Behavioral/Cognitive

Words Jump-Start Vision: A Label Advantage in Object Recognition

Bastien Boutonnet[1] **and Gary Lupyan**[2]

[1]Leiden Institute for Brain and Cognition, University of Leiden, NL-2300 RA Leiden, The Netherlands, and [2]Department of Psychology, University of Wisconsin-Madison, Madison, Wisconsin 53706

People use language to shape each other's behavior in highly flexible ways. Effects of language are often assumed to be "high-level" in that, whereas language clearly influences reasoning, decision making, and memory, it does not influence low-level visual processes. Here, we test the prediction that words are able to provide top-down guidance at the very earliest stages of visual processing by acting as powerful categorical cues. We investigated whether visual processing of images of familiar animals and artifacts was enhanced after hearing their name (e.g., "dog") compared with hearing an equally familiar and unambiguous nonverbal sound (e.g., a dog bark) in 14 English monolingual speakers. Because the relationship between words and their referents is categorical, we expected words to deploy more effective categorical templates, allowing for more rapid visual recognition. By recording EEGs, we were able to determine whether this label advantage stemmed from changes to early visual processing or later semantic decision processes. The results showed that hearing a word affected early visual processes and that this modulation was specific to the named category. An analysis of ERPs showed that the P1 was larger when people were cued by labels compared with equally informative nonverbal cues—an enhancement occurring within 100 ms of image onset, which also predicted behavioral responses occurring almost 500 ms later. Hearing labels modulated the P1 such that it distinguished between target and nontarget images, showing that words rapidly guide early visual processing.

Key words: **categorization; concepts; even-related potentials; language and thought; perception; representations**

Introduction

People use language to shape each other's behavior in highly flexible ways. Effects of language are often assumed to be "high-level" in that, whereas language clearly influences reasoning, decision making, and memory, it does not influence low-level visual processes. Here, we test the prediction that words are able to provide top-down guidance at the very earliest stages of visual processing by acting as powerful categorical cues.

The knowledge of what something looks like can be activated in a variety of ways. For example, after learning that dogs bark, hearing a bark can activate the corresponding visual knowledge as attested by facilitated visual recognition and discrimination of cued categories (Lupyan and Thompson-Schill, 2012; Edmiston and Lupyan, 2013) and activation of category-specific representations in visual cortex (Vetter et al., 2014). Another way in which visual knowledge can be activated is through words. However, unlike other perceptual cues, words are categorical and unmotivated—any utterance of the word "dog" can be used to refer to any dog. This renders words uniquely suited for activating mental states corresponding to categories.

Although no one doubts the power of language to verbally instruct to expect one thing or another and to change how one evaluates and acts on perceptual inputs, many believe that such verbal cuing of knowledge leaves perceptual processing itself unaffected (Gleitman and Papafragou, 2005; Klemfuss et al., 2012; Firestone and Scholl, 2014). Words, in this view, are "pointers" to high-level conceptual representations (Bloom, 2000; Jackendoff, 2002; Dessalegn and Landau, 2008; Li et al., 2009). On an alternative account, words can affect visual processing by setting visual priors with the effect of altering how incoming information is processed from the very start (Thierry et al., 2009; Lupyan, 2012a; Boutonnet et al., 2013; Lupyan and Ward, 2013; Francken et al., 2015; Kok and de Lange, 2014; Kok et al., 2014).

To tease apart these alternatives, we recorded brain EEGs while participants indicated whether a picture matched a previously presented verbal cue (e.g., "dog") or an equally informative nonverbal cue (e.g., dog bark). In previous studies (Lupyan and Thompson-Schill, 2012; Edmiston and Lupyan, 2013), a highly reliable label advantage was observed: people were faster in recognizing a picture of a dog after hearing "dog" than after hearing a bark.

If this label advantage derives from differences in how the two cue types activate higher-level (nonvisual) semantic representations (to which the images are matched during recognition), then verbal and nonverbal cues may elicit later ERP differences likely indexed by the N4–ERP component known to reflect semantic integration (Kutas and Federmeier, 2011).

Received Nov. 18, 2014; revised April 8, 2015; accepted May 10, 2015.

Author contributions: B.B. and G.L. designed research; B.B. performed research; B.B. and G.L. analyzed data; B.B. and G.L. wrote the paper.

The work was supported in part by the National Science Foundation (Grant BCS-1331293 to G.L.). We thank Emily J. Ward, Lynn K. Perry, Marcus Perlman, and Pierce Edmiston for valuable insight in the preparation of the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Bastien Boutonnet, University of Leiden, Postbus 9515, NL-2300 RA Leiden, The Netherlands. E-mail: bastien.b@icloud.com.

DOI:10.1523/JNEUROSCI.5111-14.2015

 0270-6474/15/359329-07$15.00/0

The alternative is that the label advantage arises because labels are especially effective at activating visual representations of features diagnostic of the cued category, thereby providing the visual system with a set of priors that bias the processing of incoming stimuli (Kok et al., 2014). Processing an incoming image in light of these priors should help to accept congruent images and reject incongruent ones (Delorme et al., 2004). If true, we expected differences in recognizing an image when its category was cued verbally versus nonverbally to be reflected in changes to early electrophysiological signals such as the P1 and/or N1 and that these modulations may be more prominent on the left hemisphere given that word recognition is strongly left lateralized and such hemispherical differences have been reported by previous investigations of language effects on color perception (Gilbert et al., 2008; Mo et al., 2011).

