Title
Now You See It, Now You Don't: Verbal but not visual cues facilitate visual object detection

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN
1069-7977

Authors
Lupyan, Gary
Spivey, Micheal J.

Publication Date
2008

Peer reviewed

Now You See It, Now You Don't: Verbal But Not Visual Cues Facilitate Visual Object Detection

Gary Lupyan (lupyan@cornell.edu)
Michael J. Spivey (spivey@cornell.edu)
Department of Psychology, Cornell University
Ithaca, NY 15850 USA

Abstract

Does knowing what one is about to see make it easier to see it? The answer may depend on the source of the knowledge. Participants completed an object detection task in which they made an object-presence or -absence decision to briefly-presented letters. Hearing the letter name prior to the detection task facilitated detection (*d*'), but seeing a preview of the to-be-detected stimulus did not. Follow-up experiments explored the role of position uncertainty and cue validity. The results suggest that auditory labels produce a modulatory effect on visual processing such that immediately after hearing a category label processing of visual items associated with the label is facilitated even when the exact position of the to-be-detected stimulus is unknown. These results indicate that cognition has a much stronger top-down influence on perceptual processing than previously thought.

Keywords: visual perception; language; labeling; object detection; crossmodal cues

Introduction

A great deal of evidence shows that allocating visual attention to a location improves reaction times to probes appearing in that location (Posner, Snyder, & Davidson, 1980), increases detection sensitivity (Hawkins et al., 1990) and even increases perceived stimulus contrast (Carrasco, Ling, & Read, 2004). In addition to its spatial properties, the spread of attention is affected by specific objects: cuing an object facilitates the detection of a probe within the cued (or even a similar uncued object) compared to equidistant positions outside of the object (e.g., Egly, Driver, & Rafal, 1994; Mozer & Vecera, 2005). These lines of evidence comprise within-vision effects: visually presented cues affect attention to visually presented stimuli. However, information from outside of vision has also been shown to affect visual processing. For instance, a tactile cue in one location can improve discrimination for visual stimuli at that location, an effect that has been shown to arise from modulation of visual cortex by multimodal parietal regions (Macaluso, Frith, & Driver, 2000).

There is now accumulating evidence that higher-level semantic information presented in the auditory modality can influence visual perception in some surprising ways. For instance, auditory processing of verbs associated with particular directions of motion (e.g., fly, bomb) increases sensitivity to the congruent motion direction in random-dot kinematograms (Meteyard, Bahrami, & Vigliocco, 2007). In visual search tasks, hearing words that label the target or distractors improves the speed and efficiency (RT slope as a function of display size) of search. For instance, when searching for the number 2 among 5's, participants are faster to find the target (whose identity is always known and remains constant) when they actually hear "find the two" immediately prior to the search trial (Lupyan, 2007a). The facilitation of visual processing by verbal labels depends on the existence of a pre-existing association between the label and the visual stimulus and is disrupted by manipulations that preserve the low-level visual features of a stimulus but alter its association with the named category (e.g., through a mirror reversal) (Lupyan, 2008). These findings leave open the question of whether hearing verbal labels can affect the visual processing at a still more basic level in tasks that neither require nor allow naming. Here, we test whether object names influence participants' ability to detect briefly presented objects. We predicted that hearing verbal labels would facilitate detection of stimuli matching the label. By contrasting the effects of auditory cues consisting of the verbal label with visual cues consisting of a preview of the actual stimulus that was to be detected, we were able to investigate whether effects of cues on object-detection were specific to spoken verbal labels.

Understanding the interaction between verbal labels and visual processing is important for a number of reasons. First, as language processing becomes better understood at a neural level, a cross-comparison between linguistic neural processes and better-understood visual neural processes will be facilitated. Second, findings that support real-time linguistic influences on visual processes encourage a rethinking of modular theoretical accounts of the visual system.

Third, understanding of how verbal labels affect visual processing can help to better understand reports of cross-linguistic differences in visual tasks (e.g., Winawer et al., 2007) and thus inform the "language and thought" debate.

