Language can boost otherwise unseen objects into visual awareness

Gary Lupyan[a,1] **and Emily J. Ward**[b]

[a]University of Wisconsin-Madison, Madison, WI 53706; and [b]Yale University, New Haven, CT 06520

Edited by James. L. McClelland, Stanford University, Stanford, CA, and approved June 25, 2013 (received for review February 20, 2013)

Linguistic labels (e.g., "chair") seem to activate visual properties of the objects to which they refer. Here we investigated whether language-based activation of visual representations can affect the ability to simply detect the presence of an object. We used continuous flash suppression to suppress visual awareness of familiar objects while they were continuously presented to one eye. Participants made simple detection decisions, indicating whether they saw any image. Hearing a verbal label before the simple detection task changed performance relative to an uninformative cue baseline. Valid labels improved performance relative to no-label baseline trials. Invalid labels decreased performance. Labels affected both sensitivity (*d′*) and response times. In addition, we found that the effectiveness of labels varied predictably as a function of the match between the shape of the stimulus and the shape denoted by the label. Together, the findings suggest that facilitated detection of invisible objects due to language occurs at a perceptual rather than semantic locus. We hypothesize that when information associated with verbal labels matches stimulus-driven activity, language can provide a boost to perception, propelling an otherwise invisible image into awareness.

vision | top-down effects | CFS | penetrability of perception

To what extent is awareness of a visual stimulus affected by factors outside of vision? Although any number of nonvisual factors can affect where one attends and thus what one is aware of, some have argued that the contents of visual perception are impenetrable to outside influences (e.g., ref. 1). However, there is growing evidence suggesting that information outside the visual modality can change what one sees (e.g., appropriately timed sounds can qualitatively change visual percepts) (2, 3). Here we demonstrate that auditory linguistic labels can affect not just what one sees (a process that likely involves some degree of higher-level interpretation) but whether one sees something. For example, we demonstrate that an otherwise invisible kangaroo can be boosted into visual awareness by language.

The idea that "higher"-level processes such as word-recognition can influence "lower"-level processes presents challenges for strictly feed-forward models of cognition and perception but is consistent with interactive-activation accounts, which posit that neural processing is intrinsically interactive (4–6): sensory input signals are only a part of what drives "sensory" neurons, "…processing stages are not like assembly line productions, [and later processing can influence earlier processing]" (7). Such interactive-activation accounts explain otherwise counterintuitive findings, such as the word-superiority effect (e.g., ref. 6) and changes to visual processing due to contextual information (8, 9) (this process of constraining lower levels of processing by higher levels is by no means unique to vision; e.g., see ref. 10 for review of interactive processes in speech perception).

Understanding the relationship between vision and language is important for several reasons. First, it speaks to the modularity thesis, one formulation of which specifically describes language and vision as independent modules (11). On this account, processes in one domain do not influence processes in the other. Although there are numerous demonstrations of such influences (refs. 12–14 for reviews), there continues to be controversy concerning the extent to which visual representations are "penetrable" by factors outside of vision. Rather than showing effects on visual awareness per se, most existing demonstrations concern subtle changes to certain qualities of the visual experience, or can be attributed to interpretive or decision biases (1, 15–18). A second reason for understanding the relationship between language and vision is that humans live in a linguistic world, with much of human behavior guided by language. Despite the ubiquity of language, the extent to which its effects on behavior include changes to ongoing perceptual processing rather than solely to higher-level decision processes is still largely unexplored. Finally, an improved understanding of interactions between language and perception clarifies and constrains theorizing about cases when learning and using different languages should or should not affect perceptual processing (e.g., refs. 19–21).