Classically, the P1 and N1 components are known to reflect processing of low-level visual features (e.g., contrast, color, luminance; Spehlmann, 1965; Mangun and Hillyard, 1991; Allison et al., 1999). Although it is known that the P1 can be modulated by, for example, attention to the visual modality (Foxe and Simpson, 2005; Karns and Knight, 2009) and that expectation of certain visual features can modulate early visual activations (Kok and de Lange, 2014; Kok et al., 2014), no prior work, to our knowledge, has demonstrated category-based modulation of the P1.

Materials and Methods

Participants

We tested 14 participants, all native English speakers (9 female, 5 male) from the School of Psychology at Bangor University, United Kingdom. All participants were given course credits or monetary compensation for their participation.

Stimuli

The visual stimuli comprised 50 pictures from 10 categories (5 per category: cat, car, dog, frog, gun, cockerel, train, cow, whistle, and motorcycle). Each of the 10 categories was represented by 5 different highly recognizable color images: one normed color drawing (Rossion and Pourtois, 2004), three photographs obtained from online image collections, and one less typical "cartoon" image (Lupyan and Thompson-Schill, 2012). Stimuli subtended ~9° of visual angle. Labels were recorded by a British male speaker and sounds were downloaded from online libraries. The mean label/nonverbal sound length was 0.67 ± 0.05 s. All pictures were easily nameable and all sounds were easily identifiable as determined by an extensive set of norming studies described in Lupyan and Thompson-Schill (2012). Participants produced the correct label in response to the sounds 89% of the time. The labels and nonverbal sounds were equated on an "imagery concordance" task (Rossion and Pourtois, 2004) in which people were instructed to visualize an image depicted by a spoken label or nonverbal sound and then to rate a subsequently appearing picture on how well it matched the image that they visualized.

Procedure

Participants completed 500 trials of a cued-picture recognition task. On each trial, participants heard a word (e.g., "dog") or a nonverbal sound (e.g., a dog bark). After a 1 s delay, a picture appeared and participants responded "yes" or "no" via button press to indicate whether the picture matched the auditory cue. In 50% of the trials (congruent trials), the picture matched the auditory cue at the category level ("dog" → dog or [bark] → dog). In the remaining 50% (incongruent trials), the image that followed was from one of the other nine categories. The picture remained visible until a response was made. Participants took a short break every 100 trials. All trial parameters were fully randomized within participants.

Data collection and EEG preprocessing

The EEG was recorded from 64 Ag/AgCl electrodes placed on the participants' scalp according to the extended 10–20 convention (American Electroencephalographic Society, 1994; Klem et al., 1999) at the rate of 1 kHz in reference to electrode Cz. Data were filtered offline with a high-pass 0.1 Hz filter and a low-pass 30 Hz filter and re-referenced to the common average of all scalp electrodes. Epochs ranging from −100 to 1000 ms relative to the onset of the target pictures were extracted from the continuous recording. Epochs with activity exceeding ±75 μV at any electrode site were automatically discarded. Independent components responsible for vertical and horizontal eye artifacts were identified from an independent component analysis (using the *runica* algorithm implemented in EEGLAB) and subsequently removed. Baseline correction was applied in relation to the 100 ms of prestimulus activity. After these steps, all remaining epochs were averaged by condition and for each participant. All signal-processing steps were performed in the MATLAB version 2013a (The MathWorks) environment using a combination of inhouse scripts and routines implemented in EEGLAB version 13.1.1 and ERPLAB version 4.0.2.3.

Analyses

Hypothesis testing. Statistical hypothesis testing on all but one of our analyses (described separately) was performed in the R environment (version 3.1.1). Linear mixed-effects modeling was performed using the lme4, R package (version 1.1–7; Bates et al., 2014) and *p*-values from those models were obtained using the Satterthwaite approximation implemented in the lmerTest, R package (Kuznetsova et al., 2014).

Behavioral data. We used linear mixed-effects models to predict reaction times (RTs, the time elapsed between target picture onset and participants' response) from the interaction between cue type and congruence with random slopes for cue type and congruence by participant.

ERP analyses. Four ERP components were identified from grand-averaged data. The P1, N1, and P2 were maximal at parietal sites in the 70–125, 130–180, and 190–230 ms range, respectively. The N4 was maximal over central cites and measured in the 300–500 ms time window. Mean ERP amplitudes were measured in regions of interest (ROIs) around the sites of maximal amplitude (PO3, PO4, PO7, PO8, PO9, PO10, O1, O2, for the P1, N1, and P2; FC1, FC2, FCz, C1, C2, Cz, CP1, CP2, CPz for the N4). We did not conduct a full-scalp analysis because the modulations of the ERP components were predicted to occur in the ROIs and statistical analyses were conducted on *a priori* determined electrodes. For the P1, N1, and P2 analyses, mean ERP amplitudes sampled in the windows and from electrodes corresponding to the two parieto-occipital ROIs listed above were subjected to a linear mixed-effects model, where mean amplitudes were predicted by the interaction between cue type, congruence, and laterality (left/right parieto-occipital ROI), with random slopes for cue type, congruence, and laterality by participant. To analyze the N4, the same model was run on the amplitudes collected from the electrodes corresponding to the centroparietal ROI listed above. There was no laterality factor in this model because the N4 was sampled over a single ROI. Trials on which participants made errors were discarded from all analyses involving electrophysiological data.

Single-trial analyses. Aggregating data often masks trial-to-trial variance in peak amplitudes and especially in peak latencies (Rousselet and Pernet, 2011), rendering attempts to correlate physiological measures with behavioral responses woefully underpowered. Mixed-effect models allow us to analyze the data at a single-trial level and to correlate the amplitude and latency of the earlier more reliably time-locked response (P1) on each trial with the participants' behavior (RTs).

