Redundant spoken labels facilitate perception of multiple items

GARY LUPYAN
University of Wisconsin, Madison, Wisconsin

AND

MICHAEL J. SPIVEY
University of California, Merced, California

Because of the strong associations between verbal labels and the visual objects that they denote, hearing a word may quickly guide the deployment of visual attention to the named objects. We report six experiments in which we investigated the effect of hearing redundant (noninformative) object labels on the visual processing of multiple objects from the named category. Even though the word cues did not provide additional information to the participants, hearing a label resulted in faster detection of attention probes appearing near the objects denoted by the label. For example, hearing the word *chair* resulted in more effective visual processing of all of the chairs in a scene relative to trials in which the participants attended to the chairs without actually hearing the label. This facilitation was mediated by stimulus typicality. Transformations of the stimuli that disrupted their association with the label while preserving the low-level visual features eliminated the facilitative effect of the labels. In the final experiment, we show that hearing a label improves the accuracy of locating multiple items matching the label, even when eye movements are restricted. We posit that verbal labels dynamically modulate visual processing via top-down feedback—an instance of linguistic labels greasing the wheels of perception.

How does language interact with perceptual processing? According to traditional accounts, language is "merely the formal and expressive medium that is [used] to describe mental representations" (Li & Gleitman, 2002, p. 290). Rather than shaping and modulating perceptual or conceptual representations, language is simply a tool for reporting them (Bloom & Keil, 2001; L. Gleitman & Papafragou, 2005; Pinker, 1994). Under this view, linguistic–perceptual interaction is seen in terms of recoding perceptual experiences into verbal ones (Dessalegn & Landau, 2008; Munnich & Landau, 2003; Paivio, 1986). In the present work, we take an opposing view, arguing that language dynamically modulates visual processing. We focus here on one aspect of this interaction: the degree to which processing spoken labels facilitates the visual processing of the named items. Recent work in cognitive neuroscience has stressed the importance of top-down feedback in perception (Enns & Lleras, 2008; Foxe & Simpson, 2002; C. D. Gilbert & Sigman, 2007; Kveraga, Ghuman, & Bar, 2007; Lamme & Roelfsema, 2000). We argue that in humans, language may be one of the components of such top-down modulation.

The notion that language is tightly integrated with perceptuo-motor activity such as eye movements is in itself uncontroversial. Hearing a word tends to elicit looks to objects associated with the verbal label (e.g., Spivey & Dale, 2005; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) evoking eye movements to the locations of relevant objects even when the scene is no longer visible (Altmann, 2004; Spivey & Geng, 2001). Deictic spatial labels such as *above* and *below* have been shown to guide spatial attention, imposing a conceptual reference frame on the visual scene (Logan, 1995). Spatial verbal labels have also been shown to guide attention even when the labels are nonpredictive: In a task that required participants to detect an X presented above, below, to the left, or to the right of fixation, centrally presented spatial terms (*up*, *down*, *left*, *right*) facilitated detection of the target in the congruent regions even though the cues were not predictive of the location of the target (Hommel, Pratt, Colzato, & Godijn, 2001).

Finding that perception and/or the deployment of attention can be altered by the informational content of language may mean that language actually modulates perceptual processing or that comprehending a spoken or written label alters one's decision criteria. Being told to search for a red vertical line segment, for example, initiates search for a target with these and not other attributes (Wolfe, 1994). Indeed, for simple stimuli like differently colored vertical and horizontal lines, a verbal cue can be almost as effective as a visual preview of the target (Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). Search

---

G. Lupyan, lupyan@wisc.edu

for more complex real-world objects (e.g., a motorcycle), can also be guided by verbal labels, although predictably, category labels like *motorcycle* are never as effective as seeing a pictorial preview of the actual target motorcycle (Vickery, King, & Jiang, 2005). These results, although relevant for understanding the types of cues that can guide the deployment of attention, do not speak directly to the effect of language on visual processing, because language is used simply as a tool to convey information regarding target identity or location.

In fact, linguistic input appears to augment visual processing in ways that extend beyond the communicated content. For instance, a visual search task requiring the search for a feature conjunction (e.g., a red vertical among green vertical and red horizontal line segments) is inefficient even when participants are informed of the target's features ahead of time. However, concurrent spoken delivery of target features can make the inefficient search into an efficient one (Reali, Spivey, Tyler, & Terranova, 2006; Spivey, Tyler, Eberhard, & Tanenhaus, 2001). Hearing the word *red* allows items to compete on the color dimension; the subsequent delivery of *vertical* then allows the already salient items to compete on the orientation dimension, effectively turning a conjunction search into two feature searches.

These types of effects of language on visual processing go beyond types of information delivery. If hearing verbal labels dynamically modulates visual processing, we may observe facilitation in attending to labeled items, even when the labels do not provide additional information. Lupyan (2007b) tested this prediction by looking at whether informationally redundant auditory labels make search more efficient. Participants searched for a 2 among 5s or vice versa (as in Wang, Cavanagh, & Green, 1994). On some trials, prior to the search display, the participants heard a recorded voice name the target identity (e.g., "find the five")—the *label* condition; on other trials, the participants heard a sound clip of equal length and amplitude, in which the target name was omitted—the *no-label* condition. Despite always knowing ahead of time what target they needed to find (the target identity was held constant for hundreds of trials), the participants were faster and more efficient (i.e., RTs had shallower search slopes) when they heard the target labeled than when they searched for a known target without hearing it labeled on that specific trial. The facilitative effect of the labels disappeared when the numerals were rotated 90º, suggesting that it depended on a preexisting association between the visual exemplars and the label (see also Lupyan, 2008a). Importantly, this facilitation was observed in a design that intermixed labeled and unlabeled trials, showing that the effects of labels on visual processing were nonstrategic and transient.

Verbal Labels As Neural Feedback

We hypothesize that words modulate the degree to which visual processing is influenced by the conceptual category of the object or objects being processed. For instance, in labeling a particular chair with the category label CHAIR, the object representation of the particular chair is augmented with the category information of previously encountered chairs (Lupyan, 2008b). This attended and augmented object representation then sends facilitatory feedback to the visual features most associated with the category, while the idiosyncrasies of a particular category exemplar are deemphasized. This process would facilitate the ability to simultaneously attend to all the items of a given category through, for example, a saliency map.

We therefore predicted that hearing category labels would enhance the saliency of the named objects beyond what is possible without hearing the label—especially when participants are required to process/attend to multiple objects simultaneously. This type of modulation hypothesis is broadly supported by what is known about the neural architecture of the primate visual system. Feedback connections from higher to more primary cortical areas are ubiquitous (Mesulam, 1998) and recurrent loops in which higher areas modulate neural activity in lower areas (e.g., prefrontal cortex and IT; MT and V1) and are well documented (Foxe & Simpson, 2002; C. D. Gilbert & Sigman, 2007; Lamme & Roelfsema, 2000). This recurrent activity is not merely epiphenomenal, but is causally implicated in visual awareness (Corthout, Uttl, Walsh, Hallett, & Cowey, 1999; Fahrenfort, Scholte, & Lamme, 2007, 2008). The implication of these findings, which place feedback processing at the forefront of visual perception, is that manipulations that affect feedback activity can influence perception.

Labels and Categories

Verbal labels denote categories. As Goldstein stated, "When we speak of 'table' we do not mean a special given table with all the accidental properties, but we mean 'table' in general. We employ the word 'table' in this categorical sense when naming a particular table" (cited in Noppeney & Wallesch, 2000, p. 376). An important goal of the present work was to investigate the function that labels serve in processing multiple objects from the same category rather than specific object instances. The study of the role of categories in visual processing has had a rich history (see Duncan, 1983; H. Gleitman & Jonides, 1976; Jonides & Gleitman, 1976; White, 1977). Drawing on these classic studies, contemporary investigations of category effects in perception have often relied on the visual search paradigm. Wolfe, Friedman-Hill, Stewart, and O'Connell (1992) found that when searching for oriented lines among differently oriented distractors, search proceeds as though there were categories of VERTICAL, HORIZONTAL, STEEP, SHALLOW, RIGHT, and LEFT. Search among targets and distractors that are within these categories is slower than that between them, even though the differences in stimulus parameters remain constant. Daoutis, Pilling, and Davies (2006) obtained similar results using color categories with perceptual distances controlled. The guided search model (Wolfe, 1994) elegantly accounts for these findings through top-down effects of channels that have peak activations at the centers of these categories (e.g., the vertical channel responds maximally to vertical lines). It is at present unclear where these channels come from and how they come to encode such information as steepness

and leftness. Although it is possible that such fundamental perceptual dimensions as verticality arise from ecological distributions, other hypothesized channels seem more arbitrary (e.g., steep, shallow, and purple channels). Might categorical effects be produced in part by linguistically derived categories? That is, might the ability to use the label *steep* facilitate selection of the steep elements (perhaps by enabling more efficient perceptual grouping among elements that comprise the STEEP and SHALLOW categories)? Supporting this possibility is a study showing that learning arbitrary associations between tilted lines and arbitrary labels can produce such categorical effects. For instance, learning to associate 45º and 90º lines with the label *pencil* and −45º and 0º lines with the label *elephant* resulted in faster search when targets and nontargets spanned the category boundary than when they were both in the same category (Smilek, Dixon, & Merikle, 2006).

Further evidence for the possible involvement of language in perceptual categories comes from the finding that categorical perception of color (i.e., greater perceptual sensitivity for items that span a category boundary than for equally spaced items within a category) is lateralized to the left hemisphere (A. L. Gilbert, Regier, Kay, & Ivry, 2006) and is disrupted when participants are placed under conditions of verbal interference (Pilling, Wiggett, Ozgen, & Davies, 2003; Roberson & Davidoff, 2000; Wiggett & Davies, 2008; Winawer et al., 2007). The involvement of language in the categorical perception of color is also supported by findings that categorical perception of color is disrupted by verbal interference. These effects are not unique to color categories and have been extended to facial expressions (Roberson & Davidoff, 2000) and familiar objects (A. L. Gilbert, Regier, Kay, & Ivry, 2008). In summary, there is suggestive evidence that category effects in visual processing may be modulated by verbal labels. We do not wish to claim that visual categorical perception is not possible without language. However, given that verbal labels typically denote categories (e.g., *green* denotes a range of colors) and given that top-down feedback is important for even the most basic visual processes, it may be the case that verbal labels can dynamically modulate visual processing, enabling improved processing of stimuli from the named category.

Predictions

In the present experiments, we test several specific hypotheses. First, we test the prediction that hearing a category name (e.g., FIVE or CHAIR) enhances the saliency of multiple exemplars of the named category, facilitating deployment of attention to those items. Second, we test the prediction that effects of labels on attention are modulated by stimulus typicality. Because category labels are most strongly associated with typical exemplars (Rosch, 1973), the effects of labels on attention should be strongest for typical items. Third, we hypothesize that the effect of labels may interact with the spatial grouping of the visual stimuli. The facilitatory effect of hearing a label may be stronger when the task requires more focused attention, such as when exemplars from different categories are spatially interspersed rather than spatially grouped. Fourth, if the effect of labels on the deployment of attention is dynamic, as we claim, it should unfold gradually in time. We investigate the time course of the label effect by parametrically varying the delay between the appearance of the visual stimuli and the appearance of the attentional probe (see Figure 1 and the Method section of Experiment 1 below). Fifth, if hearing a label enhances the saliency of multiple objects (a form of category-based attention), its effects should be observed even when the display of the objects to be attended is very brief, preventing eye movements and restricting multiple covert shifts of attention.