Verbal labels have been shown to affect performance on visual search tasks, in which completely redundant verbal labels seem to exaggerate perceived differences between targets and distractors (22, 23). Lupyan and Swingley (24) likewise showed that actually saying a label affected visual search performance over and above simply knowing what to search for. In tasks requiring identification of the direction of moving dots, performance was worse after hearing a direction-incongruent verb (25), and simply imagining or hearing a story about motion can produce motion aftereffects (26). Conversely, lexical decision times for motion words increased in the presence of congruent visual motion and decreased with incongruent motion (27), suggesting bidirectional influences between linguistic and visual–motion processing. Similar demonstrations have been shown in the domain of contrast sensitivity (28) and face processing (29).

Combined, these results suggest that words rapidly and automatically activate visual properties of the entities to which they refer, and the effects of words on perception arguably stem from their activating representations that overlap with those used in visual processing. However, an alternative interpretation is that the effects of words occur at a postperceptual decision or semantic level. According to this interpretation, words and semantic information may affect participants' judgments about a perceptual experience or their memory of a perceptual experience rather perception itself (1, 30). Furthermore, it is difficult to distinguish perception from perceptual identification (e.g., object recognition), the latter of which is itself a higher-level process affected by knowledge and experience (31, 32). A stronger test of the hypothesis that top-down cues more broadly and language more specifically can modulate a fundamental aspect of perception is to determine whether such top-down signals can make the difference between seeing and not seeing something, that is, alter visual awareness—a fundamental aspect of perception.

An effective approach for studying top-down influences on perception is to keep visual stimulation constant while manipulating the top-down influence that is hypothesized to augment the perceptual process in question. Any differences in perception due to stimulus-driven processes can then be ruled out. In

---

Author contributions: G.L. and E.J.W. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: lupyan@wisc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1303312110/-/DCSupplemental.

a previous study that examined effects of verbal labels on detecting the presence of visual stimuli, participants performed an object-detection task for briefly presented backward-masked letters that were preceded by auditory labels (e.g., "emm") (33). The authors found that hearing the letter name improved simple detection of the corresponding letters compared with hearing noise or irrelevant words. However, a postperceptual account can still explain these findings because backward-masking, although effective in impairing recognition, does little to reduce bottom-up visual processing (e.g., refs. 34 and 35), and unseen backward-masked stimuli can still be processed semantically as measured by semantic priming; there are even reports of subliminal cross-linguistic priming (36)—clear evidence that backward-masked primes are processed at a semantic level (ref. 37 for review).

Therefore, it is imperative to demonstrate top-down influences on visual awareness using a technique that suppresses explicit visual awareness while also suppressing semantic processing. One way to achieve this is through interocular rivalry. By presenting different stimuli to the two eyes, the stimuli compete and the same retinal input can give rise to different conscious percepts. Which percept gains dominance is determined by numerous visual aspects of the stimulus, such as contrast, contour density, luminosity, size, eccentricity, etc. (ref. 38 for review), but importantly, there is strong evidence that the percept of which the subject is unaware receives minimal if any semantic processing (addressed further in *Discussion* and *SI Results and Discussion, Semantic Processing of Images Made Invisible by CFS*). A version of interocular rivalry particularly well suited for examining visual awareness is continuous flash suppression (CFS). In CFS, an object is placed in interocular competition with high-contrast noise patterns that alternate at ∼10 Hz, which acts to suppress the stimulus from awareness for extended durations (39).

In the first two experiments, we tested whether hearing a verbal cue can make visible an otherwise invisible image, suppressed through CFS, by measuring visual detection (using explicit self-report, a basic measure of awareness; e.g., ref. 40). We reasoned that if processing verbal labels activates perceptual representations that contribute to awareness, then hearing a label before viewing a picture made invisible through CFS should change the likelihood of actually seeing the picture. Valid labels, for example the word "kangaroo" preceding a picture of a kangaroo, should increase the likelihood of seeing the picture. Invalid labels should decrease the likelihood. To further test the hypothesis that effects of labels on visual detection occur owing to labels activating visual features with which they are associated, in experiment 3 we examined whether the effectiveness of labels depends on the degree of correspondence between the shape denoted by the label and visual shape presented to the participants. We conclude that language-based activation of visual representations can act as a top-down "boost" to perception that, under some circumstances, can propel an otherwise invisible image into awareness.