Experiment 1

Subjects

A total of 92 Cornell University undergraduates volunteered for four experiments in exchange for course credit. Forty-two participated in Experiment 1, split randomly into a visual-cue and auditory-cue conditions. Experiments 2-4 included the auditory condition only. Twenty subjects each participated in Experiments 2 and 3, and 10 in Experiment

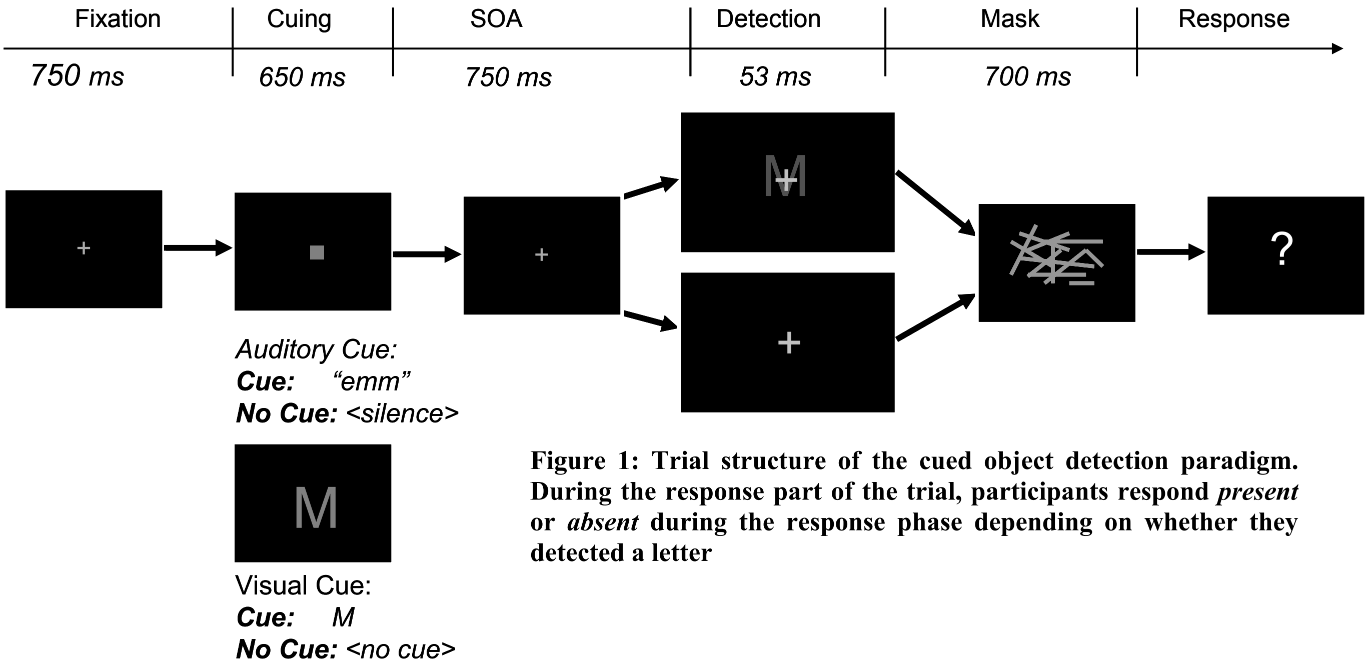

Figure 1: Trial structure of the cued object detection paradigm. During the response part of the trial, participants respond *present* or *absent* during the response phase depending on whether they detected a letter

4. All were naïve to the hypothesis and none participated in more than one experiment.

Stimuli

The stimuli were uppercase English letters, rendered in the Arial font, and subtended 2.2° (Vertical) × 1.8° (Horizontal) visual-angle. Letters were chosen as stimuli because of the strong pre-existing associations between their visual forms and their names. The letters used in the main part of the experiment were B, E, F, H, M, O, R, U, V, Y. The visual cues were identical to the stimuli to-be-detected. The auditory cues were pre-recorded letter names of a female speaker, originally designed for telephone voice XML systems. The recordings are available at: http://community.voxeo.com/library/audio/prompts/alphabet/index.jsp. The letter names for the selected stimuli, as recorded, were approximately 650 ms in duration.

Procedure

The basic trial design is illustrated in Figure 1. The participants' task was to detect uppercase letters, and respond *present* if they detected *any* letter, and *absent* if they thought no letter was present. On some trials, a cue preceded the detection task allowing us to study the effect of the cue on detection performance.