Fixation	Cue	Cue-to-stimuli delay	Stimuli-to-probe delay	Detect probe
500 msec	1,300 msec	500 msec	1,500 msec	Response or 4,000 msec

Label: "attend to two [five]"
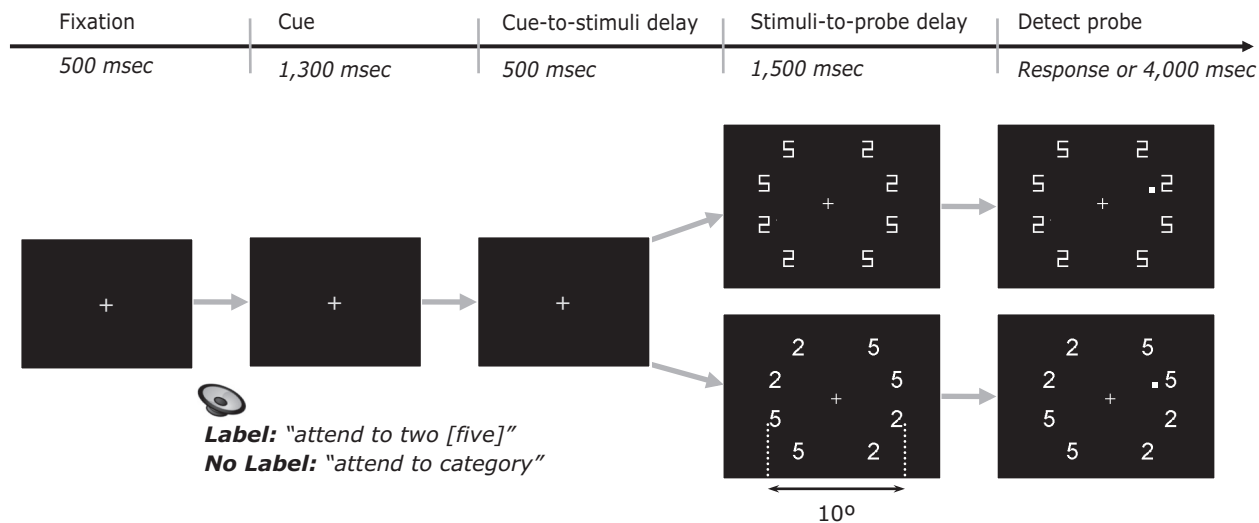No Label: "attend to category"

10º

Figure 1. The design of Experiment 1. A random half of the trials contained numerals rendered in a typical (Arial) font (bottom panel), and the remaining half contained the less typical digital-font numerals (top panel). A random half of the trials were cued with category labels.

EXPERIMENT 1

To test the hypotheses listed above, we used a variation of the classic probe-detection technique (e.g., Egly, Driver, & Rafal, 1994; Eriksen & Hoffman, 1972; Posner, Snyder, & Davidson, 1980). In this paradigm, participants report the appearance of a small visual stimulus (probe) by pressing a button as soon as they detect it. A classic finding is that manipulations that evoke covert shifts of attention from a central fixation point to a particular location—for instance, viewing a cue in the form of an arrow that points to a given location—yield shorter response times (RTs) when a probe subsequently appears at that location. More generally, manipulations that increase the salience of particular stimuli should improve the speed with which a probe near those stimuli is detected.

In the modified version of the probe-detection paradigm used here, the cues consisted of spoken category labels (the words *two* and *five* in Experiments 1, 2, 5, and 6; the words *chair* and *table* in Experiment 4). The spoken label cues were manipulated as a within-subjects factor, occurring on a random half of the trials. Following the cue, participants viewed a display of numbers (2s and 5s, Experiments 1, 2, 5, and 6) or pictures of chairs and tables (Experiment 4) and had to respond to the appearance of a probe appearing next to one of the numerals or pictures. Although the exact position of the probe was not known ahead of time, it was constrained to always appear next to a stimulus from a given category (e.g., 2s) for one half of the experiment and next to the other category (e.g., 5s) for the other half (cf. Experiments 3A and 3B). The participants were explicitly told of this constraint at the start of the experiment.

The critical test involved comparing RTs to the probe on trials during which the participants heard the category labeled (e.g., "attend to the five") with RTs when the category name was replaced by a filler word (e.g., "attend to the category"). Insofar as labels enhance the saliency of objects from the labeled category, the participants should be faster to detect the probe when the target category is labeled. That is, hearing the label should enhance the saliency of stimuli matching the named category. We hypothesized that this facilitation would be transient, and so we should observe facilitated responses on label relative to no-label trials, even when the two trial types are intermixed. Because the same category was probed for hundreds of trials, the label was redundant; it did not tell the participants anything that they do not already know. With only two categories, memory demands were minimal; the participants did not need to be reminded which was the relevant category.

Method

Participants. Eighteen Carnegie Mellon University undergraduates volunteered in exchange for course credit or $7.

Materials. The stimuli were the digits 2 and 5 presented in either standard Arial font (2, 5) or a more atypical digital font (5, 2) in which the two digits are mirror images of each other. Numeric characters were used as stimuli because they are perceptually simple and, being overlearned, have strong category representations. As is true of most categories, these stimuli can be classified at mul-

tiple levels of abstraction: A 2 can be a number, an even number, a two, and so on (Posner & Mitchell, 1967). Classifying a 2 as TWO can be thought of as a basic-level task in the sense that people are generally faster at identifying alphanumeric characters at this level than at identifying more superordinate levels (Dick, 1971; Posner, 1970), probably because a 2 is more frequently classified as a member of the category TWO than as a member of the category NUMBER.

On a given trial, all stimuli were in the same font. The characters were white on a black background and had a visual angle of $0.7° \times 0.8°$. The numerals were arranged along the circumference of an imaginary circle having a diameter of 10° around a fixation cross (0.5° diameter), making all stimuli equidistant from fixation. Each trial contained eight digits: four 2s and four 5s. On each trial, there was always an equal number of 2s and 5s on the left, right, top, and bottom parts of the display. The attention probe consisted of a small white dot (3×3 pixels) that appeared 3.85° from fixation in the direction of one of the numerals (Figure 1). Each stimulus display contained eight numerals arranged in a circle. The numerals were positioned at 18°, 54°, 126°, 162°, 198°, 234°, 306°, and 342°. No stimuli were positioned directly above or below the fixation cross in order to more easily check for possible visual field effects. The stimuli were displayed on a 17-in. CRT monitor; stimulus delivery was controlled by Presentation software (www.neurobs.com). Responses were collected using a standard USB mouse.

To equalize auditory exposure, both the label and the no-label conditions contained auditory cues. We recorded a female native English speaker producing the phrase "attend to the" and the words "two," "five," and "category." For the label trials, the carrier phrase "attend to the" was spliced with "two" or "five." For the no-label trials, the phrase was spliced with the neutral word "category." The resulting auditory cues for the label and no-label trials thus differed only in whether the relevant numeral category was labeled. Although the labels *five* and *two* technically referred to multiple 5s and 2s, respectively, the words "fives" and "twos" seemed unnatural because of the rare occurrence of plural markers on number words. Hence, the singular words "two" and "five" were used.

Procedure. The participants were instructed that they should fixate on the central cross throughout the experiment and click the mouse button as soon as they detected the probe. They were warned that on some trials (catch trials), no probe would be present, in which case, they should just wait for the next trial to start. Catch trials were equally distributed across trial types. For half of the experiment, the participants were instructed to attend to all of the 5s in the display (T = 5 trials), for the remaining half, to all the 2s (T = 2 trials). The order was counterbalanced among participants.

Each trial began with a fixation cross (500 msec) followed by a 1.30-sec audio cue. For the label trials, the audio cue was "attend to the two" for T = 2 trials or "attend to the five" for T = 5 trials. For the no-label trials, the cue was "attend to the category." The numerals appeared 500 msec after the label offset and remained on the screen until the response or trial timeout (4 sec). After 1,500-msec, a small probe appeared next to one of the 2s for the T = 2 trials or 5s for the T = 5 trials. A buzz sounded if the participants responded before the probe appeared or during a catch trial. The labeling condition was a within-subjects factor; the label and no-label trials were randomly intermixed.

There were two kinds of display types: alternating, in which the 2s and 5s alternated (2, 5, 2, 5), and clumped, in which they were clumped in groups of two (2, 2, 5, 5, 2, 2). For half of the trials, the stimuli appeared in a familiar Arial font (typical condition; Figure 1 bottom), whereas on the remaining trials, the numerals appeared in a digital font (atypical condition; Figure 1 top). Both the font type and the display type were randomly intermixed.

The rationale for manipulating the spatial grouping of the stimuli was twofold. First, it increased the uncertainty of the spatial distribution of the digits from trial to trial. Second, it allowed us to examine possible interactions between spatial grouping and conceptual grouping—the finding that visual similarity is affected by conceptual similarity such that objects from the same conceptual category become

more similar to each other, producing a type of grouping, (Lupyan, 2008a; Lupyan, Thompson-Schill, & Swingley, 2010). The prediction was that the categories (2s vs. 5s) would be especially salient when the items to which they referred were spatially proximate to each other (clumped condition). However, it is not totally clear whether this increased salience should lead to a larger or smaller facilitatory effect of labels. The label effect may be larger, because object features highlighted by the label would attract attention more reliably when they belonged to objects physically proximate to one another. Alternatively, increasing the salience of the categories through spatial grouping may lead to better performance overall, masking any facilitation due to labels. In the alternating display, the two categories (2s vs. 5s) were spatially intermixed and thus required more focused attention to separate. Hearing a label may be especially important in this condition, but its efficacy may be limited to the typical font numerals, because they possess more category-typical features (i.e., the visual features activated by hearing the label *two* are more congruent with the typical rather than atypical font stimuli.

Each block contained 71 trials (eight valid probe positions × two display types × two levels of typicality × label vs. no-label + 7 catch trials on which the probe was absent). The participants completed five blocks of 71 trials per target category—a total of 710 trials. Because a number of participants complained about the length of the task, we reduced the length to four blocks per target (568 trials) for the last 8 participants. The task took approximately 50 min to complete.

Results

Misses (1.6%) and false alarms (2.6%) were excluded from the RT analyses. Analyses of these infrequent errors failed to show any evidence of condition effects or a speed–accuracy trade-off—unsurprising for a probe-detection task—hence our discussion will be focused on RT analyses. The RT distribution was highly skewed (Kolmogorov–Smirnoff test: $KS = .201$, $p < .01$). We therefore used RT medians as the dependent measure in this and all subsequent experiments.

