engstler-Schooler, 1990). Similarly, a secondary language task of the sort used here 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 through 4 might have encouraged such strategies because participants were tested not only on actual scenes but also on descriptions of those scenes. It is possible that a linguistic secondary task might reveal interference with scene understanding if participants' performance on scenes

had been tested only with actual scenes rather than with descriptions. We tested this possibility in Experiments 5a and 5b.

Experiment 5
Method

Experiments 5a and 5b replicated Experiments 2 and 3, respectively, with three crucial changes. First and most important, participants were never tested on descriptions of scenes, but only on actual 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 participants in Experiment 2 did; these sentences were again presented word by word in the center of the scenes. 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 grid lines in a small square; however, rather than pressing a key as soon as they saw a density change, on half of 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 completed the experiments, participants were tested on the secondary task in isolation, with no primary task and no scenes shown. In addition, we increased the presentation duration to 250 ms per picture in Experiments 5a and 5b.[4]

Results and discussion

Scene-recognition performance in Experiments 5a and 5b is shown in Figure 4a (results of the analyses of scene-recognition performance are shown in Table 1, and proportions of hits and false alarms are shown in Fig. S3a). Replicating the results of Experiments 2 and 3, the data showed that participants performed better in Experiment 5a than in Experiment 5b (Table 2; see Tables S1 and S2 for similar comparisons using hits and false alarms), which suggests that the nonlinguistic secondary task of Experiment 5b interfered more with scene understanding than did the linguistic secondary task of Experiment 5a.

Performance on the secondary task was analyzed in two ways: first, when the secondary task was performed in conjunction with the primary task and, second, when the secondary task was presented alone. When the secondary task was performed
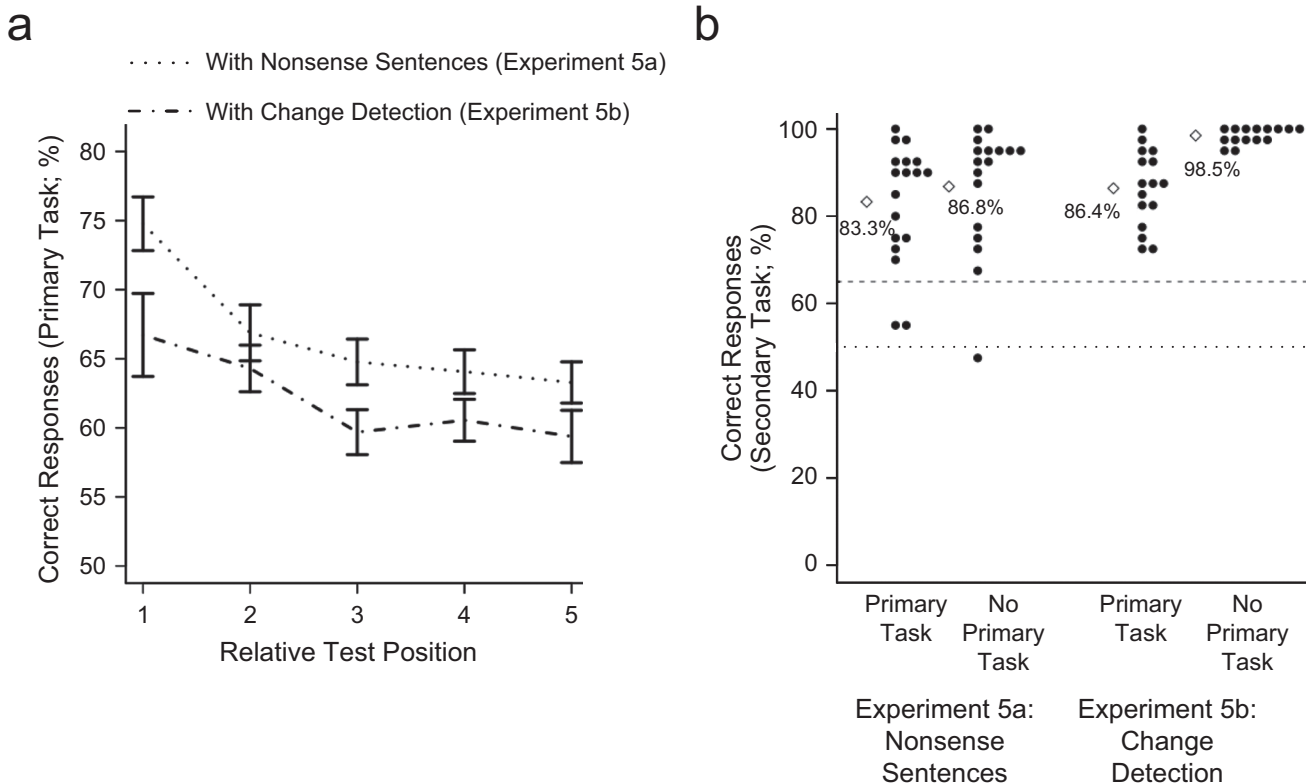

Fig. 4. Results of Experiments 5a and 5b: mean percentage of correct responses on (a) the recognition test and (b) the secondary task. Results for the recognition test are presented as a function of relative test position and the nature of the secondary task (recall of nonsense sentences in Experiment 5a and change detection in Experiment 5b). Error bars represent standard errors of the mean. Results for the secondary task are presented as a function of experiment and whether or not a primary task was performed. Dots indicate individual participants' means, and diamonds indicate averages across the sample. The dotted line denotes the chance level of performance across the sample (50%), and the dashed line shows the chance level of performance for individual participants (65%, as determined by a one-tailed binomial test).

