Working Memory Mediates the Effect of Language on Perception

Nola Klemfuss, William Prinzmetal, Richard B. Ivry

University of California, Berkeley

Correspondence should be sent to:
Richard Ivry
University of California, Berkeley
California 94720
ivry@berkeley.edu
510-642-7146
Abstract

Lupyan and Spivey (2008) recently made the rather strong claim that top-down input from linguistic representations can sharpen visual feature detectors, thus providing a mechanism for language to influence perception. We present two experiments that challenge this hypothesis. In Experiment 1, we use two versions of a visual search task to demonstrate that the benefits of linguistic priming should be attributed to a reduction in the demands on working memory rather than the modification of feature detectors. In Experiment 2, we show that visual search performance can be disrupted when the stimuli automatically activate irrelevant linguistic representations, a result consistent with the idea that linguistic and sensory representations interact at a late stage of processing, one associated with response selection. These results raise a cautionary note: While language can influence performance on a visual search, the influence need not arise from a change in perception per se.

Limit 150 words. I have 146
Language provides a medium for describing the contents of our conscious experience. We use it to share our perceptual experiences, thoughts, and intentions with other individuals. The idea that language guides our cognition was clearly articulated by Benjamin Whorf (1956) who proposed that an individual's conceptual knowledge was shaped by his or her language. There is clear evidence demonstrating that language directs thought (Ervin-Tripp, 1967), influences concepts of time and space (e.g., Boroditsky), and affects memory (e.g., Loftus & Palmer, 1974).

More controversial has been the claim that language has a direct effect on perceptual experience. Kay and Kempton (1984) found that linguistic labels influence decisions in a color categorization task. However, in such tasks, language may affect decision processes as well as perception. Moreover, while linguistic groups show wide variation in the number of basic color terms, their ability to perceive color differences is invariant (Rosch, 1973), a result at odds with the hypothesis that language alters perception.

To reduce the contribution of memory and decision processes, researchers have turned to simple perceptual discrimination tasks to explore the influence of language on perception. Consider a visual search study by Lupyan and Spivey (2008). Participants were shown an array of shapes and made speeded responses, indicating if the display was homogeneous or contained an oddball (Figure 1a). The shapes were the letters "2" and "5", rotated by 90°. In one condition, the stimuli were described by their linguistic labels. In the other condition, the stimuli were referred to as abstract geometric shapes. RTs were faster for the participants who had been given the linguistic labels or spontaneously
noticed that the shapes were rotated letters. Lupyan and Spivey concluded that “… visual perception depends not only on what something looks like, but also on what it means”, (p. 412).

While the Lupyan and Spivey results provide a compelling demonstration that language can influence performance, the mechanisms underlying this interaction remain unclear. The authors suggest that top-down feedback from linguistic representations sharpens visual feature detectors. However, we consider here an alternative account that shifts the emphasis to working memory processes. Consider the task from the participants’ point of view. The RT data indicate that the displays are searched in a serial fashion (Treisman & Gelade, 1980). When targets are familiar, participants compare each display item to an image stored in long-term memory, terminating search when the target is found. With unfamiliar stimuli, the task is much more challenging. The participant must form a mental representation of the first shape and then maintain this representation while comparing it to each successive shape. Unfamiliar shapes are undoubtedly much more difficult to remember than a verbal label, and thus during a trial, it might be necessary to re-encode items for the comparison process to proceed. Furthermore, since the unfamiliar stimuli do not have a representation in long-term memory, the first item must be encoded anew on each trial. We test the memory hypothesis in the following experiment, comparing a replication of Lupyan and Spivey (2008) with a condition in which the demands on working memory are reduced.

Experiment 1

There were four groups of participants in Experiment 1. For two of the groups, the task was similar to that used by Lupyan and Spivey (2008), with participants making
speeded responses to indicate if a display contained a homogenous set of items or contained one oddball. For the two other groups, a cue was present in the center of the display, indicating the target for that trial. Within each display type, one group was given linguistic primes by being told that the displays contained rotated 2’s and 5’s. The other group was told that the stimuli were abstract forms.

The inclusion of a cue was adopted to minimize the demands on working memory. By pairing the search items with a cue of the target, the task is changed from one requiring an implicit matching process (i.e., compare each item to a stored representation) to one requiring an explicit matching process (i.e., compare each item to the cue). If language influences perception by priming visual feature detectors, we would expect that participants given the linguistic labels would exhibit a similar advantage with both types of displays. In contrast, if the labels reduce the demands on working memory, then we would expect this advantage to be eliminated or attenuated when the displays contain an explicit cue.

Method

Participants. 48 participants from the UC Berkeley Research Participation pool were tested. They received class credit for their participation.