Several repeated measures ANOVAs were performed. The first included three within-subjects factors: target category (2 vs. 5), font (Arial vs. digital), and presence of labels (label vs. no label). No target category effects were found in any of the experiments ($Fs < 1$). The participants were quicker to detect the probe on the Arial-font typical trials ($M = 346$ msec, $SD = 74$) than on the digital-font atypical trials ($M = 372$ msec, $SD = 91$) [$F(1,17) = 28.39$, $p < .0005$] (Figure 2, top panel). The participants were faster to detect the probe on label trials than on no-label trials [$F(1,17) = 13.35$, $p = .002$]. This facilitation was not large, but highly reliable: 15 of the 18 participants showed the effect (sign test: $p = .008$). As is evident in Figure 2, top panel, the facilitation was highly reliable for the typical-font trials [$F(1,17) = 6.73$, $p = .019$]. Probe detection for the atypical trials was not significantly facilitated by hearing the label [$F(1,13) = 1.13$, $p = .302$]. However, the typicality × label presence interaction was not significant [$F(1,17) = 1.05$, $p = .323$]. The label facilitation persisted for the duration of the experiment; we did not find any interaction between experimental block and degree of label facilitation. An additional analysis including visual field (probe on the left vs. probe on the right) as a factor failed to show any significant visual field effects.

To further explore possible mediators of a relationship between typicality and label presence, we ran a second repeated measures ANOVA, which included font (i.e., typicality), label condition, and display type (alternating vs. clumped) as within-subjects factors. Main effects of typicality and label presence remained highly significant ($F = 19.61$ and 9.82, respectively). Contrary to our expectation, the main effect of display type did not approach significance [$F(1,17) < 1$], suggesting that the total number of attentional foci required (four for the alternating display, two for the clumped display) did not influence overall performance. However, there was a significant interaction among typicality, label presence, and display type [$F(1,17) = 6.65$, $p = .02$]. This three-way interaction is unpacked in the bottom panels of Figure 2. Regardless of how the numerals were arranged, the participants were faster to respond to probes on the typical-font trials, and their RTs were decreased overall by hearing the name of the numeric category. When numerals were clumped into groups of two (e.g., 2, 2, 5, 5, 2, 2), the facilitation produced by the labels was equivalent for the typical- and atypical-font conditions (Figure 2, bottom left panel). However, when the numerals were not grouped in pairs (alternating display type; Figure 2, bottom right panel), the facilitation produced by the labels was limited to the numerals rendered in a typical font. Planned t tests on the atypical and typical trials for the alternating display condition (Figure 2, bottom right panel) revealed that the labeling difference was not significant in the atypical-font condition [$t(16) = 0.13$, n.s.] but was robust in the typical-font condition [$t(16) = 3.17$, $p = .005$]. The two-way interaction between typicality and label presence was marginal [$F(1,17) = 3.55$, $p = .077$].

Discussion

Spoken category labels facilitated the deployment of attention to multiple instances of the named category, as was revealed by faster responses to probes appearing next to one of the exemplars from the named category. The observed facilitation was relative to trials in which the participants knew the relevant category (which remained constant for several hundred trials) but did not actually hear the spoken label. Moreover, the facilitation was observed when the label and no-label trials were intermixed, making unlikely an account of the labeling effect based on strategic factors. The fact that labels facilitated probe detection when label and no-label trials were intermixed also indicates that the effect of labels is transient. If it were not, the RT differences between the two trial types would have quickly dissipated. The temporal dynamics of the label effect is explored further in Experiment 5.

In addition to the effect of labels on the deployment of attention, we found a highly reliable difference in RTs between typical (Arial-font) and atypical (digital-font) stimuli. This finding is not obvious a priori: identifying the digits is ancillary to the primary task of detecting the probe; the present task of simple detection of a probe can, in principle, be accomplished without identifying or categorizing the numerical digits at all.

Our conclusion that spatial attention was in fact driven by category-level information rests on two sources of evi-
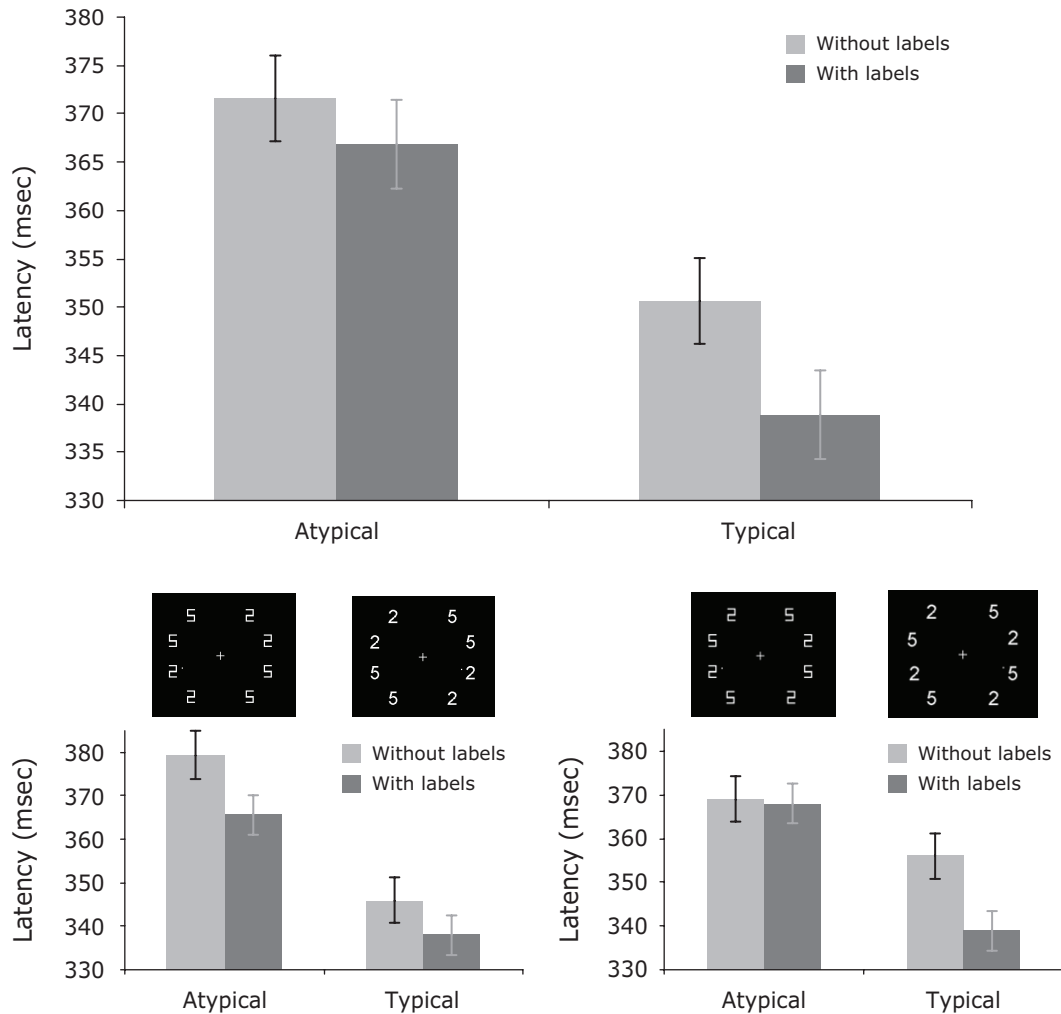
Figure 2. Results of Experiment 1. The top panel shows overall median reaction times graphed separately for the atypical (left) and typical (right) trials. The bottom panels show the same data separated by display type and typicality. Error bars represent 1 standard error of within-subjects condition differences.

dence. First, the difference in RTs between typical and atypical trials is expected if the participants were able, as they were instructed, to allocate attention to the 2s or the 5s in the display and could achieve this more effectively if the visual exemplars were highly overlearned Arial-font stimuli than if they were the less familiar and more confusable digital-font stimuli. Second, unless spoken labels facilitate probe detection irrespective of the stimuli used in the display—a possibility that is tested in Experiment 2—the finding that hearing a category label facilitates the detection of a probe appearing next to one of the named objects is only possible if the participants processed the numeric characters at a conceptual level to some degree.

As was predicted, the effect of labels was numerically greater for the typical stimuli; however, the two-way interaction between typicality and label presence was not significant. Further analyses revealed a reliable three-way interaction among typicality, label presence, and display type. One interpretation of this interaction is that the spa-

tially grouped (clumped) numerals from the same category provided a better match to the category label and that this effect is conditioned by typicality. Atypical stimuli were made more salient by the label (resulting in decreased probe-detection RTs), but only when the categorical status of the atypical stimuli was further enhanced by spatial grouping. That is, the mechanism that performs spatial grouping may not be independent of the mechanism that performs conceptual grouping. There may be partial overlap between the process that allows a spoken label to highlight atypical objects that are conceptually similar and the process that highlights visually similar objects that are positioned near one another.

EXPERIMENT 2

To further test the hypothesis that the labeling effect observed in Experiment 1 is specific to stimuli that are strongly associated with the label, we conducted a replication of Experiment 1, except that the Arial-font 2s and 5s

were mirror reversed (i.e., rotated about the *y*-axis). We reasoned that this manipulation would weaken the mapping between stimulus and category/category label, thus making these stimuli analogous to the atypical digital font. The effect of the label on the deployment of attention was thus predicted to be weakened or eliminated (see Lupyan, 2008a, for a similar manipulation).

A secondary goal of Experiment 2 was to examine whether the shorter RTs in the Arial-font (typical) trials observed in Experiment 1 were simply due to low-level perceptual differences between the typical- and atypical-font trials or whether the RT differences were indicative of visual processing being driven by the stimulus category. In the former view, performance was slower on the digital-font trials because the digital-font numerals were less physically discriminable than the Arial-font stimuli. Alternatively, the overall longer RTs to the digital-font stimuli may have been due to their status as less typical category exemplars, which slowed the categorization process. On the latter view, the significantly faster detection of the probe next to the visually typical numerals was due to their being more quickly categorized.

Note that although the digital-font 2s and 5s are indeed more difficult to discriminate from each other than 2s and 5s rendered in a standard font, it is not clear, a priori, that discriminability should be at all relevant in this task. In fact, the main effect of font type on performance, whether or not it is due simply to discriminability differences, provides evidence that visual saliency is manipulated to some degree by conceptual categories.

To disentangle the effects of low-level discriminability differences from category effects, we replaced the typical-font numerals with their mirror-reversed versions. This had the effect of maintaining all the low-level features of the Arial-font stimuli, while weakening the association between the visual form of the exemplar and the category or category label (e.g., it is slower to name or classify a mirrored numeral than its canonical form). If the effect of font on probe detection had a basis in low-level physical differences (e.g., physical discriminability), the mirror reversal should have no effect on the demonstrated advantage of detecting probes next to Arial-font stimuli. Conversely, if the robust RT difference between typical- and atypical-font trials demonstrated in Experiment 1 was related to typicality (the association strength between the visual exemplars and the category), this mirror reversal should eliminate or lessen the advantage demonstrated for the Arial-font stimuli.

The mirror reversal also allowed us to test the specificity of the labeling advantage found in Experiment 1. Insofar as labels facilitate the processing of associated stimuli, manipulations that disrupt the association between the label and the visual exemplar should eliminate or lessen the benefit of labels (see also Lupyan, 2007b, 2008a).