Most single-trial analyses implement component analyses (Gerson et al., 2005; Philiastides and Sajda, 2006; Saville et al., 2011) to alleviate limitations imposed by a low signal-to-noise ratio, but this was not needed here because the P1 was highly reliable and the number of trials (~125 per condition) was more than sufficient for our statistical models. Our only step to improve the signal-to-noise ratio consisted of creating a single virtual "optimized" electrode from linear derivations of the electrodes in the two parieto-occipital ROIs mentioned above. The optimized signal from the 65–130 ms poststimulus onset was submitted to a peak-finding algorithm (based on the *findpeaks* MATLAB function)

along a series of five 13 ms sliding windows, which returned a peak location (in milliseconds) and its amplitude (in millivolts) for each trial.

Predicting cue–picture congruence from P1 single-trial activity
To determine whether properties of the P1 distinguished between trials in which the cue matched the target and those in which it did not, we used a generalized linear mixed-effects model to predict whether a given trial was congruent (i.e., the cue matched or mismatched the target image at a category level) from the interaction of single-trial peak latency, amplitude, and cue type with random slopes for cue type by participant and by item category.

Predicting behavior from P1 single-trial activity
To relate the electrophysiological data to the responses that our subjects made, we predicted single-trial RTs from single-trial P1 peak latencies and amplitudes using linear mixed-effects models (with cue type and congruence as covariates) and random slopes of the main effects of cue type and congruence by participant and by item category.

Results

Behavioral results

Label cues led to faster recognition of target image RTs than nonverbal sound cues (Fig. 1), as indicated by a highly reliable main effect of cue type ($b = -10.4$, $t = -5.4$, $p \ll 0.0001$). Congruent trials led to faster RTs compared with incongruent trials—a commonly found "yes" response advantage ($b = -30$, $t = -15.5$, $p \ll 0.0001$). The label advantage was also reflected in accuracy (Mlabel = 97%, Msound = 95%, $b = 0.01$, $t = 4.5$, $p \ll 0.0001$). Cue type and congruence did not interact reliably in the analyses of RT and accuracies ($t < 0.1$).

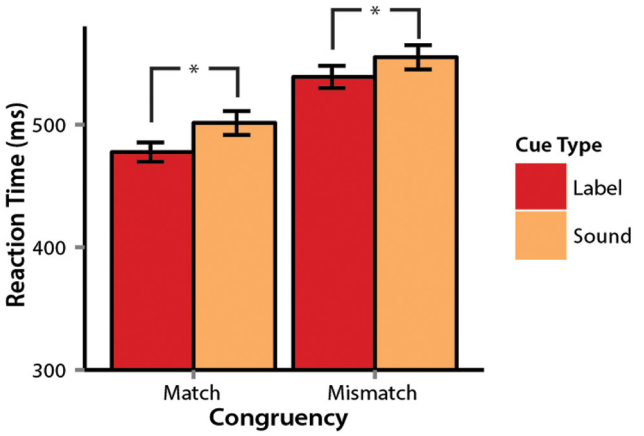

Figure 1. Mean RTs (correct responses only) showing effects of cue type and congruence. Error bars indicate 95% confidence intervals.

Electrophysiological results

Effects of cues on peak amplitudes
Pictures that were cued by labels elicited more positive P1 peak amplitudes than when the same pictures were cued by nonverbal sounds (MD = 0.45 μV; $b = 0.3$, $t = 2.37$, $p = 0.02$). The effect of cue type was numerically larger over the left hemisphere ($MD_{LEFT} = 0.57$ μV, $MD_{RIGHT} = 0.32$ μV) electrodes, but the side-by-cue type interaction was not reliable ($b = 0.2$, $t = 0.9$, $p = 0.34$; Fig. 2*A*). Congruence did not affect P1 amplitudes and did not interact with cue type ($t < 0.1$).

Pictures cued by labels also elicited more positive P2s compared with pictures cued by nonverbal sounds (MD = 0.62 μV; $b = 0.31$, $t = 2.75$, $p = 0.006$). The P2 was also modulated by cue–picture congruence, showing larger amplitudes on incongruent than congruent trials (MD = 1.04 μV; $b = -0.5$, $t = -4.59$, $p \ll .0001$; Fig. 2*A*,*B*). There was no interaction between cue type and congruence ($t < 0.1$). Neither the cue type nor the congruence manipulations or any interactions between those factors affected the N1 (Fig. 2*A*,*B*).

As in past work showing that unexpected semantic information elicits a larger N4, incongruent trials (e.g., hearing "dog" and seeing a motorcycle) considerably increased N4s compared with congruent trials (MD = −1.7 μV; $b = -0.82$, $t = -6.48$, $p \ll 0.0001$). Importantly, the N4 was not modulated by cue type ($b = 0.1$, $t = -1.1$, $p = 0.27$). There was also no reliable cue type by congruence interaction ($t < 1$; Fig. 2*C*,*D*; see below for discussion).

Relationship of the P1 to behavior
If the cues modulate directly perceptual processes brought to bear in recognizing the pictures, we expected to see a relationship between the latency or amplitude of the P1 and the behavioral response. Indeed, over and above the effects of cue type and congruence, behavioral RTs were reliably predicted by the electrophysiological measures. Specifically, earlier peak latencies of the P1 led to faster RTs ($b = 0.23$, $t = 2.38$, $p = 0.02$; Fig. 3*A*). A relationship between the P1 and RTs was also found in the amplitude domain. Larger peak amplitudes predicted faster RTs ($b = -0.56$, $t = -2.53$, $p = 0.01$; Fig. 3*B*). Behavioral RTs were predicted independently by peak latencies and amplitudes, which were correlated at $r = -0.08$ (nonsignificant).