Stimuli. The visual search arrays consisted of 4, 6, or 10 white characters, presented on a black background. The characters were arranged in a circle. The characters were either a "5" or "2", rotated 90° clockwise. The characters fit inside a rectangle that measured 9 cm x 9 cm, and participants sat approximately 56 cm from the computer monitor. For the No Cue conditions, a fixation cross was presented at the center of the display. For the Cue groups, the fixation cross was replaced by a cue.
Procedure. The participants were randomly assigned to one of four groups (12/group). The two No Cue (NC) groups provided a replication of Lupyan and Spivey (2008). They were presented with stimulus arrays (Fig 1a) and instructed to identify whether they were all the same character or whether one was different. One group was told that the display contained 2’s and 5’s whereas the other group was told that the displays contained abstract forms. For the two Cue groups, the fixation point was replaced with a visual cue (Fig 1b). For these participants, the task was to determine if an array item matched the cue. As with the NC conditions, one group was told that the display consisted of 2’s and 5’s and the other half were told that the display contained abstract forms.

Each trial started with the onset of either a fixation cross (NC groups) or cue (CUE groups). 300 ms later, the search array was added to the display. Participants responded on one of two keys, indicating if the display contained one item that was different than the other display items. Following the response, a feedback screen was presented for 1000 ms with the word "Great!" on correct trials and "Incorrect" on error trials. The screen was then blanked for a 500 ms inter-trial interval. Average RT and accuracy were displayed at the end of each block.

The experiment consisted of a practice block of 12 trials and four test blocks of 60 trials each. At the beginning of each block, participants in both the NC and Cue groups were informed which character would be the target for that block of trials. Each character served as the oddball for two of the blocks. The oddball was present on 50% of trials, positioned on the right and left side of the screen with equal frequency.
Results and Discussion

Overall, participants were correct on 89% of the trials and there was no indication of a speed-accuracy trade-off. Excluding incorrect trials, we analyzed the RT data (Fig 2) in a 3-way ANOVA with two between-subject factors (1) task description (linguistic vs. abstract) and (2) task set (NC vs. Cue), and one within-subject factor, (3) set size (4, 6, or 10 items). The effect of set size was highly reliable, consistent with a serial search process, $F(2,88) = 289.35, p < .0001$. Importantly, the two-way interaction of task description and task set was reliable, $F(1,44) = 4.96, p<.05$

For the NC groups, the data replicated Lupyan and Spivey (2008). Participants instructed to view the characters as rotated numbers (linguistic description) performed 303 ms faster than participants for whom the characters were described as abstract symbols, $F(1.22) = 10.12, p<.001$. This linguistic advantage disappeared (-45.6 ms) in the Cue groups, $F(1,22) = 0.072, ns$. Thus, when the demands on working memory were reduced by the inclusion of a cue, we observed no linguistic benefit. In the overall ANOVA, the three-way interaction was reliable, $F(2,88) = 6.23 p < .005$, reflecting the fact that the linguistic advantage was greatest for the largest set size, but only for the NC group.

The results of Experiment 1 challenge the hypothesis that linguistic labels provide a top-down priming input to perceptual feature detectors. If this were so, then we would expect to observe a linguistic advantage regardless of whether the task involved a standard visual search (oddball detection) or our modified, matching task. With either display, the linguistic description of the characters should have provided a similar priming signal.
In contrast, the results are consistent with our working memory account. By this hypothesis, the linguistic advantage in the NC condition (as well as in the Lupyan and Spivey, 2008 study) arises from the fact that participants must compare items in working memory during serial search. As predicted by the working memory hypothesis, the linguistic advantage was abolished when subjects are reminded of the target with a visual cue.

Experiment 2

We take a different approach in Experiment 2, testing the prediction that linguistic labels disrupt processing when this information is task irrelevant. To this end, we had participants make an oddball judgment based on a physical attribute, line thickness. We presented upright or rotated 2s and 5s, assuming that upright numbers would be encoded as linguistic symbols, while rotated numbers would not. If language enhances perception, performance should be better for the upright displays. Alternatively, the automatic activation of linguistic codes for the upright displays may produce response conflict given that this information is irrelevant to the task.

Method

Participants: Twelve participants received class credit for completing the study.

Stimuli: Thick and thin versions of each character were created. The thick version was the same as in Experiment 1. For the thin version, the stroke thickness of each character was halved.

Procedure: Each trial began with the onset of a fixation cross for 300 ms. An array of four characters was then added to the display and remained visible for 450 ms
Participants were instructed to indicate whether the four characters had the same thickness, or whether one was different. The characters were either displayed in an upright orientation or rotated, with the same orientation used for all four items in a given display. Upright and rotated trials were randomized within a block. Each participant completed four blocks of 80 trials each. All other aspects were identical to Experiment 1.

Results and Discussion

Participants were slower when the characters were upright compared to when they were rotated, $F(1,11) = 7.67$, $p < .01$. The mean RT was 375 ms for the upright displays and 348 ms for the rotated displays, for an average cost of 27 ms ($SE_{diff} = 5.6$ ms). Participants averaged 92% correct, and there was no evidence of a speed accuracy trade-off.