Method

Participants. Fourteen Carnegie Mellon undergraduate students were recruited and compensated $7 for their participation.

Materials and Procedure. The materials and procedure were identical to those in Experiment 1, with the exception that the typical (Arial-font) stimuli were mirrored (rotated about the *y*-axis).

Results

Misses (1.0%) and false alarms (3.0%) did not differ between conditions and were excluded from the RT analyses. The RT data are shown in Figure 3. Median RTs were numerically faster in this experiment than those in Experiment 1 (322 vs. 355 msec), but this difference was not reliable [$F(1,30) = 1.61, p > .2$]. Responses were still faster to the now-rotated typical stimuli than to the atypical stimuli [$F(1,13) = 13.42, p = .003$], although the difference was significantly smaller than that in Experiment 1. The Arial-font advantage was reduced from 25.1 msec in Experiment 1 to 9.7 msec in the present experiment. This difference was significant by a two-sample, two-tailed *t* test [$t(30) = 2.67, p = .014$].

There was no overall effect of labels [$F(1,13) = 1.60, p > .2$] and no interaction between label presence and font [$F(1,13) = 1.10, p > .3$]. Planned comparisons examining the effect of labels on the atypical (digital-font) and mirrored (Arial-font) trials revealed that, as in Experiment 1, labels did not significantly affect RTs for the atypical-font trials [$F(1,13) = 3.17, p = .100$] and also had no effect on the now mirror-reversed Arial-font trials [$F(1,13) < 1$], even though the mirrored numerals had the same low-level perceptual properties of the original numerals. Also as in Experiment 1, there was no interaction between experiment block and the presence of labels ($F < 1$).

A direct comparison of Experiments 1 and 2 was conducted by entering typicality and labeling condition as within-subjects factors and experiment as a between-subjects factor. We found a significant interaction between label-presence and experiment [$F(1,30) = 8.89, p = .004$], indicating a significantly smaller effect of labels in Experiment 2 than in Experiment 1. The three-way interaction among typicality, label presence, and experiment did not reach significance [$F(1,30) = 1.65, p > .1$]; however, as can be observed by comparing Figures 2 and 3, the difference in the labeling effect between the two experiments was limited to the Arial-font trials: Mirroring them rendered the labels ineffective. A direct comparison of the effect of labels on the Arial-font trials in Experiment 1 with that on the mirrored Arial-font trials in Experiment 2 yielded a marginally significant experiment × label presence interaction [$F(1,30) = 4.04, p = .055$].
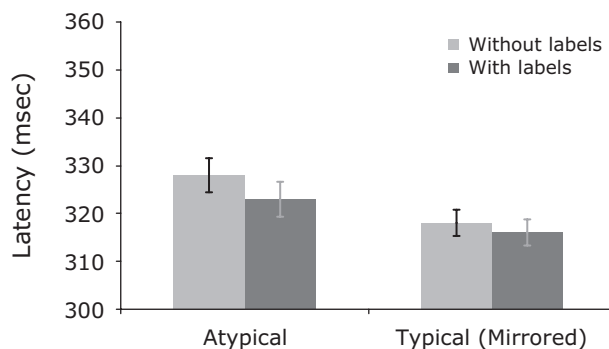

Figure 3. Results of Experiment 2. Mirror-reversing the typical stimuli eliminated the facilitating effect of labels. Error bars represent 1 standard error of within-subjects condition differences.

Discussion

The finding that a simple physical transformation (a mirror reversal) significantly reduced the RT difference between the typical- and atypical-font trials and that mirroring eliminated the effect of hearing category labels on visual processing has two immediate implications. First, the highly reliable RT difference in detecting a probe appearing next a typical versus atypical category exemplar cannot be reduced to a simple difference in the physical discriminability between the two categories. When low-level differences between the exemplars are preserved but the association between the typical exemplars and their category representations (or prototypes) is disrupted, the effect of typicality on overall RTs is significantly reduced. Second, the present results confirm that the degree to which labels facilitate the deployment of attention to multiple instances of the named category appears to depend on stimulus typicality (operationalized by the association strength between the visual exemplar and the label), rather than on any feature-based visual differences (Lupyan, 2008a).

Although the overall RT difference between the fonts was reduced in the present experiment, responses were still significantly faster on the now-mirrored Arial-font trials than on the digital-font trials. This suggests that stimulus discriminability may indeed be a factor in producing the RT difference between the Arial and digital fonts. Notably, the digital-font stimuli did not differ from each other on any simple visual features, whereas the Arial-font stimuli, although made less familiar by a mirror reversal, were still discriminable on the basis of simple visual features (e.g., the presence of a vertical line segment in the 5s but not in the 2s). Nevertheless, the finding that the font effect was reduced by more than a factor of two between Experiments 1 and 2 suggests that the primary origin of the RT difference lies in category typicality.

EXPERIMENT 3A

We have claimed that labels facilitate the deployment of attention even when they are entirely redundant. In Experiment 1, we described the spoken labels as redundant, because they did not communicate any information not contained in the block-wide instructions that the participants were given at the start of the task informing them of the target category (which remained constant). However, previous work has shown that trial-by-trial cues are more effective than block-wide cues (Posner et al., 1980). Thus, it may be that the participants did not make use of the block-wide cues, which would make the labels quite informative, rather than redundant. Experiments 3A and 3B were designed to test whether participants can make use of block-wide category cues (with no spoken labels delivered within the block). If under these circumstances, trials that are congruent with the cue (valid trials) elicit shorter RTs than invalidly cued trials, we can conclude that the participants are using their block-wide knowledge of the cue even on trials that do not contain a spoken label reminder and, through inference, that the spoken labels in Experiment 1 were indeed redundant, as we have claimed.

Method

Participants. Ten University of Pennsylvania undergraduate students completed the experiment in exchange for course credit.

Materials and Procedure. The stimuli were identical to those in Experiment 1, except that only the typical (Arial-font) numbers were used. As in Experiment 1, the participants were instructed to attend to all of the 5s (T = 5 trials) in the display for one half of the experiment and to all of the 2s in the remaining half (T = 2 trials). The order was counterbalanced among participants. Because the goal of the present experiment was to test the effectiveness of a block-wide cue, the cues consisted of simply instructing people to attend to the 2s (T = 2) or 5s (T = 5) only at the start of the experiment. The trials were divided into validly cued trials (72%), invalidly cued trials (18%), and catch trials (10%). That is, of the noncatch trials, 80% were valid and 20% invalid. A valid trial was defined as one on which the cue appeared next to one of the 2s in a T = 2 block and next to one of the 5s in a T = 5 block. If instructing the participants to attend to stimuli from a particular category is effective, we should find shorter RTs on the valid than on the invalid trials.

Results and Discussion

Misses (0.15%) and false alarms (2.5%) did not differ between conditions and were excluded from the RT analyses. A repeated measures ANOVA, with target category (2 vs. 5) and validity as fixed factors and subject as a random effect, revealed significantly shorter RTs on the validly cued trials ($M = 382$ msec, $SD = 85$) than on invalidly cued trials ($M = 395$ msec, $SD = 95$) [$F(1,9) = 6.82$, $p = .028$]. There were no other reliable effects ($Fs < 1$).

We found a significant, albeit small, RT advantage for validly cued relative to invalidly cued trials. This difference indicates that the participants were using their knowledge of what the cue was during the entire block of trials, even when there were no spoken label reminders. Consequently, we conclude that the spoken label cues that accompanied some individual trials in Experiment 1 and that imbued an exceptional facilitation of RTs were indeed redundant cues.

It is perhaps noteworthy that the median RTs for the valid trials in the present study were, on average, 44 msec greater than those for the typical no-label trials of Experiment 1 (Figure 2). This difference did not reach significance as tested by a cross-experiment t test, but the longer RTs observed in the present experiment support previous evidence that block-wide cues are less effective than cues that are delivered on a trial-by-trial basis (Posner et al., 1980).

EXPERIMENT 3B

In Experiment 3A, we found that block-wide category cues facilitated responses on valid relative to invalid trials. However, in Experiment 3A, we used stimuli that were readily discriminable (the typical-font numerals), raising the possibility that participants attend more to the relevant digits in this experiment than in Experiments 1 and 2, which included trials on which the numerals were less discriminable. If true, it remains possible that block-wide cues would not be effective if discrimination between 2s and 5s on some trials was made more difficult, as it was in Experiment 1. Experiment 3B was a replication of Experiment 3A, but the typical and atypical (digital-font) numerals were intermixed. A finding that valid cues still

sped responses, particularly for the typical-font stimuli, would strengthen the claim that the labels used in Experiment 1 were truly redundant, affecting attention over and above block-wide cues.

Method

Participants. Eleven University of Pennsylvania undergraduate students completed the experiment in exchange for course credit.

Materials and Procedure. The stimuli were identical to those in Experiment 1. The procedure was identical to that in Experiment 3A, except that both typical and atypical digits were used and that to keep the length of the experiment comparable to that of Experiment 3A, each participant was instructed to attend to only a single target category. The T = 2 and T = 5 trials were counterbalanced among participants.

Results and Discussion

Misses (0.2%) and false alarms (5.0%) did not differ reliably between target categories (T = 2 vs. T = 5); hits were marginally higher for valid than for invalid trials (valid, $M = 99.4\%$; invalid $M = 99.9\%$) [$F(1,10) = 4.06$, $p = .072$]. Error trials were excluded from the RT analysis. Median RTs were analyzed using a repeated measures ANOVA. Initial analyses failed to show any effects of the target category ($F < 1$), and so this factor was dropped from further analyses. The factors of interest were validity (valid vs. invalid) and typicality (typical font vs. digital font). There were no main effects of validity ($F < 1$), or typicality [$F(1,10) = 1.04$, $p > .3$], but there was a reliable validity × typicality interaction [$F(1,10) = 7.89, p = .02$]. Planned comparisons showed that the participants responded more quickly on valid than on invalid typical trials (valid, $M = 331$ msec; invalid, $M = 340$ msec) [$t(10) = 2.66$, $p = .024$]. There was no reliable effect of validity on atypical (digital-font) trials. In fact, for these trials, valid cues slightly slowed down responses (valid $M = 336$ msec; invalid, $M = 344$ msec) [$t(10) = 1.70$, $p = .12$].

These results replicate and extend the finding of Experiment 3A. Block-wide category cues (i.e., the instruction to attend to just the 2s or just the 5s) at the start of a block produced faster responses on valid than on invalid trials, but only for the typical-font stimuli. This result supports our claim that verbal labels modulate visual processing of labeled items over and above block-wide cues.

EXPERIMENT 4

Although the stimuli used in Experiments 1–3B arguably comprise basic-level categories in Rosch's sense (Mervis & Rosch, 1981), it is important to test the generality of the labeling effect with more ecologically valid categories. It is also important to demonstrate that the label-facilitation effect is present when each trial includes visually heterogeneous exemplars from the named category.

Method

Participants. Eighteen Carnegie Mellon University undergraduates volunteered in exchange for course credit or $7.

