Acta Psychologica 170 (2016) 19–31

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

I can see what you are saying: Auditory labels reduce visual search times☆

Kit W. Cho

University of Houston-Downtown, Department of Social Sciences, 1 Main St., Houston, TX 77002, United States

ARTICLE INFO

Article history:
Received 16 September 2015
Received in revised form 16 May 2016
Accepted 12 June 2016
Available online xxxx

Keywords:
Auditory label
Language
Object recognition
Perception
Visual search

ABSTRACT

The present study explored the self-directed-speech effect, the finding that relative to silent reading of a label (e.g., DOG), saying it aloud reduces visual search reaction times (RTs) for locating a target picture among distractors. Experiment 1 examined whether this effect is due to a confound in the differences in the number of cues in self-directed speech (two) vs. silent reading (one) and tested whether self-articulation is required for the effect. The results showed that self-articulation is not required and that merely hearing the auditory label reduces visual search RTs relative to silent reading. This finding also rules out the number of cues confound. Experiment 2 examined whether hearing an auditory label activates more prototypical features of the label's referent and whether the auditory-label benefit is moderated by the target's imagery concordance (the degree to which the target picture matches the mental picture that is activated by a written label for the target). When the target imagery concordance was high, RTs following the presentation of a high prototypicality picture or auditory cue were comparable and shorter than RTs following a visual label or low prototypicality picture cue. However, when the target imagery concordance was low, RTs following an auditory cue were shorter than the comparable RTs following the picture cues and visual-label cue. The results suggest that an auditory label activates both prototypical and atypical features of a concept and can facilitate visual search RTs even when compared to picture primes.

1. Introduction

Features associated with a concept can be activated in a variety of ways. According to Lupyan and colleagues (see, Lupyan, 2012a, 2012b, for reviews), the features associated with an object activated by a label will be different from the features of that same object when activated by a nonlabel that is directly associated with that object (e.g., hearing the sound that the object typically makes—"WOOF WOOF" for the concept DOG). Specifically, they (e.g., Lupyan & Swingley, 2012; Lupyan & Thompson-Schill, 2012) assert that a label will activate the visual and non-visual features that are most closely associated with (or more prototypical of, see Rosch, 1975) its referent relative to a nonlabel. Thus, when the label DOG is presented, a person is more likely to think of a dog whose features resemble those of a Golden Retriever or Labrador (i.e., large body, large tail, hairy) than of a Chihuahua (i.e., a small body, minimal body hair). Along with the activation of prototypical features associated with a concept, Lupyan and colleagues assert that the activation of features not correlated with the concept is minimized (Harnad, 2005; Lupyan, 2008; Lupyan & Spivey, 2012a). For example, if a person is told that an unfamiliar object is a CHAIR, features of that object will be more likely to be encoded with chair-like characteristics and any idiosyncratic visual features associated with the object (e.g., an atypical upholstery pattern) will be lost (Lupyan, 2008). Thus, labels provide top-down knowledge that can influence the mental representations of unfamiliar objects.

Labels can also have a top-down influence on visual search. Using a visual world paradigm, Dahan and Tanenhaus (2005) showed that participants are more likely to look at a picture of a rope when asked to look for a snake. Huettig and Altmann (2011) showed that eye-movements were more likely to be directed at distractors that shared the same diagnostic color as the target (e.g., hearing the word "FROG" directs attention to object pictures that are green). Top-down knowledge from labels also reduces RTs to recognize object pictures as recently studied when they are presented in a viewpoint (orientation) different from how they were learned (Collins & Curby, 2013). Such top-down benefits produced by labels occur in the extrastriate ventral area (Mazer & Gallant, 2003) and occur rapidly, with electrophysiological studies showing that top-down processing can affect visual processing as early as 10 ms after stimulus onset (Gilbert & Sigman, 2007; Lamme & Roelfsema, 2000).

☆ This article is based on a PhD dissertation submitted to the University at Albany, State University of New York. I thank the members of my dissertation committee James H. Neely (chair), W. Trammell Neill, and Laurie B. Feldman, for their advice and guidance on the project. I am indebted to Nicole A. Arnold, Vanessa Chan, Conor Egan, Andrew Holman, Briana F. Moglia, and Lukas K. Sotola, who assisted me with data collection for this project and to Gregory Simmons and Stephanie Crocco for their assistance with recording the stimuli.

http://dx.doi.org/10.1016/j.actpsy.2016.06.003
0001-6918/

2. "Uninformative" labels affect the activation of features associated with a concept

Lupyan and Thompson-Schill (2012) showed that a label (e.g., DOG) produces greater activation of features associated with a concept than do nonlabels that refer to the same concept (e.g., hearing a barking noise produced by a dog) in an object-identification task. However, their study does not provide compelling evidence that labels, per se, more strongly activate concepts than nonlabels. It could instead be that the activation of a concept occurs more quickly when participants receive a label (e.g., DOG). To address this, Lupyan and Thompson-Schill (2012, Experiments 1A-1C) varied the interval between the cue's *offset* and the target's onset from 400 to 1000 to 1500 ms and showed that the results were the same for these intervals. However, it could be that the shortest 400 ms interval was not short enough to detect a difference in the rates of activation buildup produced by a label or nonlabel.