Results

Experiment 1. We investigated the effect of cue type on detection and recognition of suppressed visual objects (Fig. 1*A*) by analyzing several dependent measures of image detection. First we calculated detection accuracy in terms of hit rate and false alarms and calculated a signal detection sensitivity measure (d'). We also analyzed participants' correct-trial reaction times (RTs). The experimental design crossed object-presence (present, absent) and cue type [absent, valid, invalid (Fig. 2*B*)]. All factors were within-subjects.

Overall, participants correctly detected 49.8% of the presented objects and false alarmed (responding "present" on object-absent trials) 15% of the time. Verbal cues had no effect on false alarms, $F < 1$, but had a reliable effect on the hit rate, $F_1(2,38) = 4.88$, $P = 0.013$ (subject-based), $F_2(2,14) = 5.47$, $P = 0.018$ (item-based) (Fig. 2*A*). Relative to the no-cue baseline ($M = 0.48$), valid cues increased correct detection ($M = 0.53$), $t(19) = 2.52$, $P = 0.021$. Detection after invalid cues ($M = 0.46$) was reliably lower than after valid cues, $t(19) = 2.74$, $P = 0.013$, but not reliably lower than on the baseline trials, $t < 1$ (Fig. 2*A*). A signal detection analysis likewise showed a significant effect of cuing condition, $F(2,38) = 4.93$, $P = 0.012$ (Fig. 2*B*): d' was significantly lower on invalid trials ($M = 1.18$) relative to valid trials ($M = 1.44$), $t(19) = 2.94$, $P = 0.008$. Performance on the no-cue baseline was intermediate ($M = 1.32$); pair-wise comparisons between baseline and valid, and baseline and invalid were marginal. Cues had no effect on response criteria (Ps > 0.3).

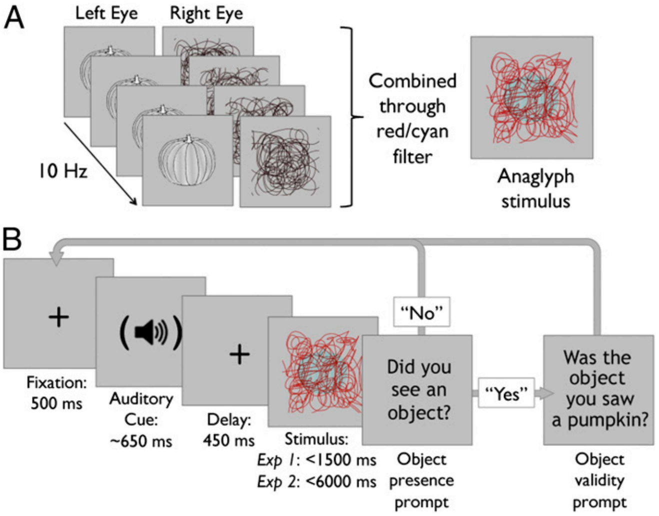

Fig. 1. (*A*) Stimulus creation using (CFS). (*B*) Basic procedure of experiments 1 and 2.

An analysis of median RTs on correct-response trials mirrored the hit-rate/false-alarm analysis. Verbal cues did not affect RTs on object-absent trials, $F(2,38) = 1.26$, $P = 0.29$ (Fig. 2*C*) but reliably affected RTs on object-present trials, $F(2,38) = 6.49$, $P = 0.004$. Relative to the no-cue baseline ($M = 1{,}886$ ms), valid cues decreased RTs ($M = 1{,}782$ ms), $t(19) = 2.41$, $P = 0.026$. RTs after invalid cues were considerably slower ($M = 1{,}991$ ms) than those after valid trials, $t(19) = 2.95$, $P = 0.008$, and marginally slower relative to the no-cue baseline, $t(19) = 1.85$, $P = 0.08$.