Materials and Procedure. The procedure was identical to that in Experiment 1, with the following exceptions. The numerals 2 and 5 were replaced with images of chairs and tables. The images were obtained from the IKEA online catalog, converted to grayscale, and calibrated to have equivalent brightness and contrast in order to minimize effects of low-level perceptual differences in drawing attention to individual objects. Each trial contained 4 chairs and 4 tables, selected at random with replacement from a larger set of 20 stimuli. The images were positioned on the screen in a manner identical to that in Experiment 1. The participants were told in one part, to attend to the chairs and, in the second part, to attend to the tables. Order was counterbalanced. The participants completed five blocks for each target category. Within each block, half of the trials contained a label cue (e.g., "attend to the chairs [tables]") and half continued a no-label cue ("attend to the category"). The two trial types were intermixed. The present stimuli had more heterogeneous contours than the 2s and 5s, and some stimuli tended to mask the probe dot in their contours, making detection difficult. For this reason, the probe consisted of a thin gray frame appearing around the target stimulus.

Results

Accuracy was very high (misses, $M = 2\%$; false alarms, $M = 2\%$) and did not differ between conditions, so our analyses were focused on median RTs. An initial repeated measures analysis using label presence as a within-subjects factor showed a significant main effect of label presence [$F(1,17) = 11.96$, $p = .003$]. Exploratory analyses revealed that the effects of labels were strongest at the start of each half of the experiment and wore off after several blocks. We therefore added a second factor to the ANOVA, coding Blocks 1 and 2 as the early blocks and Blocks 3–5 as the late blocks. The main effect of label presence remained highly significant [$F(1,18) = 8.95$, $p = .008$]. Within each object category, there was no overall change in RTs between early and late blocks [$F(1,17) < 1$]. As is shown in Figure 4, labels facilitated responses only for the early blocks as is shown by a significant label presence × block (early vs. late) interaction [$F(1,17) = 5.16$, $p = .036$]. Having found this interaction, we conducted an identical analysis on the data from Experiment 1. This analysis failed to show an interaction between the magnitude of the labeling effect and the time of the experiment (early vs. late blocks) [$F(1,17) < 1$] and no significant effect of experiment block on overall RTs [$F(1,17) = 1.41, p > .2$].

Discussion

Experiment 4 replicated the main result of Experiment 1: Labels facilitated the deployment of attention to multiple exemplars associated with the category label. Experiment 4 extended this result in two ways. First, we showed that the result generalizes to a much richer category structure: realistic pictures of chairs and tables, which are difficult to distinguish as a group using basic perceptual features (Wolfe & Horowitz, 2004). Although it may be argued that the numerical stimuli 2 and 5 can be discriminated on the basis of simple visual features (e.g., the presence of a vertical line segment in a 5 but not a 2 rendered in an Arial font), such a strategy was not possible in the present experiment, because the chair and table pictures did not differ as a category on any simple visual features.

Second, we showed that spoken labels facilitate the deployment of attention to multiple different exemplars of a given category. In contrast to Experiment 1, in which
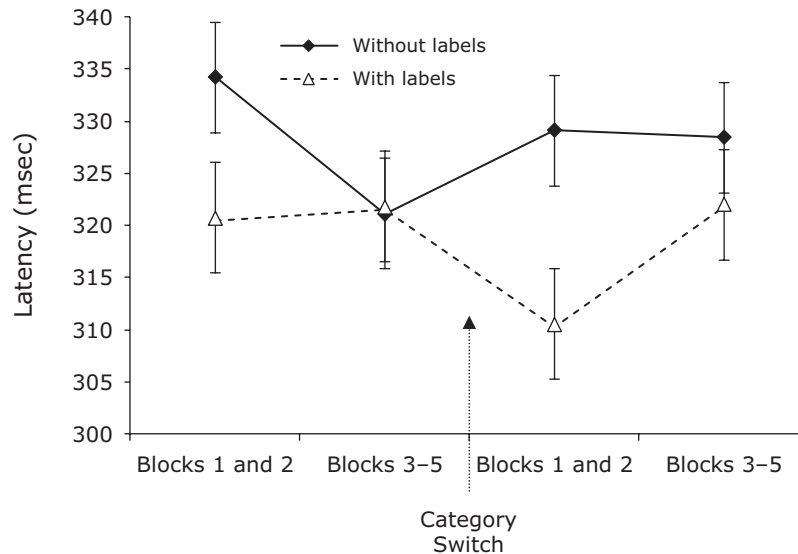
Figure 4. Results of Experiment 4. The facilitatory effect of labels, generalized to a richer stimulus set (chairs and tables). The facilitation of RTs due to labels for this stimulus set was only present for the first several blocks of each target category, reappearing when the category was switched. Error bars represent 1 standard error of within-subjects condition differences.

stimuli from the same category were identical to one another on a given trial, in the present experiment, multiple different exemplars were shown on a single trial. The finding that RTs on the label trials were still reduced relative to those on the no-label trials provides additional evidence that hearing a spoken label facilitates the visual processing of multiple different stimuli that are associated with the label.

Unlike in Experiment 1, in which labels facilitated performance for the entire duration of the task, in the present study, labels only facilitated performance for the first several blocks of each target category. Following the category switch (from chair to table or table to chair) the labels once again facilitated performance, with the effect dissipating over time. The disappearance of the facilitation may be due to semantic satiation (a phenomenon in which repetition of a word causes increased difficulty in processing it; e.g., Smith & Klein, 1990), which may be stronger for these more complex picture stimuli than for numeric characters. The rather complex pattern of results evident in Figure 4 is likely a product of semantic satiation, general practice effects, and switching the target category (i.e., switching from attending to chairs to attending to tables). Further work is necessary to tease apart the contributions of these factors.

EXPERIMENT 5

The proposed account of the effects of labels on the deployment of attention posits that it is transient. To further explore the temporal aspects of the effect of labels on attention, we manipulated the duration of the stimulus-to-probe delay. We predicted that the facilitatory effect of the label should unfold in time, as would be expected

if the effect is produced through recurrent processes that involve top-down feedback. Knowing the time frame at which the labeling effect is at its peak will be useful for guiding future experimentation.

Method

Participants. Twenty-five Cornell University undergraduate students completed the experiment in exchange for course credit.

Materials and Procedure. Experiment 5 was identical to Experiment 1, with the following exceptions: First, only typical (Arial-font) stimuli were used. Second, the stimulus-to-probe delay—the interval between the appearance of the numeral display, and the appearance of the probe—was not fixed to 1,500 msec, but ranged from 750 to 2,000 msec in increments of 250 msec. There was an equal number of trials at each delay duration. The choice of delays was informed by a pilot experiment in which we tested shorter delays of 350 and 500 msec. Labels did not affect RTs for these short delays.

Results

Accuracy was very high (misses, $M = 1\%$; false alarms, $M = 2\%$) and did not vary as a function of the stimulus-to-probe delay ($F < 1$). The subsequent analyses are focused on median RTs. A repeated measures ANOVA with probe-onset delay and label presence as within-subjects factors revealed significant main effects of the probe-onset delay [$F(5,120) = 14.33$, $p < .0005$]. As before, labels facilitated the deployment of attention [$F(1,24) = 7.89$, $p = .010$] (Figure 5). The label × delay interaction was not significant ($F < 1$). Planned comparisons of the effect of labels at each probe-onset delay revealed that labels facilitated responses only for the intermediate probe onsets at 1,250 msec [$t(24) = 1.87$, $p = .037$] and 1,500 msec (the delay used in Experiment 1) [$t(24) = 2.08$, $p = .024$, pairwise t tests, one-tailed]. As in Experiment 1, there was no interaction between experiment block and the presence of labels ($F < 1$).

Discussion

Our goal in Experiment 5 was to investigate the time course of the label facilitation effect reported in Experiment 1. The present experiment yielded two main findings: First, overall RTs decreased rapidly from a delay of 750 to 1,250 msec (Figure 5). Second, the facilitatory effect of labels was apparent only for the two intermediate delays—1,250 and 1,500 msec.

The finding that overall RTs decreased with greater stimuli-to-probe onset delays supports the notion that prior to the appearance of the probe, the participants are (explicitly or implicitly) categorizing the stimuli. The decrease in overall RTs with increasing probe-onset delays may arise from this categorization process, which allows for the deployment of attention to the targets and away from the distractors. Indeed, this decrease may reflect the effect of the block-wide cues.

The finding that the effect of labels is transient is expected given the results of Experiment 1. As was mentioned earlier, if hearing a label produced a long-term effect in this task (e.g., if the labels produced a type of repetition priming), the difference in RTs between the label and no-label trials would rapidly disappear, since the benefit of hearing a label on trial n would carry over to trial $n+1$. The present experiment provides direct support for the transient nature of the labeling effect observed in Experiment 1 and reveals the temporal range in which the labeling effect is most effective.

The failure to find a facilitation of labels at short probe-onset delays was not due to insufficient time to process the cue (a 750-msec probe-onset delay translates to a 1,250-msec interval between label offset and the probe, providing ample time for processing the verbal cue; Gibson & Kingstone, 2006; Vickery et al., 2005; Wolfe et al.,

2004). The failure to find an effect of labels at this delay suggests that facilitation due to labels requires that the bottom-up activation from the stimulus display interact with the top-down activity induced by the label (see Experiment 6 for an elaboration). The reported time course is no doubt dependent on the choice and number of stimuli, suggesting a number of follow-up experiments investigating the contribution of factors such as typicality, stimulus number, visual complexity, and spatial proximity on the time course of the labeling effect.

Before concluding that redundant verbal labels indeed facilitate the deployment of attention to multiple objects that match the label, we thought it would be prudent to attempt to generalize the effect to a somewhat different experimental paradigm, one that allows us to examine whether the effect of labels can be observed through an accuracy measure, as well as RT, to determine whether the effect of labels remains when eye movements are restricted, and to elucidate the nature of the rather long (>1 sec) delay that seems to be necessary for labels to affect visual processing.