We designed this experiment under the assumption that the upright displays would produce automatic and rapid activation of the lexical codes associated with the numbers. There are at least two distinct ways in which linguistic codes might disrupt performance. Perceptually, linguistic encoding encourages holistic processing. If parts of a number are thick, there is a bias to assume the whole number is thick. (Fuchs, 1923; Khurana, 1998; Prinzmetal & Keysar, 1989). This bias may be reduced for the less familiar, rotated shapes, which may be perceived as separate lines.

Alternatively, linguistic codes may constitute a form of irrelevant and potentially disruptive input to decision processes (e.g., response selection). Decision-based interference from irrelevant dimensions is ubiquitous in cognitive psychology (Kornblum, Hasbroucq, & Osman, 1990). Indeed, the Stroop effect (Macleod, 1991), is
generally accepted to reflect interference from the automatic activation of the lexical
codes of word names when the task requires judging the stimulus color.

General Discussion

In the current study, we set out to sharpen the focus on how language influences
perception. We considered different ways in which language might influence perceptual
performance. The first experiment was inspired by a recent study in which performance
on a visual search task was found to be markedly improved when the participants had
been instructed to view the search items as linguistic entities, compared to when the
instructions led the participants to view the items as abstract shapes (Lupyan & Spivey,
2008). The authors of that study had championed an interpretation, and in fact, provided
a computational model, in which over-learned associative links between linguistic and
perceptual representations allowed top-down effects of a linguistic cue to sharpen
perceptual analysis.

While this idea is certainly plausible, we offered an alternative hypothesis: The
linguistic cues reduced the burden on working memory. We tested this hypothesis by
using identical search displays, with the one addition of a visual cue, assumed to
minimize the demands on working memory. Under these conditions, we failed to
observe any performance differences between participants given linguistic and non-
linguistic prompts. These results challenge the perceptual account, given that top-down
priming effects should have remained operative. Instead, the results favor our working
memory account, pointing to subtle ways in which performance entails a host of complex
operations.
We exploited this idea in a second experiment, designing a task in which language might hinder perceptual performance. Here we again used a visual search task, but one in which participants had to determine if a display item had a unique physical feature (i.e., font thickness). For this task, linguistic representations were irrelevant. Nonetheless, when the shapes were oriented to facilitate reading, a cost in RT was observed, presumably due to the automatic activation of irrelevant linguistic representations. Our study here is reminiscent of the Stroop effect, widely regarded as reflecting the competition of different representations for response selection (MacLeod, 1991). A related account also extends to recent work, including research from our lab, showing that the effects of language on visual search performance is more pronounced in the right visual field (Gilbert, Regier, Kay, & Ivry, 2006; 2008). We assume that the activation of (left hemisphere based) linguistic representations is faster from the right visual field, and thus, exert a stronger influence on performance, compared to inputs arising in the left visual field.

This pair of experiments has taught us that linguistic coding can be a useful tool to aid processing. However, the influence that language has on our experience can both facilitate and impede performance. In cases where there are memory demands, language provides a way to categorize familiar stimuli, and in visual search, linguistic coding would provide an efficient mechanism to encode and compare the display items. However, when the linguistic nature of the stimulus is irrelevant to the task, language may also hurt performance. These findings enrich how we think about the interaction between language and perception. The words we speak simultaneously reinforce and compete with the dynamic world we perceive and experience. However, the results they
also provide a cautionary note warning: while there is evidence for an interaction between language and perception, we must be vigilant in characterizing the manner in which language and perception interact.
Figure Captions

Figure 1. Sample stimulus displays for the No Cue (left) and Cue (right) conditions in Experiment 1. Participants made speeded responses, indicating if the display items were homogenous or contained an oddball. In the Cue conditions, the oddball matched the central cue.

Figure 2. Reaction time data for Experiment 1. Confidence intervals in the figure were calculated using the 3-way interaction as described by Loftus and Masson (1994).

Figure 3. Sample stimulus displays for Experiment 2. The four display items were letters, rotated 90° clockwise (left) or upright (right). Participants made speeded responses, indicating if the font thickness for the display items were homogenous or contained an oddball.
Figure 1. Sample stimulus displays for the No Cue (left) and Cue (right) conditions in Experiment 1. Participants made speeded responses, indicating if the display items were homogenous or contained an oddball. In the Cue conditions, the oddball matched the central cue.
Figure 2. Reaction time data for Experiment 1. Confidence intervals in the figure were calculated using the 3-way interaction as described by Loftus and Masson (1994).
Figure 3. Sample stimulus displays for Experiment 2. The four display items were letters, rotated 90° clockwise (A) or upright (B). Participants made speeded responses, indicating if the font thickness for the displays items were homogenous or contained an oddball.
References