Participants were randomly assigned to an *auditory* or *visual* cue condition. The conditions differed only in what happened during the cuing part of the trial. In the *visual* condition, a letter cue was presented on half of the trials alerting the participants to the identity of the to-be detected stimulus. On the remaining trials, the fixation cross was replaced by a gray square for a duration identical to the cue duration (650 ms). The auditory condition was logically identical except the cue was auditory, consisting of the letter name of the to-be detected letter (e.g., “emm” for M). The cue was presented on exactly half the trials. During the cue interval, the fixation cross was replaced by a gray square for a constant 650 ms. The display then reverted back to the fixation cross for 750 ms after which the detection part of the trial began. On exactly half of the trials a faint uppercase letter was flashed for 53 ms and was masked by randomly oriented line segments. On the remaining half of the trials, no letter was present during the 53 ms interval. The mask for each trial was selected randomly from 100 pre-generated masks, ensuring participants could not anticipate the perceptual details of the masking stimulus.

To observe the effect of the cue on object detection, the task had to be difficult enough to avoid ceiling-level performance. Pilot work revealed that participants were able to detect single letters rendered in a white font on a black background even when they were presented for a single screen refresh (13.3 ms). We thus decided to manipulate the contrast of the stimuli relative to the black background. Because we expected large individual differences in detection ability, we adjusted the contrast level for each participant by using a brief staircasing procedure during which the contrast of the to-be-detected stimulus was lowered following a correct response and increased following an incorrect response.

Each experimental session began with the staircasing procedure starting with plainly visible letters, and lasting 75 trials, sufficient to produce final hit rates of approximately 60%. The first 15 trials were considered practice. Feedback in the form of a buzzing sound was provided following incorrect responses for these practice trials only. During staircasing the detection trials were not cued and all 26 letters (randomly selected on each trial) were used as stimuli.

The main part of the experiment consisted of 6 blocks of 40 trials (stimulus-present vs. stimulus-absent × auditory (visual) cue vs. no cue × stimulus identity). Trial order was random with the target present on exactly half of the trials.

On exactly half of the target-present trials, the target was preceded by a cue. Participants gave 2-alternative target *present* / *absent* responses using a gamepad controller. Response mapping (left hand *present* vs. right-hand *present*) was counterbalanced between participants.

Results and Discussion

There were no overall differences in hit rates or false alarms between the visual and auditory conditions: $Hits_V$=.58, $Hits_A$=.60, two-tailed t-test, $t(40) < 1$; FA_{Vis}=.18, FA_{Aud}=.12, $t(40) = 1.16$, $p = .26$. Auditory cues increased the hit rates from .56 to .64, paired t-test, $t(20) = 3.03$, p=.007. There was no corresponding increase in hits in the visual condition, $t(20)$=1.18, n.s. Auditory cues marginally increased false alarms from .10 to .13, $t(20) = 2.08$, $p = .05$. Visual cues did not affect false alarms, $t(20) < 1$.

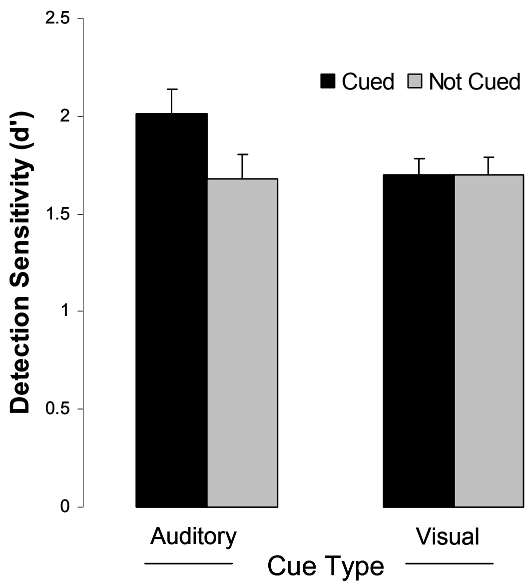

Figure 2: Effects of auditory and visual cues on the detection of cued visual objects (Experiment 1). Bars indicate 1 SE of the within-subject difference between the means.