EXPERIMENT 6

We have interpreted the finding that verbal labels decrease RTs to probes appearing next to cued numerals as evidence that hearing a label—even one that is informationally redundant—facilitates the perception of all the stimuli that match the label. This is perhaps accomplished by the label's preactivating the associated visual representations, enabling more efficient deployment of attention to the matching stimuli. In effect, hearing the word *five* temporarily makes the visual system a better detector of 5s (see Iordanescu, Guzman-Martinez, Grabowecky, & Su-
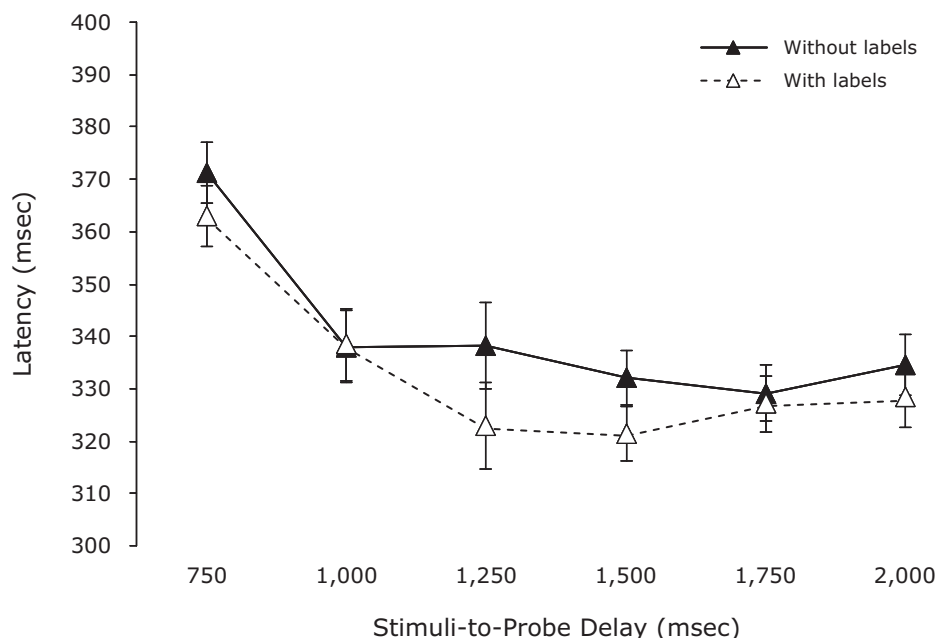

Figure 5. Results of Experiment 5. Manipulating the delay between the onset of the numerical stimuli and the onset of the probe revealed that the facilitatory effect of labels was transient.

zuki, 2008, for similar reasoning). In this experiment, we aimed to test this account in a new way. Instead of being instructed to detect a probe near an item from a labeled category, the participants heard a verbal cue followed by a briefly presented display. Their task was to click on the locations of all of the items (now off-screen) that matched the category relevant to the task (the number 2 for one half of the participants and the number 5 for the remaining half). Our primary dependent measure was the number of correctly identified locations. As before, the labels, presented randomly on a trial-by-trial basis, were informationally redundant. The target category stayed constant throughout the experiment.

A secondary aim of Experiment 6 was to provide better control over eye movements. The brief stimulus presentations ensured that at least for the shortest display duration, the participants did not have sufficient time to move their eyes while the target items were in view. Finally, the experiment elucidates the nature of the time course of the effect of labels on attentional deployment. Experiment 5 showed that labels facilitated probe detection only when the probe appeared 1,250–1,500 msec after the onset of the target/distractor stimuli (the numerals 2 and 5). One possibility is that this much time is required for the top-down cues (labels) to interact with the bottom-up information (the eight digits surrounding the central fixation point; see Di Lollo, Kawahara, Zuvic, & Visser, 2001, for a relevant discussion). Although there is no baseline against which to compare these >1 sec delays, these times are longer than the time it typically takes to generate an object template from a text label (Vickery et al., 2005; Wolfe et al., 2004) or to evoke a spatial shift following a word like *above* (Hommel et al., 2001). A possibility investigated in this experiment is that verbal labels can facilitate processing of matching stimuli even when the latter are very briefly presented, as long as the participants have information regarding the positions of potential targets (Burkell & Pylyshyn, 1997). To this end, in Experiment 6, we used empty placeholders to mark (precue) the locations of potential targets and limited exposure to the target and distractor stimuli to 100–500 msec.

Method

Participants. Twenty-two University of Pennsylvania undergraduate students completed the experiment in exchange for course credit. The participants were randomly assigned to the 2 or the 5 condition.

Materials and Procedure. The materials were identical to those of Experiment 1, except only the typical (Arial-font) stimuli were used. To accommodate the demands of the present task, the imaginary circle along which the stimuli were arranged was reduced to 7.4°, and the stimuli now appeared inside of placeholder rectangles outlined in white. Rather than detect an attention probe, the participants' task was to click on all of the locations in which a stimulus from a specific category appeared (Figure 6). Thus, during the presentation of the stimuli, the participants had to deploy attention to all of the stimuli matching the target category.

Each trial began with a fixation cross. After 500 msec, it was surrounded by placeholders (Figure 6). After another 500-msec delay, an auditory label was heard, identical to that used in Experiment 1. One second after label offset, the placeholders were filled with randomly arranged 2s and 5s (four of each). The digits were visible for one of three durations (100, 300, or 500 msec), chosen pseudorandomly for each trial. All trial types had equal frequencies. After the digits disappeared, a mouse cursor became visible at the fixation point, and the participants clicked on the locations (indicated by the empty placeholders) at which they remembered seeing a particular digit (2 or 5, depending on the condition). The next trial was initiated automatically when the participants clicked on four unique locations. No feedback was provided. Chance level was 50% (two out of four correct).

The participants completed as many trials as possible in a single 45-min session (range = 250–405 trials, $M = 355$ trials). The design ensured that each position was equally likely to be occupied by a 2 or a 5 and that correlation between label presence and stimulus-duration was exactly or very close to 0.

Results

Preliminary analyses revealed that performance was significantly worse during the first 50 trials; these were therefore considered practice and were not analyzed further. There were no significant differences in performance between the 2 and 5 conditions [$F(1,20) < 1$], and the conditions did not interact with display duration [$F(2,40) = 1.75$, $p = .19$]. All subsequent analyses were collapsed across the target category.

As is shown in Figures 7A and 7B, performance was remarkably good. Of the almost 7,000 total trials, there

Fixation	Cue	Cue-to-stimuli delay	Stimulus display	Click on four target positions
500 msec	1,300 msec	1,000 msec	100–500 msec	Until response

Label: "attend to two [five]"
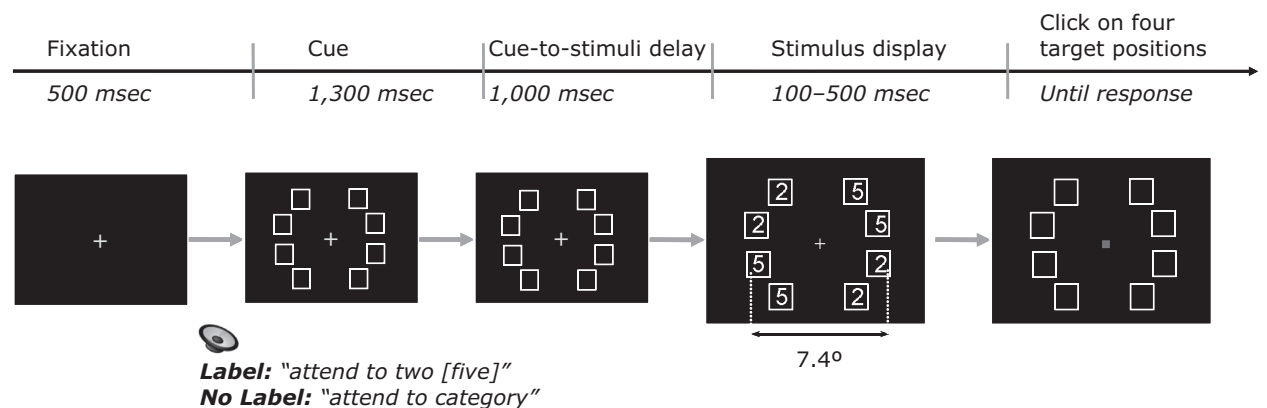No Label: "attend to category"

7.4°

Figure 6. The design of Experiment 6. A random half of the trials was cued with category labels. The trials were split evenly into stimulus durations of 100, 300, and 500 msec.

were a total of 2 on which a participant clicked on none of the correct locations. Even when the eight digits were shown for 100 msec—too short to permit any eye movements—overall performance was 77.4% (3.096/4), well above the chance value of 50%. To test whether labels affected accuracy in locating the target stimuli, we conducted a repeated measures ANOVA with label presence and duration as within-subjects factors. Because of greater variability in the shortest (100-msec) duration condition, the ANOVA was weighted so as not to violate the equal variances assumption. Unsurprisingly, performance improved dramatically for longer exposure duration, as measured by accuracy [$F(2,42) = 241.55$] (Figure 7C) and total RT [$F(2,42) = 49.22$] (Figure 7D). Critically, relative to the intermixed no-label trials, responses during the label trials were significantly faster [$F(1,21) = 4.46, p = .047$] and more accurate [$F(1,21) = 7.04, p = .015$] (Figures 7A–7D). These effects remained significant with an unweighted ANOVA. The label presence × stimulus duration interaction was not significant for either RTs or accuracy, but as can be observed in Figure 7C, the effect of labels on accuracy was primarily driven by the

shorter stimulus durations, with a likely ceiling effect at the 500-msec duration for accuracy.[1]

Discussion

The results of Experiment 6 extended the results of Experiments 1, 2, 4, and 5 to an experimental paradigm requiring participants to explicitly note the location of all of the stimuli matching the target category. It was found that hearing an informationally redundant label (presented on a random 50% of the trials) improved the participants' ability to correctly identify the locations of the stimuli from the target category. We also observed shorter RTs on the label trials than on the no-label trials. In the context of this experiment, RTs reflect the time it took the participants to move the mouse to the four locations in which they thought the target stimuli were presented. It is reasonable to interpret shorter RTs as resulting from greater confidence in the location of the targets.

Although the participants in Experiments 1–5 were instructed not to move their eyes, the design of these experiments provided ample opportunities for eye movements, making it unclear whether the facilitatory effects of hear-
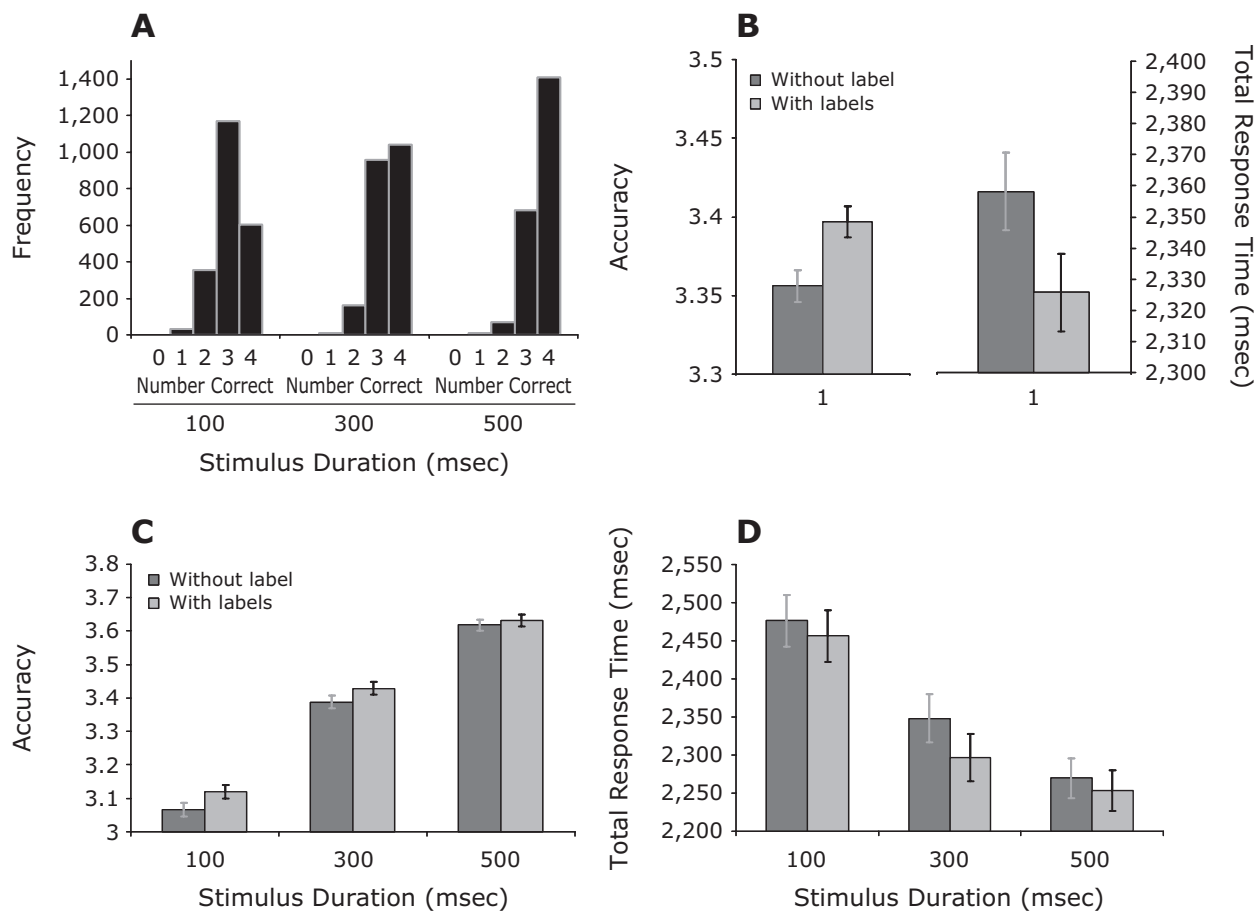