Selective modulation of the P1 by labels
If labels are especially effective at activating category-diagnostic visual features, as we have hypothesized, then it may be possible to distinguish targets (containing such features) from nontargets (lacking these features) based on the properties of the P1, specifically on the label trials. We therefore investigated whether the modulation of the P1 by verbal labels was limited to gross-level amplitude differences (i.e., labels leading to larger/earlier P1s) or if these changes yielded selective modulation of processing, helping to distinguish matching from nonmatching images.

An analysis predicting the category-level congruence (match/mismatch) between the cue and the picture from the latency and amplitude of the P1 revealed a reliable interaction between cue type and P1 latency ($b = -0.42$, $t = -3.79$, $p = 0.0002$). *Post hoc* analyses showed that congruence was reliably predicted in label-cued trials ($b = -0.59$, $t = -3.8$, $p = 0.0002$). In contrast, the P1 did not differ between congruent and incongruent trials when people were cued by nonverbal sounds ($b = 0.24$, $t = 1.6$, $p = 0.12$; Fig. 4). Label cues reliably sped up the P1 relative to sound cues on congruent trials (MD = 1 ms; $b = -0.44$. $t = -2.27$, $p = 0.023$).

Congruence was also predicted reliably by a significant interaction between cue type and single-trial peak amplitudes ($b = 0.56$, $t = 2.23$, $p = 0.03$) whereby more positive amplitudes predicted matching trials, but the *post hoc* main effects were not individually reliable.

The analyses above show that labels modulate the P1 within 100 ms of the appearance of the target picture, allowing for discrimination between congruent and incongruent trials earlier than when people are cued nonverbally. The absence of any effects of the cue on the N4 suggests that both cue types were equally well matched to the pictures at the level of semantic congruence.

Effects of cue length and pretarget differences
Unlike previous research (Lupyan and Thompson-Schill, 2012; Edmiston and Lupyan, 2013), the length of the verbal and non-

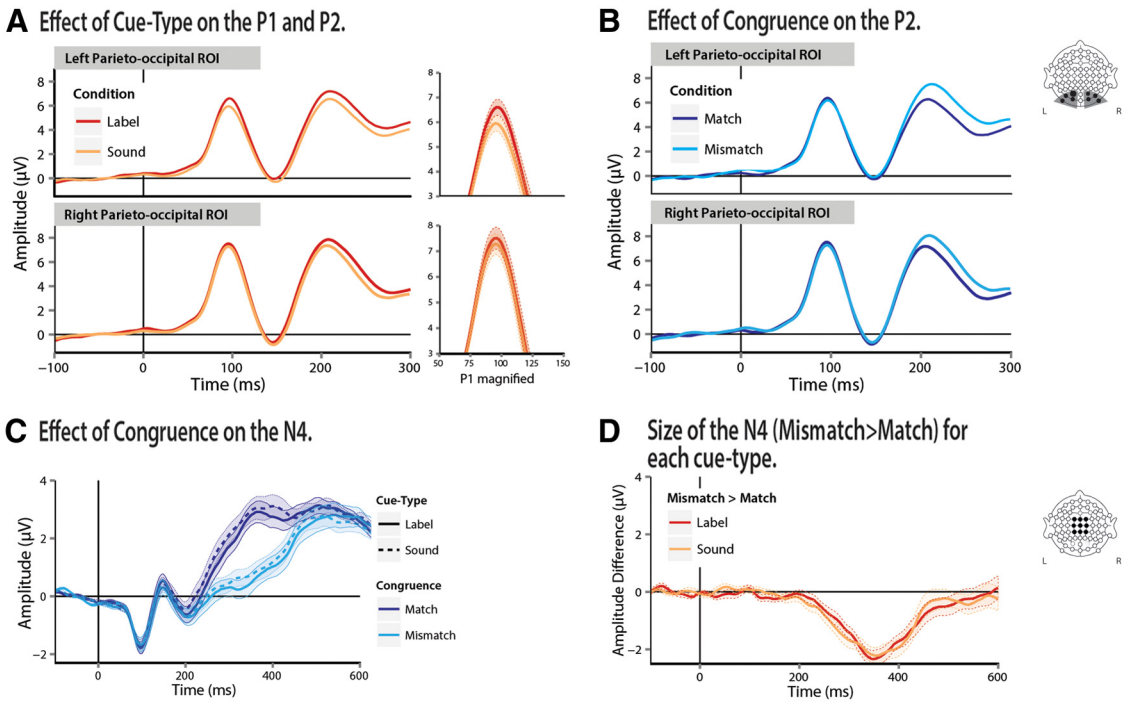

Figure 2. ERPs for the two experimental factors across the parieto-occipital and central ROIs (linear derivations of PO3, PO7, PO9, O1 for the left and PO4, PO8, PO10, O2 for the right). ***A***, P1 and P2 ERPs across the two levels of cue type and magnification of the P1 wave. In red are ERPs elicited by label-cued pictures and in orange are ERPs elicited by sound-cued pictures. ***B***, P1 and P2 ERPs across the two levels of congruence. In dark blue are ERPs elicited by pictures that were congruent with target and in light blue are ERPs elicited by pictures that were incongruent with the cue. ***C***, N4 ERPs across the two levels of cue type and congruence. In dark blue are matching trials. In light blue are mismatching trials. Solid lines represent label-cued trials and dashed lines represent sound-cued trials. ***D***, Difference waves of the effect of congruence on the N4 for each type of cue. Red shows the difference between mismatching and matching trials cued by a label. Orange shows the difference between mismatching and matching trials cued by a nonverbal sound.

verbal cues was not fully equated in the present study. These small differences meant that the delay between cue and picture onset was slightly longer for some cues compared with others. To rule out the possibility that the effects reported above are due to differences in cue length, we recomputed the analyses above partialing out effects of sound length by adding item sound length to the fixed-effect structure of the linear model. The effects of congruence on the P1 presented above were not altered.