To determine the effect of cues on detection sensitivity, we computed *d'* for each of the four cuing conditions (visual-cued vs. uncued and auditory cued vs. uncued). The results of this signal detection analysis are shown in Figure 2. Auditory cues significantly increased detection sensitivity, $t(20) = 2.64$, $p = .016$. Visual cues did not, $t(20)<1$. The difference in the cuing effect was reflected in a significant cue-type × cue-presence interaction, $t(40) = 2.22$, $p = .032$.

Both the visual and auditory cues informed the participants of what letter needed to be detected in the upcoming trial. The visual cues additionally provided participants with an actual preview of the to-be-detected stimulus. The auditory cues in contrast required participants to "translate" the auditory information (letter name) into a visual code. One would therefore expect that if cuing can affect object detection—itself an unanswered question—then cues identical to the target stimulus should be more effective, as is the case in more complex tasks like visual search (Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). Yet, visual cues did not affect visual detection performance in this task. The reason for the failure to find facilitation following visual cuing is not fully understood, but some possibilities are discussed in the General Discussion.

Experiment 2

A possible explanation for both the effect of cues on subsequent object detection and the finding that only auditory cues improved detection is that detection ability is improved simply by the attentional arousal induced via the auditory stimulation itself, rather than the information it conveys. For example, it may be that hearing sounds produces a transient improvement in performance by increasing vigilance (e.g., Pollack & Knaff, 1958). Indeed, distinctive sounds, such as a high tone embedded in a sequence of low tones have been shown to improve detection of visual targets, although only when the targets were presented in synchrony with the tone (Vroomen & de Gelder, 2000). In any case, if the effect of auditory cues in Experiment 1is a simple consequence of hearing sounds rather than a result of letter names affecting perception, then including an irrelevant sound stimulus during the no-cue trial should eliminate the advantage observed during the letter-name cue trials.

Stimuli

The stimuli were identical to Experiment 1 except for the inclusion of a new auditory stimulus used during the no-cue trials. The stimulus consisted of a female audio recording of the word "ready."

Procedure

The procedure was identical to the auditory condition of Experiment 1 except that now both the cue and no-cue detection trials were preceded by auditory stimuli. The cue trials included letter-names, as before. For the remaining trials, participants heard the uninformative word "ready" during the cuing interval.

Results and Discussion

The results were very similar to those of Experiment 1 with participants demonstrating superior detection sensitivity on the trials in which they heard a letter-name cue compared to those in which they heard the uninformative word "ready," $t(19) = 2.25$, $p = .036$. Cuing had no effect on false alarms (FA_{cue}=.19, $FA_{no\text{-}cue}$=.21), $t(19) < 1$, but a highly significant effect on hit rates (H_{cue}=.66, $H_{no\text{-}cue}$=.56), $t(19) = 2.73$, p =.013 (see Figure 3, left).

These results allow us to rule out the possibility that the detection advantage on the cued auditory trials arose simply from the alerting nature of the auditory cue. However, the cuing effect in the present experiment was somewhat smaller than that observed in Experiment 1, suggesting that general arousal following auditory stimulation may contribute to the facilitatory effect of the auditory cue.

Experiment 3

One way in which the auditory cues may have increased detection sensitivity is by encouraging participants to mentally image the named letter. Indeed, instructing participants to image a specific letter in a specific location has been shown to increase detection sensitivity (Farah, 1985; cf. Segal & Fusella, 1970). The instruction to imagine a specific letter was effective only when there was an exact match between the imaged and actual stimulus location (a finding that was used to support a common locus of perception and mental imagery). This finding is in line with later studies showing that mental imagery produces a local lowering of detection criterion inside the contours of the imaged figure (Farah, 1989). If auditory cues facilitate object detection by encouraging mental imagery, then the advantage should be specific to the position in which the stimulus is imaged and should disappear with spatial uncertainty of the