Figure 7. Results of Experiment 6. (A) The distribution of responses as a function of stimulus duration. (B) Overall effects of labels on accuracy (left, out of four) and response times (right). (C) The effect of labels on accuracy (out of four) as a function of stimulus duration. (D) The effect of labels on response time as a function of stimulus duration. Error bars represent 1 standard error for within-subjects condition differences.

ing labels are observable in the absence of eye movements. In the present study, the short display durations limited eye movements. We cannot be sure that eye movements did not contribute to the results of the earlier studies, but the finding of facilitated performance on the trials on which the stimuli were displayed for only 100 msec means that the facilitatory effects of redundant labels can occur in the absence of eye movements. The finding of facilitated performance on the label trials (both in accuracy and RT) suggests that labels can affect the processing of multiple stimuli throughout a visual scene in parallel (i.e., when the participants have only a very limited opportunity or no opportunity to shift their attention from item to item after hearing the label). As we discuss below, we believe that this is achieved through a preactivation of visual representations associated with the verbal label.

In Experiments 1, 2, 4, and 5, there was a long delay after the offset of the label, during which the participants could examine the visual scene. In the present study, this delay was replaced by one with empty placeholders, which the participants could use to select the regions of the display that might contain target stimuli (Burkell & Pylyshyn, 1997). Although the category representations evoked by the verbal labels may have some degree of spatial invariance (Lupyan & Spivey, 2010), the placeholders explicitly demarcated the possible positions. In this context, hearing the word two may activate visual representations of category-typical 2s in the positions corresponding to the empty placeholders. When the 2s actually appear in a subset of those positions, there is a match between the bottom-up, stimulus-driven representation and the top-down templates set up by the label and a mismatch (a prediction error) in the position occupied by the nontargets (5s in this example). This account is fully compatible with predictive vision frameworks (Enns & Lleras, 2008; Kveraga et al., 2007; Rao & Ballard, 1999).

GENERAL DISCUSSION

Spoken category labels facilitated the deployment of attention to multiple category exemplars, as revealed by shorter RTs to a visual probe appearing near a stimulus belonging to the named category (Experiments 1, 2, 4, and 5). The facilitation was observed for numerals (2s and 5s) and for more complex objects (chairs and tables). In the former case, the facilitation lasted for the duration of the experiment. For more complex objects, the facilitation dissipated over time and was restored after a category switch (e.g., switching from attending to chairs to attending to tables). More research is required to fully understand the cause of this dissipation.

Hearing category labels improved performance, even though the labels were informationally redundant. The relevant category was kept constant for hundreds of trials and was known to the participants ahead of time. Nevertheless, actually hearing the category label facilitated probe detection. Experiment 6 extended this result to a paradigm requiring the participants to actually note the locations of all of the relevant stimuli (which occupied four out of eight possible spatial positions). Hearing an information-

ally redundant label improved the participants' accuracy in identifying the targets—that is, facilitated deployment of attention to stimuli matching the label—when the targets/distractors were presented for as briefly as 100 msec.

The claim that the labels were informationally redundant (a claim further strengthened by the results of Experiment 3) is critical. A number of previous studies have established that verbal object labels help to set up templates that can guide visual search (Vickery et al., 2005; Wolfe et al., 2004). However, in these studies, the labels informed the participants of the upcoming target. In contrast, in the present experiments, the participants always knew what the relevant category was, both because they were explicitly told what it was and because it remained constant for hundreds of trials. The finding of a transiently facilitated performance following the presentation of an object label (a within-subjects, mixed-trial manipulation) suggests that actually hearing a label enables the visual system to process stimuli more effectively than can be accomplished without the label (Lupyan, 2007b, 2008a). In effect, language is "greasing the wheels of perception" (R. Goldstone, personal communication, July 25, 2008).

The facilitatory effect of labels was modulated by the typicality of the exemplars, becoming weaker when the digits were rendered in an atypical font (Experiment 1) and disappearing when the stimuli were mirror reversed (Experiment 2)—a manipulation that was designed to keep constant the low-level features of the stimuli but disrupt the association between the visual stimuli and the label.

These results are unexpected under a conception of verbal labels (and language more broadly) as simple outputs of a system designed to translate concepts to their equivalent linguistic symbols (Dessalegn & Landau, 2008; Li & Gleitman, 2002). On such a view, the concept of which object to attend should have been equivalently activated in all conditions of these experiments, but the presence of the redundant linguistic cue was found to enhance performance.

What can explain this effect of language on vision? We theorize that the effect is a type of top-down modulation of ongoing activity in the visual system. Contemporary accounts of the primate visual system stress the importance of feedback—the modulation of earlier areas by later areas (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; C. D. Gilbert & Sigman, 2007; Lamme & Roelfsema, 2000; Mesulam, 1998; Mumford, 1992). The consequence of massive feedback processes is that the clear *anatomical* hierarchy of the visual system (Felleman & Van Essen, 1991) does not result in a clear *functional* hierarchy of its subprocesses. For instance, the very late prefrontal areas of cortex can at times respond to the presence of a visual stimulus before early visual cortex (V2; see Lamme & Roelfsema, 2000, for a review). Visual processing can be modulated by top-down feedback extremely rapidly. The presence of fast-conducting magnocellular pathways between V1 and MT enables the latter to modulate activity in V1 via feedback within 10 msec (Hupé et al., 2001; Vidyasagar, 1999). Vidyasagar and Hupé et al. speculated that the extremely rapid conduc-

tion velocities of the V1↔MT pathway mean that some signals from MT can be transmitted back to V1–V3 in as little as 1–2 msec. One effect of this feedback activity is to dynamically reshape receptive fields. For instance, in V1, cells are retuned from reflecting simple orientation features within classically small and simple receptive fields to reflecting figure–ground relationships over a much larger area about 100 msec after stimulus onset (Lamme, Rodriguez-Rodriguez, & Spekreijse, 1999; Olshausen, Anderson, & Van Essen, 1993).

Effects of verbal labels on vision can be seen as embodying a more complex type of perceptual modulation (which correspondingly takes more time to achieve). In this view (the label feedback hypothesis of Lupyan, 2007a), processing a category label initiates a volley of feedback activity to object-selective regions of the cortex, such as IT (Logothetis & Sheinberg, 1996), producing a predictive signal or head start to the visual system (Esterman & Yantis, 2008; Kveraga et al., 2007; Puri & Wojciulik, 2008). In the biased competition theory of attention (Desimone & Duncan, 1995), these predictive signals would enable those neurons with receptive fields lying within the named object to gain a competitive advantage (Deco & Lee, 2002; Kramer, Weber, & Watson, 1997; Kravitz & Behrmann, 2008; Vecera & Farah, 1994). Given feedback from object-selective cortical regions, winning objects can bias earlier spatial regions of visual cortex. Iordanescu et al. (2008) invoked a similar mechanism to explain why presenting sounds characteristic of target objects, such as the jingling of keys or the meowing of a cat, facilitates reporting the location of the associated objects in a visual search task even though the sound cues do not themselves provide any spatial information.

The observed effects of labels on attentional deployment occur sooner than eye movements can be generated (Experiment 6), are manifested in a within-subjects design on a trial-by-trial basis (all of the present experiments) and comprise facilitation over and above that provided by block-wide knowledge of the cue (e.g., Experiment 3). Therefore, we are inclined to characterize this finding of redundant spoken labels influencing visual perception as a relatively automatic top-down influence of linguistic–conceptual representations on visual representations (Lupyan, 2007a, 2007b, 2008a; Lupyan et al., 2010). The present results do not allow us to make conclusions about the neural locus of the labeling effect. At present, it remains unclear whether labels directly affect visual representations that comprise an attentional saliency map or whether they affect an object-based attentional process that in turn influences a visual saliency map. In either case, the present findings comprise an effect of verbal labels on the ability to attend to multiple objects matching the label.

We interpret the present effects as being automatic, rather than strategic. Our use of the rather ambiguous term *automatic* is similar to the definition of automaticity in the context of the Stroop effect. On being presented with a word, participants tend to automatically process its meaning. However, this type of automaticity does not imply that the dependent measure cannot be influenced by various experimental manipulations just as the degree of interference in the Stroop paradigm can be influenced by numerous factors (MacLeod, 1991). Nevertheless, because in the present studies the labels predicted the target category, we cannot conclude that labels automatically direct attention regardless of their predictive power (in the same way that a flashing light captures attention even when it is counterpredictive of the target location; e.g., Posner et al., 1980). At present, the finding that attentional deployment to the named items is affected by redundant verbal labels requires that the participants attend to the verbal cue (cf. Hommel et al., 2001, who found that verbal effects on spatial attention can be observed even when the verbal cues are unpredictive or counterpredictive).

We predicted that, insofar as verbal labels are associated with object categories rather than with single objects, verbal cues would affect not just the processing of single objects, but would have scene-wide effects on multiple objects that match the label. This claim is at odds with a single-spotlight model of attention. Despite the continued prevalence of the single-spotlight metaphor, there is strong evidence that people can attend simultaneously to multiple noncontiguous locations or objects (Malinowski, Fuchs, & Müller, 2007; McMains & Somers, 2004; M. M. Müller & Hübner, 2002; Scholl & Pylyshyn, 1999), even for sustained periods of time (McMains & Somers, 2005; M. M. Müller & Hübner, 2002; M. M. Müller, Malinowski, Gruber, & Hillyard, 2003; N. G. Müller & Kleinschmidt, 2003). The results of Experiment 6, in which brief presentations restricted eye movements, provide additional support to the claim that the participants were able to rapidly process multiple items from a common category—a form of category-based attention. This ability was facilitated by the participants' actually hearing a verbal label immediately prior to the appearance of the visual scene.