An additional potential concern is that any observed electrophysiological differences between verbally cued trials and nonverbally cued trials may stem from differences in neural activity elicited by the two auditory cues rather than from the cues affecting the visual processing of subsequently shown pictures. If true, the observed differences in the P1 may reflect the difference between, hearing speech versus nonspeech. To address this concern, we analyzed the EEG signal before the target picture for each cue type. Epochs corresponding to 1 s of activity preceding target onset were averaged for each cue type. Differences between cue types were assessed using a paired, two-tailed permutation test based on the *t*-max statistic (Blair and Karniski, 1993) with a familywise α level of 0.05 at all time points of the 1 s pretarget onset activity and from all electrodes corresponding to the 3 ROIs used in the ERP analyses. This statistical analysis was performed in the Mass Univariate MATLAB toolbox (Groppe et al., 2011). The analysis of the pretarget activity failed to detect any significant differences at any time points between the activity generated by labels and sounds in the −1000 to 0 ms time window.

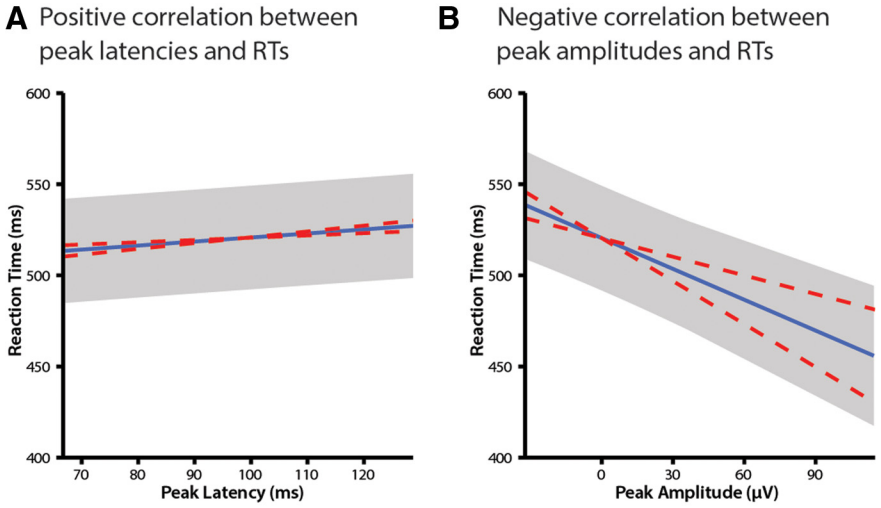

Figure 3. P1 peak latencies (***A***) and peak amplitudes (***B***) were predictive of RTs. Lines depict best linear fits. Error bars indicate ±1 SE of the intercept and slope. Dashed red lines show ±1 SE of the slope.

Discussion

At its most basic, language allows people to "verbally point" to something. When asked to "look at the car," we expect people to do just that. This type of linguistic control of behavior is often assumed to happen at a relatively "high" semantic level (Jackend-

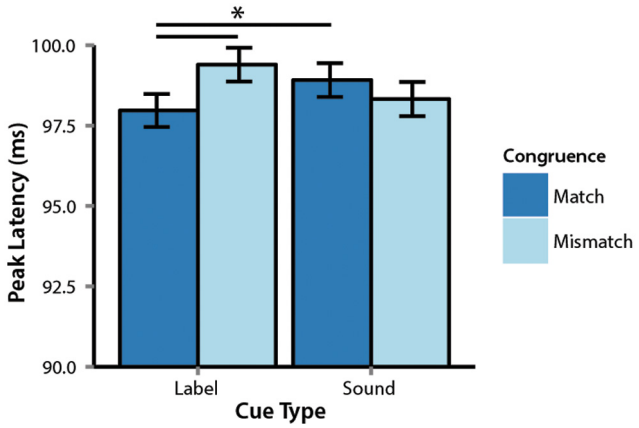

Figure 4. Interaction between cue type and congruence on P1 peak latencies. Error bars indicate 95% confidence intervals. Vertical bars indicate significant differences.

off, 2002; Li et al., 2009). Effects of language on cognition are thus viewed as deriving from changes to such higher-level processes, like working memory or decision making, which are typically thought to happen after visual processing is completed (Pylyshyn, 1999; Klemfuss et al., 2012). In this "labels-as-pointers" view, the idea that people recognize images more quickly and accurately when they are cued by words than by equally familiar and unambiguous sounds (Lupyan and Thompson-Schill, 2012; Edmiston and Lupyan, 2013) is understood in terms of labels being more effective at activating high-level, and putatively amodal, semantic representations than are sounds.

On an alternative account, this label advantage has a perceptual locus: verbal and nonverbal cues provide different top-down signals to the visual system, changing how subsequently incoming (bottom-up) information is processed (Delorme et al., 2004; Kok et al., 2012a, 2014).