Object detection, measured both by detection accuracy and RTs, was affected by verbal labels. Labels did not affect criteria, false alarms, or RTs on object-absent trials, replicating earlier results from a backward-masking paradigm (ref. 33; *Discussion*) and supporting the hypothesis that labels provide a top-down boost, which, under some circumstances, can propel an otherwise invisible image into awareness. In the absence of a bottom-up input to receive the boost (object-absent trials), the label has no effect on detection rates. Although some of the differences

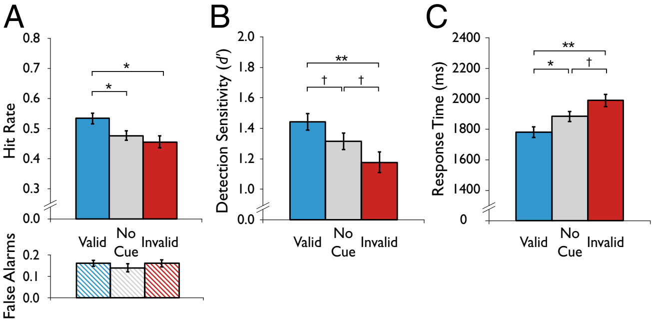

Fig. 2. (*A*) Hit rates and false alarms, (*B*) detection sensitivity, and (*C*) response times for detecting objects suppressed through CSF in experiment 1. Error bars show ±1 SEM with between-subject variance removed (74).

between the baseline and cuing conditions were marginal, we were primarily interested in the main effect of cuing and the critical comparison between valid and invalid cues—both of which were consistently reliable, showing the power of cuing despite identical visual inputs in the three conditions.

In addition to affecting detection rates, verbal labels also affected recognition responses as tested by verification prompts shown when participants claimed to see an object. When asked whether the image was, for example, a zebra, participants were more likely to respond "yes" on valid trials (i.e., when they heard the word "zebra") than on no-cue baseline trials, $M_{\text{no-cue}} = 0.69$, $M_{\text{valid-cue}} = 0.88$, $t(19) = 4.03$, $P = 0.001$, and marginally less likely to respond "yes" on invalid trials ($M = 0.55$) than on no-cue trials, $t(19) = 1.94$, $P = 0.07$. The simplest explanation for these results is a bias to interpret an otherwise indeterminate image as matching a previously heard label. This interpretation also receives support from false-alarm trials. When participants were asked if the (nonexistent) image they "saw" was a zebra, they were much more likely to respond "yes" when they previously heard the word "zebra" than when they did not, $M_{\text{no-cue}} = 0.36$, $M_{\text{valid-cue}} = 0.80$, $t(15) = 3.98$, $P = 0.001$ (four participants had no false alarms for valid/no-cue trials). Although such a bias explanation is sufficient to explain the effect of labels on recognition responses, it cannot explain how labels increased the likelihood of detecting the mere presence of a stimulus.

A rarely noted property of CFS is that accuracy can improve substantially over time, suggesting a kind of "learning to see" takes place. In the present study the hit rate increased from 36% in the first 40 trials to 64% in the last 40 trials. Cuing effects tended to become larger over the course of the experiment, but these changes were not reliable.