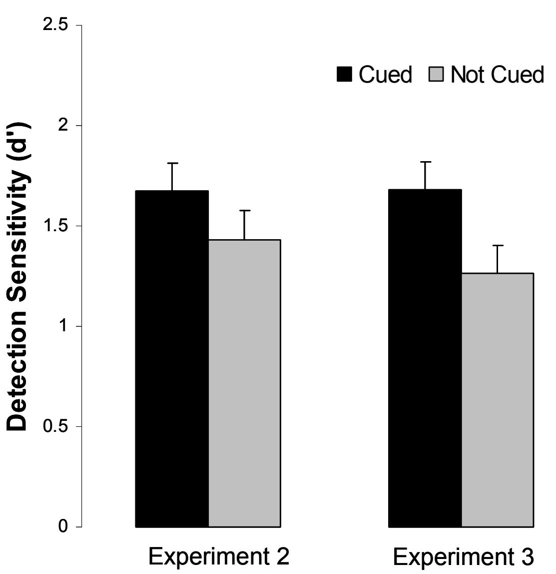

Figure 3: Effects of auditory labels on visual object detection in Experiment 2 which contrasts informative auditory cues with an uninformative auditory sound, and Experiment 3 which adds stimulus jitter. Bars indicate 1 SE of the within-subject difference in the means.

to-be-detected object. Alternatively, if the effects of auditory names on object detection have as their locus a more position-invariant stimulus representation, then varying the position should have no effect on the advantage conferred by auditory cues.

Stimuli

The stimuli were identical to Experiment 1.

Procedure

The procedure was identical to the auditory condition of Experiment 1 except the to-be-detected stimulus was now displayed with some spatial uncertainty. All stimuli were still displayed well within foveal vision, but their position was randomly jittered in the horizontal and vertical dimensions from a minimum of 0.5^o from fixation to a maximum of 1.5^o (measured from fixation to the center of the letter).

Results and Discussion

The results mirrored those of Experiment 1 with detection performance on the cued trials exceeding performance on the non-cued trials, $t(19)$=2.99, p = .007 (see Figure 3, right). As in Experiment 1, the sensitivity advantage arose from greater hit rates: auditory cues increased hit rates from .47 to .60, $t(19)$=3.40, p = .003. Cuing had no reliable effect on false alarms (FA_{cue}=.16, $FA_{no\text{-}cue}$=.13), $t(19) = 1.11$, p = .28.

Varying the position of the to-be-detected stimulus did not eliminate the facilitatory effect of auditory cues on object detection. This result suggests that even though the cues may encourage participants to maintain a mental image of the cued letter (indeed, many participants reported using this strategy in both the auditory and visual conditions of Experiment 1), the cuing effect has as its locus a somewhat position-invariant representation.

Experiment 4

A critical limitation of Experiments 1-3 is that the cues always validly predicted the to-be-detected stimulus. Although the cues did not predict stimulus-presence, when present, the cue and stimulus always matched. The goal of Experiment 4 was to assess the specificity of the cuing effect by contrasting valid cues (those that matched the to-be-detected stimulus) with invalid cues (those that did not match the to-be-detected stimulus.

Stimuli

The stimuli were identical to Experiment 1.

Procedure

The procedure was identical to the auditory condition of Experiment 1 with the exception that the cued stimulus-present trials were evenly divided into cue-valid and cue-invalid trials. Thus, in this experiment cues not only did not predict stimulus presence, but also did not predict the identity of the stimulus should it appear.

Results and Discussion

Only valid cues improved detection sensitivity (Figure 4). Planned comparisons using pairwise t-tests showed that sensitivity (*d'*) was significantly higher in valid trials than invalid trials, $t(9) = 2.41$, $p = .039$. A comparison of valid and no-cue trials once again revealed a significant advantage for the former, $t(9) = 3.10$, $p = .013$. There was no significant difference between invalid and no-cue trials, $t(9) = 1.65$, $p = .13$. As in Experiments 1-3, the difference in *d'* arose from differences in hit rates: $H_{valid\text{-}cue}$=.73, $H_{invalid\text{-}cue}$=.64, $H_{no\text{-}cue}$=.52. Paired t-tests of hit-rates mirrored the *d'* analysis.