To tease apart these two alternatives, we used EEG to measure real-time brain activity to determine the precise timing of the effects underlying the label advantage. We expected that, if the advantage had a semantic locus, then differences in brain activity elicited by label-cued versus sound-cued pictures would occur late in time and be reflected by ERP components commonly associated with semantic integration—namely the N4 (Kutas and Federmeier, 2011). If instead labels potentiate visual processes, then we expected to see modulations of early electrophysiological signals classically associated with bottom-up processes (Spehlmann, 1965; Mangun and Hillyard, 1991; Allison et al., 1999).

Our results unambiguously support the hypothesis that the label advantage has a perceptual locus. Labels led to more positive and earlier P1s (the generators of which are linked to extrastriate cortex; Di Russo et al., 2003). Importantly, the peak latency of the P1 was sensitive to the congruence between the cue and the target, but only when people were cued by a word. This suggests that, after hearing a word, the neural processes responsible for generating the P1 were already sensitive to the object's category. Our analysis of the pretarget activity suggests that the observed differences do not arise before target presentation, confirming that the effects observed on the P1 are due to genuine cue–picture visual integration. Moreover, we found a strong correlation between P1 activities and behavior (earlier P1s predicted shorter correct RTs), indicating that the processes generating the P1 were not incidental to the behavioral response.

Furthermore, whereas incongruent cues led to greater N4s compared with congruent ones (a classic N4 semantic incongruity effect; Kutas and Hillyard, 1984; Kutas and Federmeier, 2011), these effects were equivalent for verbal and nonverbal cues, supporting our contention that the two cue types were equivalently informative/unambiguous at a semantic level in the context of this task. Both the cue type and congruence effects were also reflected on the P2, an ERP component known to index matching processes between sensory inputs and memory (Luck and Hillyard, 1994; Freunberger et al., 2007). Together, these results suggest a perceptual locus for the label advantage.

The present findings apply directly to the larger question of the relationship of language, cognition, and perception. In the view advocated here, words do not simply point to preexisting semantic categories, but help to reify the categories that they denote (James, 1890; Lupyan et al., 2007; Lupyan, 2012b). In this view, comprehending a word such as "dog" activates visual properties diagnostic of dogs (Simmons et al., 2007; Evans, 2009; Kuipers et al., 2013; Pulvermüller, 2013). Subsequently presented visual inputs are processed in light of these activated representations (i.e., words act as high-level priors on the visual system; Lupyan and Clark, 2015). The functional consequence is that hearing a word allows for more effective representational separation of category members and nonmembers. Hearing "dog" effectively turns one into a better "dog-detector." The same mechanism can explain why hearing an informationally redundant word improves visual search (Lupyan, 2008), how words can unsuppress visual representations masked through continuous flash suppression (Lupyan and Ward, 2013), why labels make differences between objects more or less salient to the visual system (Boutonnet et al., 2013), and why processing motion-related words such as "float" or "dive" affects discrimination of visual motion (Meteyard et al., 2007; Francken et al., 2015). The same mechanism may also underlie the kind of warping of cortical representations after verbal instructions reported by Çukur et al. (2013).

In addition to informing the mechanisms by which words can influence visual processing, our work shows that cues can affect the P1 in a semantically coherent way on a trial-by-trial basis. Although prior work has shown that the P1 is modulated by cues that signal the task-relevant modality (e.g., visual vs auditory; Foxe and Simpson, 2005; Foxe et al., 2005; Karns and Knight, 2009) and that the P1 is altered by learning the name or the function of the depicted object (Rahman and Sommer, 2008; Maier et al., 2014), our findings are, to our knowledge, the first to show that the P1 can be selectively modulated online by category-specific cues and that the P1 predicts overt visual recognition responses occurring 500 ms later.

What is special about labels?

What makes labels more effective than nonverbal sounds at cuing visual knowledge? As discussed by Lupyan and Thompson-Schill (2012), the label advantage cannot simply be explained by differences in cue familiarity because it persists at longer cue-to-target delays, leaving enough time for the potentially unfamiliar cue to be processed. In fact, even newly learned, and thus completely unfamiliar, labels ("alien musical instruments") still have an advantage over nonverbal associates (i.e., the sound of these instruments, experiment 4 in Lupyan and Thompson-Schill, 2012), suggesting that those new labels inherit the categorical properties of familiar ones. The label advantage remains even when the congruence between sounds and pictures is maximized (e.g., an electric guitar sound cuing an electric guitar vs "guitar" cuing an electric guitar; P. Edmiston and G. Lupyan, unpublished data). That both cue types yield virtually identical N4s further suggests

that the label advantage is not simply to labels being more familiar or strongly associated with the target images. Rather, the difference between the two cue types stems from the relationship between words and referents (Edmiston and Lupyan, 2013; Lupyan and Bergen, 2015). Words denote categories. The word "dog" denotes all dogs, abstracting over the idiosyncrasies of particular exemplars. In contrast, nonverbal cues, no matter how unambiguous and familiar, are necessarily linked to a particular dog (e.g., a high-pitched bark is produced by a smaller dog), making them less categorical.

Therefore, although it is possible to convey information about specific exemplars both through verbal and nonverbal means, language may be uniquely well suited for activating categorical states, which in this case enable people to distinguish more effectively between category members and nonmembers.