Experiment 2. In experiment 1, only eight labels and eight unique objects were presented to participants (Table S1; each image was presented 40 times). This leaves open the possibility that, particularly later in the experiment, on hearing a particular label, participants may have shifted their attention, overtly or covertly, to parts of space most likely to contain relevant visual information. To help rule out such object-specific strategies, we replicated experiment 1 using more complex and variable images (photographs from many categories, with multiple exemplars from each). This introduced more uncertainty into the object properties and also allowed us to investigate the effect of labels on detecting naturalistic stimuli. Results for experiment 2 were very similar to those for experiment 1 except for considerably higher overall performance, presumably due to longer display time (6 s vs. 1.5 s) (Fig. 1*B*). Overall hit rates were 80.2% and false alarms 3.7%. As in experiment 1, verbal cues had a reliable effect on hit rates, $F_1(2{,}38) = 5.47$, $P = 0.008$ (subject-based), $F_2(2{,}78) = 6.30$, $P = 0.003$ (item-based) (Fig. 3*A*), but not false alarms, $F\text{s} < 1$. Hit rates on valid trials ($M = 0.82$) were greater than on invalid trials ($M = 0.75$), $t(19) = 2.98$, $P = 0.008$. A signal detection analysis showed a marginal overall effect of cuing condition, $F(2{,}38) = 2.71$, $P = 0.079$ (Fig. 3*B*), but the overall pattern was very similar to that in experiment 1: significantly greater sensitivity on valid ($M = 3.11$) than invalid trials ($M = 2.80$), $t(19) = 2.56$, $P = 0.019$, with a smaller (and this time marginal) difference between valid and no-cue trials ($M = 2.89$), $t(19) = 1.89$, $P = 0.08$, and a nonreliable decrease in sensitivity for invalid trials, $t < 1$. There were no differences in criterion using either c or $\log(\beta)$, $t < 1$. In sum, recognition responses revealed a pattern very similar to that in experiment 1.

Analysis of median RTs showed no effect on object-absent trials, $F < 1$, but a robust effect on object-present trials, $F(2{,}38) = 7.75$, $P = 0.002$ (Fig. 3*C*). Relative to the no-cue baseline ($M = 3{,}163$ ms), valid cues decreased RTs ($M = 2{,}735$ ms), $t(19) = 2.71$, $P = 0.014$. RTs after invalid cues were considerably slower ($M = 3{,}366$ ms) than those after valid trials, $t(19) = 3.84$, $P = 0.001$, although not reliably slower than the no-cue baseline, $t(19) = 1.20$, $P = 0.243$.

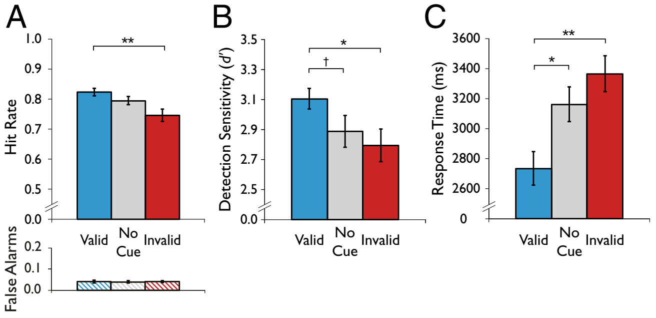

Fig. 3. (*A*) Hit rates and false alarms, (*B*) detection sensitivity, and (*C*) response times for detecting objects suppressed through CSF in experiment 2, which contained 160 separate objects in 40 categories, with each image presented for a maximum of 6 s. Error bars show ±1 SEM with between-subject variance removed (74).

Not surprisingly, effectiveness of CFS depends on stimulus-driven factors, such as size and contrast. The more signal there is, the harder it is to suppress. Indeed, using a very coarse measure of signal strength (the number of pixels per image), we find that hit rate is predicted by the number of pixels per image, $r = 0.43$, $b = 0.18$, $P = 0.006$. If verbal cues provide top-down support for detecting images, they may help override such purely stimulus-driven relationships. Indeed, as shown in Fig. 4*A*, the relationship between physical size and detection rates was affected by the cuing condition. Greater slopes indicate that detection rates increase more steeply with an increase in the size (in pixels) of the suppressed image. The relationship is strongest for the no-cue trials and decreases in the cuing conditions. This effect is clarified in Fig. 4*B*, showing that when preceded by valid labels, detection for images containing the fewest pixels (e.g., music stand, corkscrew, eyeglasses), which is very low in the no-cue condition, is increased by the greatest degree. That is, labels are more helpful for images with the weakest bottom-up signal.