Detection sensitivity was improved only when the auditory cues matched the to-be-detected stimulus (i.e., when the cues were valid). This result further supports the hypothesis that auditorily presented object names have a facilitatory effect on the subsequent detection of objects matching the verbal label. Many questions remain regarding both cue specificity and cue validity. For example, would hearing "emm" facilitate the detection of both uppercase and lowercase M's? Finally, although in this experiment, no significant difference was found between the invalid-cue and no-cue conditions, it appears that invalid cues may offer a slight benefit to object detection over no cues at all. Preliminary studies indicate that the relationship between valid and invalid cues changes over the course of the experiment, with invalid cues becoming increasingly *more* effective over time.

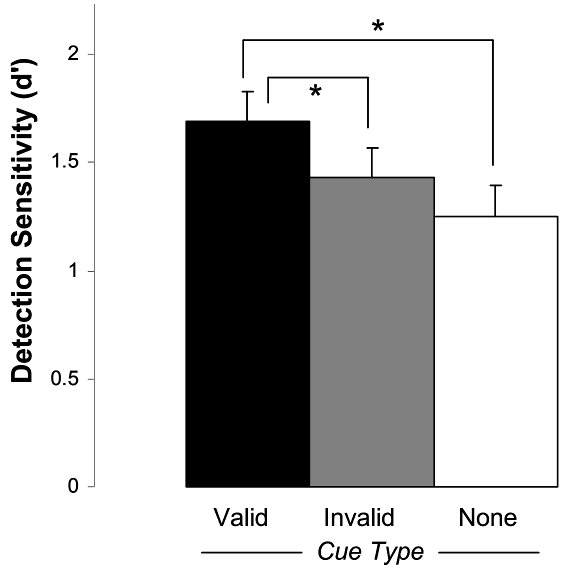

Figure 4: Results from Experiment 4. Bars indicate 1 SE of the within-subject difference in the means. Asterisks indicate significant differences between condition means at *p* < .05.

General Discussion

Knowing what stimulus needs to be detected improved detection sensitivity. This finding alone is a critical challenge to claims of the cognitive impenetrability of early vision (Pylyshyn, 1999) because it provides a demonstration of information outside of the visual system affecting performance on a rather low-level visual task. Recall that participants did not need to identify the stimuli, merely detect them, though casual inspection revealed that correct detection generally engendered correct recognition as well, confirming the findings of Grill-Spector and Kanwisher (2005). Surprisingly, only auditory cues naming the to-be-detected object improved detection performance. Getting a preview of the actual stimulus that was to be detected had no effect on detection sensitivity. Might the advantage of auditory cues arise from auditory stimulation inducing a general enhancement in visual detection? Experiment 2 contrasted stimulus-relevant cues (letter names) with stimulus-irrelevant auditory cues (a "ready" prompt). This manipulation failed to eliminate the detection-advantage of stimulus-relevant auditory cues. Experiment 3 showed that the detection advantage following auditory cues persisted even when the exact location of the to-be-detected stimulus was not determined, suggesting that the effect induced by the auditory labels has a degree of position invariance (cf. Farah, 1985; Farah, 1989). In Experiments 1-3, the cues were always valid. Although the cues did not predict whether a stimulus would be present, if a stimulus was present, it was always congruent with the cue. Experiment 4 explored the effects of invalid cues on detection performance and showed that only valid cues reliably improved detection sensitivity. At present, the effect of invalid visual cues is not well understood.

One way to understand the present findings is by conceiving of verbal labels as providing modulatory feedback to the visual system (Lupyan, 2007b). Feedback connections are omnipresent in the visual system; task demands and visual context have been shown to affect response properties even of neurons in the primary visual cortex (Lamme & Roelfsema, 2000). We think it is unlikely that auditory labels affect the visual system at the lowest levels, if only because such effects would likely be location specific (contra Experiment 3). A more likely possibility is that the verbal labels modulate processing in the visual areas of the (highly polymodal) orbitofrontal cortex (OFC) and thereby provide a prediction signal which then affects detection of stimuli matching the label through feedback connections to inferotemporal cortex (IT). Such real-time modulation is made possible by fast-conducting projections between OFC and IT (see Kveraga, Ghuman, & Bar, 2007 for a discussion of the role of OFC in visual prediction). Interestingly, OFC appears to be most involved in modulation of stimuli containing low spatial frequencies (subserved by the magnocellular visual stream) (Kveraga, Boshyan, & Bar, 2007), leading to the prediction that the facilitation effect induced by verbal labels may also be limited to the low-spatial frequency components of the named stimuli (the achromatic and low contrast letters used here fulfilled this criterion).