Role of mental imagery

Hearing cues—whether verbal or nonverbal—triggers rapid and largely automatic activations of visual representations (a type of implicit mental imagery). The neural machinery underlying such activations is likely to be substantially shared with perception, visual working memory, and, indeed, explicit mental imagery (Pearson et al., 2008; Pratte and Tong, 2014). However, we think that the hundreds of trials and the rapid pacing of the task make it unlikely that participants are engaging in the sort of strategic and explicit imagery explored by, for example, Kosslyn et al. (2006) and Farah (1989).

Interpreting the P1 enhancement within a predictive processing framework

The finding that hearing a label enhanced the P1, modulating it differentially for targets and nontargets, is well accommodated by predictive processing frameworks (Kok et al., 2012a; Clark, 2013). The auditory cue that people hear ahead of the picture activates visual predictions. For example, the label "dog" or the barking sound may activate visual representations corresponding to a dog shape. The picture (target) that subsequently appears is processed in light of these predictions. Insofar as the predictions are accurate, they will help to recognize an image from the cued category or reject an image from a nonmatching category. More specifically, the cues may increase the weighting of incoming sensory evidence consistent with the predictions (Kok et al., 2012b). As detailed above, what distinguishes the two cue types is that labels are uniquely suitable for generating categorical predictions.

Conclusions

People are faster to recognize an image at a categorical level after hearing a label (e.g., "dog") than after hearing a nonverbal cue (e.g., dog bark). We provide evidence that this label advantage stems from labels setting category-level priors that alter how a subsequent image is visually processed. Labels appear to activate category-specific visual features altering visual processing within 100 ms of stimulus onset. These early modulations were strongly predictive of behavioral performance occurring almost half a second later showing that the P1 is sensitive to semantic information. The ability of words to act as categorical cues that rapidly set perceptual priors has consequences for understanding aspects of human cognition thought to depend on categorical representations such as inference, compositionality, rule following, and formal reasoning.

References

Rahman RA, Sommer W (2008) Seeing what we know and understand: how knowledge shapes perception. Psychon Bull Rev 15:1055–1063. CrossRef Medline

Allison T, Puce A, Spencer DD, McCarthy G (1999) Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cereb Cortex 9:415–430. CrossRef Medline

American Electroencephalographic Society (1994) Guideline thirteen: guidelines for standard electrode position nomenclature. J Clin Neurophysiol 11:111–113. CrossRef Medline

Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4, Ed 1. Available from: http://CRAN.R-project.org/package=lme4.

Blair RC, Karniski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30:518–524. CrossRef Medline

Bloom P (2000) How children learn the meanings of words. Cambridge, MA: MIT.

Boutonnet B, Dering B, Viñas-Guasch N, Thierry G (2013) Seeing objects through the language glass. J Cogn Neurosci 25:1702–1710. CrossRef Medline

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204. CrossRef Medline

Çukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. Nat Neurosci 16:763–770. CrossRef Medline

Delorme A, Rousselet GA, Macé MJ, Fabre-Thorpe M (2004) Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. Brain Res Cogn Brain Res 19:103–113. CrossRef Medline

Dessalegn B, Landau B (2008) More than meets the eye: the role of language in binding and maintaining feature conjunctions. Psychol Sci 19:189–195. CrossRef Medline

Di Russo F, Martínez A, Hillyard SA (2003) Source analysis of event-related cortical activity during visuo-spatial attention. Cereb Cortex 13:486–499. CrossRef Medline

Edmiston P, Lupyan G (2013) Verbal and nonverbal cues activate concepts differently, at different times. Proceedings of the 35th Annual Conference of the Cognitive Science Society, Berlin, July–August.

Evans V (2009) How words mean. Oxford: OUP.

Farah MJ (1989) Mechanisms of imagery-perception interaction. J Exp Psychol Hum Percept Perform 15:203–211. CrossRef Medline

Firestone C, Scholl BJ (2014) "Top-down" effects where none should be found: the El Greco fallacy in perception research. Psychol Sci 25:38–46. CrossRef Medline

Foxe JJ, Simpson GV (2005) Biasing the brain's attentional set: II. Effects of selective intersensory attentional deployments on subsequent sensory processing. Exp Brain Res 166:393–401. CrossRef Medline

Foxe JJ, Simpson GV, Ahlfors SP, Saron CD (2005) Biasing the brain's attentional set: I. Cue driven deployments of intersensory selective attention. Exp Brain Res 166:370–392. CrossRef Medline

Francken JC, Kok P, Hagoort P, de Lange FP (2015) The behavioral and neural effects of language on motion perception. J Cogn Neurosci 27:175–184. CrossRef Medline

Freunberger R, Klimesch W, Doppelmayr M, Höller Y (2007) Visual P2 component is related to theta phase-locking. Neurosci Lett 426:181–186. CrossRef Medline

Gerson AD, Parra LC, Sajda P (2005) Cortical origins of response time variability during rapid discrimination of visual objects. Neuroimage 28:342–353. CrossRef Medline

Gilbert AL, Regier T, Kay P, Ivry RB (2008) Support for lateralization of the Whorf effect beyond the realm of color discrimination. Brain Lang 105:91–98. CrossRef Medline

Gleitman L, Papafragou A (2005) Language and thought. In: Cambridge handbook of thinking and reasoning (Holyoak KJ, Morrison B, eds), pp 633–661. Cambridge: Cambridge University.

Groppe DM, Urbach TP, Kutas M (2011) Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48:1711–1725. CrossRef Medline

Jackendoff R (2002) Foundations of language: brain, meaning, grammar, evolution. Oxford: OUP.