with the primary task in Experiment 5a, participants successfully discriminated words that had occurred in the sentence from words that had not (Table 3 and Fig. 4b; see Fig. S3b in the Supplemental Material for proportions of hits and false alarms). When the secondary task was performed with the primary task in Experiment 5b, participants successfully detected density changes in the grid lines (Fig. 4b and Table 3; see Fig. S3b for proportions of hits and false alarms); this performance did not differ from that on the secondary task in Experiment 5a, $F(1, 32) = 0.6$, $p = .447$, $\eta^2 = .018$, although performance in the sentence task in Experiment 5a was numerically worse. Hence, although the two secondary tasks were matched for difficulty when used as secondary tasks, the nonlinguistic secondary task interfered more with scene processing than the linguistic secondary task did.

To investigate the possibility that one task might be easier than the other when tested in isolation (i.e., without a primary task), we had participants in Experiments 5a and 5b complete their respective secondary tasks without any interfering primary tasks after they finished the main experiment. Performance on the linguistic secondary task was similar when used as a secondary task and when it was presented in isolation (Table 3). In contrast, performance on the change-detection task was almost perfect in the absence of interfering scenes. An ANOVA (excluding 2 participants, 1 in each experiment, who did not complete the second presentation of the secondary tasks) with task type (nonsense sentences vs. change detection) as a between-subjects predictor and task status (secondary task vs. sole task) as a within-subjects predictor revealed main effects of both task type, $F(1, 30) = 4.4$, $p = .044$, $\eta_p^2 = .129$, and task status, $F(1, 30) = 29.2$, $p < .0001$, $\eta_p^2 = .438$, as well as an interaction between these factors, $F(1, 30) = 7.5$, $p = .01$, $\eta_p^2 = .112$.

Although performance on the linguistic task differed only marginally depending on whether participants had to complete a concomitant primary task, $F(1, 16) = 3.9$, $p = .065$, $\eta_p^2 = .197$, performance on the nonlinguistic secondary task was markedly improved when the task was presented in isolation, $F(1, 16) = 36.0$, $p < .0001$, $\eta_p^2 = .720$. In other words, not only did the nonlinguistic secondary task interfere more with scene understanding than did the linguistic secondary task, but scene understanding also interfered more with the nonlinguistic secondary task than with the linguistic secondary task. Remarkably, performance on the linguistic secondary task was almost unaffected by concomitant scene understanding; likewise, the comparison of Experiments 1 and 2 revealed 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 scene understanding, even if both the scenes and the linguistic stimuli are presented visually.

General Discussion

In the experiments presented here, we probed the relation between language and conceptual processing at the earliest stages using stimuli 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 paradigm, we showed that participants' grasp of the conceptual meaning of scenes is almost unaffected by a linguistic secondary task and vice versa, but that scene understanding and a nonlinguistic secondary task mutually interfere.

There are three reasons why these results are not simply due to greater use of visual-processing resources or memory resources in the nonlinguistic secondary tasks than in the linguistic secondary tasks. First, stimuli for the linguistic secondary task occluded at least as much surface area in the scenes as did stimuli 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 descriptions, 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, but 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 also 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), but languages such as Turkish have the word order subject-object-verb (e.g., Mary John sees). However, when people have to gesture to communicate 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). This finding suggests 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 that occur 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 et al., 2007) because listening to such verbs might activate conceptual representations that exert top-down influences on motion perception. This could be true even though the processes used to understand verbs and to perceive motion are distinct and independent. In the absence of such top-down effects, linguistic stimuli appear to be analyzed by dedicated linguistic processors at the

earliest processing stages. This finding provides further evidence for the remarkable modularity of processes that analyze different aspects of people's environments.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

Notes

1. Most of the key findings in this experiment held across the other five experiments in this study. Participants performed better when tested on scenes than when tested on descriptions and performed worse on items in later test positions than on items in earlier test positions. However, both when tested on scenes and when tested on descriptions, participants performed significantly above chance on items in all test positions. As Figure 2, Figure 4a, and Table 1 show, analogous effects of test modality and test position were observed in each of the other experiments (these results are not reported in the text).

2. As Table 2 shows, analogous effects of test modality and test position were observed in each of the other between-experiments comparisons. The effects of test position and modality were more pronounced in some experiments than in others, but these differences are not relevant to the main questions addressed in this article.

3. The effect of test modality was somewhat more pronounced in Experiment 4 than in Experiment 2, and the effect of test position somewhat less pronounced. Although multiple factors might have contributed to these differences (e.g., the difficulty of the secondary tasks, the use of everyday meanings in Experiment 4), Experiments 2 and 4 both replicated our crucial result that scene recognition is better with a linguistic secondary task than with a nonlinguistic secondary task.

4. 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 task difficulty between the linguistic and the nonlinguistic secondary task. We increased the presentation duration to 250 ms to vary the stimulus parameters.

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*, 390–412.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.

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*, 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*, 893–924.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

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, USA*, *105*, 9163–9168.

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*, 985–1018.

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

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, *60*, 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, USA*, *99*, 9596–9601.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science*, *18*, 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*, 1203–1209.

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*, 509–522.

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*, 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*, 478–489.

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*, 1163–1175.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 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*, 358–374.

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

Spivey, M., Tyler, M., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, *12*, 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*, 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*, 285–318.

Whorf, B. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). Cambridge, MA: MIT Press.

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*, 992–996.