Several unanswered questions remain: What are the temporal dynamics of the facilitation effect of auditory labels on object detection? Because facilitated detection is observed in a design that intermixes cued and uncued trials, the facilitation must have a transient component (see also Lupyan, 2008). The duration and temporal profile of this putatively transient facilitation is unknown. Differences in the dynamics of the auditory and visual-cue conditions may explain why we failed to find a facilitation of visual cues on object detection. It may be that visual cues do in fact facilitate object detection, but this facilitation does not endure the 750 ms interval between cue offset and stimulus onset.

Familiar letters—the stimuli used in the present work—have strong auditory associations (i.e., a salient feature of the stimulus M is the sound "emm"). It remains to be seen whether similar effects can be obtained for familiar stimuli with less salient associations between the category label and the visual properties of the labeled stimulus (e.g., "chair," "flower", "insect").

The cued object-detection paradigm introduced here promises to be a useful tool for exploring the role of language in perceptual processing. For example, an additional line of questioning to be explored is whether learning to associate unfamiliar stimuli with novel verbal labels facilitates detection of these stimuli. A positive finding would further illuminate the mechanisms by which learning different languages can induce differences in perceptual processing and experience (Roberson, Davidoff, Davies, & Shapiro, 2005; Winawer et al., 2007).

Acknowledgments

The first author thanks Noa Holtzman and Stephanie Feit for their indispensable help with data collection.

References

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7,* 308-313.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting Visual-Attention Between Objects and Locations - Evidence from Normal and Parietal Lesion Subjects. *Journal of Experimental Psychology-General, 123,* 161-177.

Farah, M. J. (1985). Psychophysical Evidence for A Shared Representational Medium for Mental Images and Percepts. *Journal of Experimental Psychology-General, 114,* 91-103.

Farah, M. J. (1989). Mechanisms of Imagery Perception Interaction. *Journal of Experimental Psychology-Human Perception and Performance, 15,* 203-211.

Grill-Spector, K. & Kanwisher, N. (2005). Visual recognition - As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152-160.

Hawkins, H. L., Hillyard, S. A., Luck, S. J., Downing, C. J., Mouloua, M., & Woodward, D. P. (1990). Visual-Attention Modulates Signal Detectability. *Journal of Experimental Psychology-Human Perception and Performance, 16,* 802-811.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular Projections as the Trigger of Top-Down Facilitation in Recognition. *The Journal of Neuroscience, 27,* 13232-13240.

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition, 65,* 145-168.

Lamme, V. A. F. & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571-579.

Lupyan, G. (2007a). Reuniting Categories, Language, and Perception. In D.S.McNamara & J. G. Trafton (Eds.), *Twenty-Ninth Annual Meeting of the Cognitive Science Society* (pp. 1247-1252). Austin, TX: Cognitive Science Society.

Lupyan, G. (2007b). *The Label Feedback Hypothesis: Linguistic Influences on Visual Processing.* PhD. Thesis. Carnegie Mellon University.

Lupyan, G. The Conceptual Grouping Effect: Categories Matter (and named categories matter more). *Cognition,* (in press).

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of Human Visual Cortex by Crossmodal Spatial Attention. *Science, 289,* 1206-1208.

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

Mozer, M. C. & Vecera, S. P. (2005). Object- and space-based attention. In L.Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 130-134). New York: Elsevier.

Pollack, I. & Knaff, P. R. (1958). Maintenance of Alertness by A Loud Auditory Signal. *Journal of the Acoustical Society of America, 30,* 1013-1016.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the Detection of Signals. *Journal of Experimental Psychology-General, 109,* 160-174.

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

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50,* 378-411.

Segal, S. J. & Fusella, V. (1970). Influence of Imaged Pictures and Sounds on Detection of Visual and Auditory Signals. *Journal of Experimental Psychology, 83,* 458-&.

Vroomen, J. & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology-Human Perception and Performance, 26,* 1583-1590.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 7780-7785.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44,* 1411-1426.