James W (1890) Principles of psychology. New York: H. Holt.
Karns CM, Knight RT (2009) Intermodal auditory, visual, and tactile attention modulates early stages of neural processing. J Cogn Neurosci 21:669–683. CrossRef Medline
Klem GH, Lüders HO, Jasper HH, Elger C (1999) The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol Suppl 52:3–6. Medline
Klemfuss N, Prinzmetal W, Ivry RB (2012) How does language change perception: a cautionary note. Front Psychol 3:78. Medline
Kok P, de Lange FP (2014) Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. Curr Biol 24:1531–1535. CrossRef Medline
Kok P, Jehee JF, de Lange FP (2012a) Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75:265–270. CrossRef Medline
Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP (2012b) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197–2206. CrossRef Medline
Kok P, Failing MF, de Lange FP (2014) Prior expectations evoke stimulus templates in the primary visual cortex. J Cogn Neurosci 26:1546–1554. CrossRef Medline
Kosslyn SM, Thompson WL, Ganis G (2006) The case for mental imagery. Oxford: OUP.
Kuipers JR, van Koningsbruggen M, Thierry G (2013) Semantic priming in the motor cortex. Neuroreport 24:646–651. CrossRef Medline
Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol 62:621–647. CrossRef Medline
Kutas M, Hillyard SA (1984) Brain potentials during reading reflect word expectancy and semantic association. Nature 307:161–163. CrossRef Medline
Kuznetsova A, Bruun Brockhoff P, Haubo Bojesen Christensen R (2014) lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package) R package version 2.0–6. Available from: http://CRAN.R-project.org/package=lmerTest.
Li P, Dunham Y, Carey S (2009) Of substance: the nature of language effects on entity construal. Cogn Psychol 58:487–524. CrossRef Medline
Luck SJ, Hillyard SA (1994) Electrophysiological correlates of feature analysis during visual search. Psychophysiology 31:291–308. CrossRef Medline
Lupyan G (2008) The conceptual grouping effect: categories matter (and named categories matter more). Cognition 108:566–577. CrossRef Medline
Lupyan G (2012a) Linguistically modulated perception and cognition: the label-feedback hypothesis. Front Psychol 3:54. Medline
Lupyan G (2012b) What do words do? toward a theory of language-augmented thought. In: Psychology of learning and motivation (Ross BH, ed), pp 255–297. New York: Elsevier.
Lupyan G, Bergen B (2015) How language programs the mind. Topics in Cognitive Science. New Frontiers in Language Evolution and Development. In press.
Lupyan G, Clark A (2015) Words and the world: predictive coding and the language- perception-cognition interface. Current Directions in Psychology. In press.
Lupyan G, Thompson-Schill SL (2012) The evocative power of words: Activation of concepts by verbal and nonverbal means. J Exp Psychol Gen 141:170–186. CrossRef Medline
Lupyan G, Ward EJ (2013) Language can boost otherwise unseen objects into visual awareness. Proc Natl Acad Sci U S A 110:14196–14201. CrossRef Medline
Lupyan G, Rakison DH, McClelland JL (2007) Language is not just for talking: redundant labels facilitate learning of novel categories. Psychol Sci 18:1077–1083. CrossRef Medline
Maier M, Glage P, Hohlfeld A, Abdel Rahman R (2014) Does the semantic content of verbal categories influence categorical perception? An ERP study. Brain Cogn 91:1–10. CrossRef Medline
Mangun G, Hillyard SA (1991) Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. J Exp Psychol Hum Percept Perform 17:1057–1074. CrossRef Medline
Meteyard L, Bahrami B, Vigliocco G (2007) Motion detection and motion verbs: language affects low-level visual perception. Psychol Sci 18:1007–1013. CrossRef Medline
Mo L, Xu G, Kay P, Tan LH (2011) Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. Proc Natl Acad Sci U S A 108:14026–14030. CrossRef Medline
Pearson J, Clifford CW, Tong F (2008) The functional impact of mental imagery on conscious perception. Curr Biol 18:982–986. CrossRef Medline
Philiastides MG, Sajda P (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 16:509–518. CrossRef Medline
Pratte MS, Tong F (2014) Spatial specificity of working memory representations in the early visual 1 cortex. J Vis 14:22. Medline
Pulvermüller F (2013) How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends Cogn Sci 17:458–470. CrossRef Medline
Pylyshyn Z (1999) Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. Behav Brain Sci 22:341–423. Medline
Rossion B, Pourtois G (2004) Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. Perception 33:217–236. CrossRef Medline
Rousselet GA, Pernet CR (2011) Quantifying the time course of visual object processing using ERPs: it's time to up the game. Front Psychol 2:107. Medline
Saville CW, Dean RO, Daley D, Intriligator J, Boehm S, Feige B, Klein C (2011) Electrocortical correlates of intra- subject variability in reaction times: average and single-trial analyses. Biol Psychol 87:74–83 CrossRef Medline
Simmons WK, Ramjee V, Beauchamp MS, McRae K, Martin A, Barsalou LW (2007) A common neural substrate for perceiving and knowing about color. Neuropsychologia 45:2802–2810. CrossRef Medline
Spehlmann R (1965) The averaged electrical responses to diffuse and to patterned light in the human. Electroencephalogr Clin Neurophysiol 19:560–569. CrossRef Medline
Thierry G, Athanasopoulos P, Wiggett A, Dering B, Kuipers JR (2009) Unconscious effects of language-specific terminology on preattentive color perception. Proc Natl Acad Sci U S A 106:4567–4570. CrossRef Medline
Vetter P, Smith FW, Muckli L (2014) Decoding sound and imagery content in early visual cortex. Curr Biol 24:1256–1262. CrossRef Medline