The present studies may be construed as a type of cross-modal facilitation. Although online effects of vision on audition are well known (Ma, Zhou, Ross, Foxe, & Parra, 2009; McGurk & MacDonald, 1976), less is known about the degree to which auditory information affects online visual processing and visual attention. There is now accumulating evidence that sounds affect visual perception (Sekuler, Sekuler, & Lau, 1997; Shams, Kamitani, & Shimojo, 2002), with modulations of early visual cortex by sounds detected in as little as 35–65 msec (Shams, Iwaki, Chawla, & Bhattacharya, 2005). The finding that spoken labels facilitated the deployment of visual attention can be viewed as an instance of such cross-modal facilitation (Molholm, Ritter, Javitt, & Foxe, 2004), albeit one in which verbal labels do not merely facilitate the recognition of a single object, but facilitate the visual processing of multiple objects from the named category in parallel.

We did not compare spoken cues with other types of cues, and therefore, the present results do not speak to the question of whether these cuing effects are special to auditory labels. The facilitatory effect of spoken labels observed in the present study may well generalize to written words and may indeed even be observed when nonspeech

sounds strongly associated with the target category (e.g., "meow" for CAT) are used (Iordanescu et al., 2008).[2] Such a finding would not detract from our conclusion that language modulates visual processing. Learning a language involves, among other things, learning a mostly arbitrary mapping between sounds and classes of objects. When this association is learned, labels can provide top-down input to the visual system, modulating its activity. One thing that makes verbal labels special is their ubiquity: The association between the word *chair* and visual instances of chairs is present in the mind of every speaker of English (although its strength is perhaps conditioned by individual experience).

In a very real sense, hearing the word *chair* temporarily enabled the listener to more effectively process and locate images of chairs. By virtue of the learned associations between words and their referents, words appear to shape the perceptual representations that underlie our conceptual knowledge. Rather than simply a tool for reporting experiences (Bloom & Keil, 2001; L. Gleitman & Papafragou, 2005; Pinker, 1994) and more than providing a separate verbal code (Dessalegn & Landau, 2008; Munnich & Landau, 2003; Paivio, 1986), language augments ongoing perceptual processing.

AUTHOR NOTE

REFERENCES

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm." *Cognition*, **93**, B79-B87.

Bloom, P., & Keil, F. (2001). Thinking through language. *Mind & Language*, **16**, 351-367.

Burkell, J. A., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision*, **11**, 225-258.

Corthout, E., Uttl, B., Walsh, V., Hallett, M., & Cowey, A. (1999). Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *NeuroReport*, **10**, 2631-2634.

Daoutis, C., Pilling, M., & Davies, I. (2006). Categorical effects in visual search for color. *Visual Cognition*, **14**, 217-240.

Deco, G., & Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing*, **44**, 775-781.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, **10**, 204-211. doi:10.1016/j.tics.2006.03.007

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193-222.

Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science*, **19**, 189-195.

Dick, A. (1971). Processing time for naming and categorization of letters and numbers. *Perception & Psychophysics*, **9**(3B), 350-352.

Di Lollo, V., Kawahara, J., Zuvic, S. M., & Visser, T. A. (2001). The preattentive emperor has no clothes: A dynamic redressing. *Journal of Experimental Psychology: General*, **130**, 479-492.

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. *Perception & Psychophysics*, **34**, 221-232.

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.

Enns, J. T., & Lleras, A. (2008). What's next? New evidence for prediction in human vision. *Trends in Cognitive Sciences*, **12**, 327-333. doi:10.1016/j.tics.2008.06.001

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characterisics of selective encoding from visual displays. *Perception & Psychophysics*, **12**, 201-204.

Esterman, M., & Yantis, S. (2008). Category expectation modulates object-selective cortical activity [Abstract]. *Journal of Vision*, **8**(6), 555a.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, **19**, 1488-1497.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision*, **8**(1, Art. 12), 1-12. doi:10.1167/8.1.12

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, **1**, 1-47.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. *Experimental Brain Research*, **142**, 139-150.

Gibson, B. S., & Kingstone, A. (2006). Visual attention and the semantics of space: Beyond central and peripheral cues. *Psychological Science*, **17**, 622-627.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, **103**, 489-494.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorfian effect beyond the realm of color discrimination. *Brain & Language*, **105**, 91-98.

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, **54**, 677-696.

Gleitman, H., & Jonides, J. (1976). Cost of categorization in visual search: Incomplete processing of targets and field items. *Perception & Psychophysics*, **20**, 281-288.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633-661). Cambridge: Cambridge University Press.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, **12**, 360-365.

Hupé, J.-M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology*, **85**, 134-145.

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, **15**, 548-554. doi:10.3758/PBR.15.3.548

Jonides, J., & Gleitman, H. (1976). The benefit of categorization in visual search: Target location without identification. *Perception & Psychophysics*, **20**, 289-298.

Kramer, A. F., Weber, T. A., & Watson, S. E. (1997). Object-based attentional selection—Grouped arrays or spatially invariant representations? Comment on Vecera and Farah (1994). *Journal of Experimental Psychology: General*, **126**, 3-13.

Kravitz, D. J., & Behrmann, M. (2008). The space of an object: Object attention alters the spatial gradient in the surround. *Journal of Experimental Psychology: Human Perception & Performance*, **34**, 298-309.

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

Lamme, V. A., Rodriguez-Rodriguez, V., & Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex*, **9**, 406-413.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision

offered by feedforward and recurrent processing. *Trends in Neurosciences*, **23**, 571-579.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, **83**, 265-294.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, **28**, 103-174.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, **19**, 577-621.

Lupyan, G. (2007a). *The label feedback hypothesis: Linguistic influences on visual processing*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Lupyan, G. (2007b). Reuniting categories, language, and perception. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 1247-1252). Austin, TX: Cognitive Science Society.

Lupyan, G. (2008a). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, **108**, 566-577.

Lupyan, G. (2008b). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, **137**, 348-369.

Lupyan, G., & Spivey, M. (2010). Making the invisible visible: Auditory cues facilitate visual object detection. *PLoS ONE*, **5**(7), e11452. doi:10.137/journal.pone.0011452

Lupyan, G., & Thompson-Schill, S. (2010). The evocative power of words: Activation of visual information by verbal and nonverbal means. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 883-888). Austin, TX: Cognitive Science Society.

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, **21**, 682-691.

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: A Bayesian explanation using high-dimensional feature space. *PLoS ONE*, **4**, e4638. doi:10.1371/journal.pone.0004638

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**, 163-203.

Malinowski, P., Fuchs, S., & Müller, M. M. (2007). Sustained division of spatial attention to multiple locations within one hemifield. *Neuroscience Letters*, **414**, 65-70.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron*, **42**, 677-686.

McMains, S. A., & Somers, D. C. (2005). Processing efficiency of divided spatial attention mechanisms in human visual cortex. *Journal of Neuroscience*, **25**, 9444-9448.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, **32**, 89-115.

Mesulam, M.-M. (1998). From sensation to cognition. *Brain*, **121**, 1013-1052.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, **14**, 452-465. doi:10.1093/cercor/bhh007

Müller, M. M., & Hübner, R. (2002). Can the spotlight of attention be shaped like a doughnut? Evidence from steady-state visual evoked potentials. *Psychological Science*, **13**, 119-124.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, **424**, 309-312.

Müller, N. G., & Kleinschmidt, A. (2003). Dynamic interaction of object- and space-based attention in retinotopic visual areas. *Journal of Neuroscience*, **23**, 9812-9816.

Mumford, D. (1992). On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biological Cybernetics*, **66**, 241-251.

Munnich, E., & Landau, B. (2003). The effects of spatial language on spatial representation: Setting some boundaries. In D. Gentiner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 113-155). Cambridge, MA: MIT Press.

Noppeney, U., & Wallesch, C. W. (2000). Language and cognition—Kurt Goldstein's theory of semantics. *Brain & Cognition*, **44**, 367-386.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, **13**, 4700-4719.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Pilling, M., Wiggett, A., Ozgen, E., & Davies, I. R. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition*, **31**, 538-551.

Pinker, S. (1994). *The language instinct*. New York: HarperCollins.

Posner, M. I. (1970). On the relationship between letter names and superordinate categories. *Quarterly Journal of Experimental Psychology*, **22**, 279-287.

Posner, M. I., & Mitchell, R. (1967). Chronometric analysis of classification. *Psychological Review*, **74**, 392-409.

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

Puri, A. M., & Wojciulik, E. (2008). Expectation both helps and hinders object perception. *Vision Research*, **48**, 589-597.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience*, **2**, 79-87.

Reali, F., Spivey, M. J., Tyler, M. J., & Terranova, J. (2006). Inefficient conjunction search made efficient by concurrent spoken delivery of target identity. *Perception & Psychophysics*, **68**, 959-974.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, **28**, 977-986.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, **38**, 259-290.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: An MEG study. *Neuroscience Letters*, **378**, 76-81.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, **14**, 147-152.

Smilek, D., Dixon, M. J., & Merikle, P. M. (2006). Revisiting the category effect: The influence of meaning and search strategy on the efficiency of visual search. *Brain Research*, **1080**, 73-90.

Smith, L., & Klein, R. (1990). Evidence for semantic satiation: Repeating a category slows subsequent semantic processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 852-861.

Spivey, M. J., & Dale, R. (2005). On the continuity of mind: Toward a dynamical account of cognition. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 45, pp. 85-142). San Diego: Elsevier.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, **65**, 235-241.

Spivey, M. J., Tyler, M. J., 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.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, **123**, 146-160.

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, **5**, 81-92. doi:10:1167/5.1.8

Vidyasagar, T. R. (1999). A neuronal model of attentional spotlight: Parietal guiding the temporal. *Brain Research Reviews*, **30**, 66-76.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception & Psychophysics*, **56**, 495-500.

White, M. J. (1977). Identification and categorization in visual search. *Memory & Cognition*, **5**, 648-657.

Wiggett, A. J., & Davies, I. R. L. (2008). The effect of Stroop interference on the categorical perception of color. *Memory & Cognition*, **36**, 231-239.

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*, **104**, 7780-7785.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 34-49.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, **5**, 495-501.

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.

NOTES

1. As in Experiment 1, there was no reliable interaction between time on task (divided into three bins) and label presence ($F \leq 1$). We observed a marginal three-way interaction among subject, time on task, and label presence [ANCOVA: $F(21,42) = 1.77$, $p = .058$], suggesting that for some but not other participants, the labels facilitated performance throughout the experiment. This analysis is complicated by the fact that the faster participants completed more trials than the slower participants, making this experiment poorly suited for answering questions regarding the long-term time course of the labeling effect.

2. Current work in our lab (Lupyan & Thompson-Schill, 2010) shows that verbal labels (e.g., *cat*) facilitate the identification of pictures of the named category and facilitate the discrimination of typical from atypical instances (e.g., an upright cat vs. an upside-down cat) more strongly than hearing an equally familiar nonverbal cue (e.g., "meow").