A more compelling method for demonstrating that labels can affect the mental representation of an object would involve showing that the presence of a label can improve performance even when the label provides no additional information about the target's location or its identity. Lupyan and Spivey (2010b, Experiment 1) administered a task that required participants to respond as soon as a probe (a dot) appeared next to one of four 2s and four 5s arranged in a circular array. For the Label condition block of trials, prior to each trial, participants were told which of two specific numbers (i.e., a 2 or 5) the dot probe would appear next to and to press a response key as soon as they saw the dot probe. Although the Label condition always informed participants about the number the dot probe would be beside, it did not indicate its exact location because it could appear next to any one of the four instances of that number. In the two No-label condition blocks, prior to each trial participants were told to "attend to the category," which participants were told at the beginning of the block would always be a 2 or would always be a 5. Thus, in the No-label condition the label "category" was just as informative regarding the location of the dot probe as was the 2 or 5 in the Label condition. Nevertheless, RTs were shorter in the Label condition than in the No-label condition. This effect was replicated using pictures of basic-level, common objects (e.g., chairs and tables) as probes, rather than dots (Experiment 6). The authors asserted that their results suggest that specific labels per se facilitate the deployment of spatial attention to the features of objects that are associated with the label.

Vales and Smith (2015) extended Lupyan and Spivey's (2010a) effects. They had 3-year-olds (rather than college undergraduates) complete a conjunctive search task rather than an object search task. In this task, the pictorial target (BED or COUCH) was embedded in an array of 2 to 12 pictorial distractors. All trials began with a picture cue that was exactly the same as the to-be-searched-for target picture. Critically, for half of the participants the picture cue was presented simultaneously with its auditory word label (a pre-recorded stimulus of someone saying BED or COUCH); the other half of the participants received no label. Participants responded by pointing to the target on the computer monitor. As in Lupyan and Spivey (2010a), the label was uninformative because: (a) the same target was repeated throughout the whole experiment, and (b) the picture cue, which was given to all participants on all trials depicted the exact object picture participants had to locate.

Vales and Smith (2015, Experiment 1) reported that the presence of the label reduced RTs. There was no interaction between the effects of the number of distractors and label presence/absence. The linear increase in RTs as the number of distractors increased was of the same magnitude for the Label and No-label groups. In their Experiment 2, to rule out the explanation that their label effect was due to the auditory cue enhancing arousal, which in turn enhanced visual search performance, the researchers included a group in which the label was replaced with the pre-recorded auditory stimulus "go!" This condition yielded results comparable to the No-label group in Experiment 1.

In summary, the results of Lupyan and Spivey (2010a) and Vales and Smith (2015) provide strong evidence that labels enhance performance on visual search tasks. In both studies, the addition of the label can be interpreted as being uninformative because it neither provided information about the target's identity or its exact location. Specifically, in Lupyan and Spivey (2010a) participants always knew the number the dot probe would appear next to because they completed hundreds of consecutive trials in which the dot appeared next to that specific number. In Vales and Smith (2015), the picture cue, which was given in both the Label and No-label condition, always informed participants of the identity of the target. Furthermore, the same target was given to participants throughout the whole session. Nevertheless, these uninformative labels enhanced performance in non-verbal tasks.

3. Self-directed speech enhances performance on a visual search task

The present research focuses on Lupyan and Swingley's (2012) results, which suggest that saying a visual label aloud enhances visual search performance relative to silently reading a label. In three experiments, their participants had to click on a target object picture (or, in the case of Experiment 3, multiple target object pictures) embedded in an array of object-picture distractors, when cued by the target's silently read written label or by the same read-aloud label (hereafter the "Speech" condition). In all experiments, which of these two types of cue participants received for each trial was randomly determined. The targets and distractors were colored line drawings of basic-level, concrete, common object pictures (e.g., BANANA, GIRAFFE, WINDMILL). Distractors were sampled without replacement within trials and with replacement across trials from a pool of 240 different items. In Experiment 1 the number of distractors in the search array was either 17or 35. Overall, search RTs were faster (by about 50 ms) and more accurate in the Speech condition. (Hereafter, this will be called the self-directed-speech effect.) Although RTs increased as the number of distractors increased, there was no Number of Distractors × Type of Cue interaction, suggesting that there was no difference in serial search rates.

Experiments 2 and 3 differed from Experiment 1 in two aspects: (1) the Speech condition required participants to repeatedly say the cue aloud while they searched for the target and more important (2) the targets in these experiments varied on imagery concordance, operationalized as the degree to which the target picture matched the mental image that is activated by the written label for that target. According to Lupyan and Swingley (see also Lupyan & Thompson-Schill, 2012), if an auditory word label activates a more prototypical representation of its referent, then saying the cue aloud should produce a benefit for targets high in imagery concordance, but a null (or reversed) effect for targets low in imagery concordance. A reversed effect is predicted because if saying the cue aloud more strongly activates visual features associated with the prototypical representation of an object in a category, when participants are presented with a picture of an atypical dog, there will be a mismatch in the features of the picture with those activated by the cue, thereby increasing RTs. For example, if saying "DOG" aloud activates a Golden Retriever (a hairy dog with a large body), its features might compete with the processing of distinctly different visual features associated with a target that is an atypical member of the category such as a Chihuahua (a nearly hairless dog with a small body).

In Experiment 2, the self-directed-speech effect only occurred in the last half of the experiment. The failure to observe a self-directed-speech effect in the first few blocks could have been due to participants' adjusting to the demands of the Speech condition (i.e., repeating the target aloud while they try to locate it). In Experiment 3, there was no overall self-directed speech effect for RTs, and participants were slightly *less* accurate in the Speech condition and therefore their results should be interpreted with some caution. However, most important is that in both experiments speech significantly decreased RTs to targets high in imagery concordance but slightly increased RTs to targets low in imagery concordance. According to Lupyan and Swingley (2012), their