In experiment 2, we drastically increased the number of stimuli, with each image appearing just once, making stimulus-based strategies ineffective. Verbal cues still affected the likelihood of seeing suppressed images. Although the 6-s stimulus duration improved overall performance, on almost 30% of the trials participants waited the full 6 s before responding, highlighting the depth of suppression that can occur with the present method. As shown in Fig. 4*A*, valid cues reduced the dependence of detection on physical properties such as size/density, which, without valid cues, are highly predictive of suppression efficacy.

Experiment 3. Experiments 1 and 2 demonstrated that labels affect detection of suppressed visual stimuli. To test whether this effect occurred because of labels activating visual features with which they are associated, in experiment 3, we parametrically varied the overlap between the shapes associated with the label

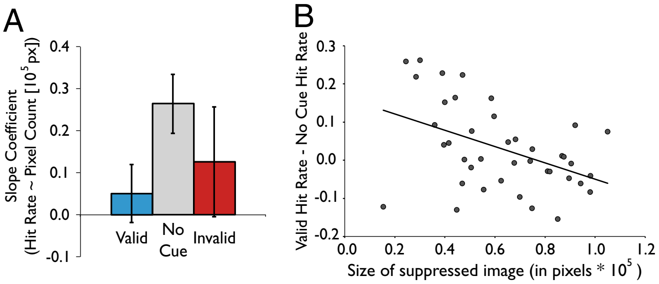

Fig. 4. (*A*) Relationship between image size (in pixels × 10^5) and image detection rates. (*B*) Plot showing the advantage of valid trials over no-cue trials as a function of image size. Errors bars show ±1 SEM coefficient.

and the shape actually shown, using the words "square" and "circle" and using visual stimuli that varied on a square-to-circle continuum.

Overall hit rates were 65.3% and false alarms 4.8%. Hit rates on label trials were reliably higher ($M = 0.68$) than on no-label trials ($M = 0.63$)*, $F(1,9) = 10.53$, $P = 0.01$, with no reliable differences observed on target-absent trials, $F(1,9) = 1.82$, $P = 0.21$. The critical question, however, was whether the effect of labels on stimulus detection reflected the degree of match between the label and shape. We hypothesized that labels would progressively improve visual detection as the correspondence between the label and the target's shape increased, and decrease detection rates as the label progressively mismatched the target. To test the hypothesis we conducted a general linear model (GLM) with hit rates as the dependent variable, cue ("square," "circle," noise) as a dependent factor, and the place of the target on the square-to-circle continuum as a covariate (11 levels). As predicted, there was a reliable interaction between cue and target, $F(2,18) = 7.70$, $P = 0.001$. As shown in Fig. 5*A*, hearing the word "square" tended to increase hit rates for the more square stimuli, whereas hearing "circle" tended to increase hit rates for the more circular shapes. (The interaction is reported for the full continuum using place on continuum as a covariate. The 11-shape continuum is collapsed into three categories in Fig. 5 for ease of interpretation.) Planned comparisons showed that hearing "square" helped to see the more square stimuli compared with hearing "circle," $t(9) = 3.29$, $P = 0.009$ [the comparison of "square" to noise trials was marginally significant, $t(9) = 2.1$, $P = 0.07$)]. Hearing "circle" helped detect the more circular stimuli both relative to hearing "square," $t(9) = 3.09$, $P = 0.01$, and hearing noise, $t(9) = 2.85$, $P = 0.02$. Consider now the intermediate stimuli. These shapes are not typical of circles or squares, but they share visual properties with both (e.g., they have intermediate curvature and angularity). Consistent with the prediction that labels activate visual features diagnostic of the denoted category, hearing either "square" or "circle" prior to seeing one of the intermediate shapes improved detection relative to the no-label noise trials, $F(1,9) = 6.29$, $P = 0.03$. As in experiments 1 and 2, all the effects were observed for hit rates with no effects on the (very low) false alarms.

A signal detection analysis mirrored the pattern of hits described above: detectability (d') showed a highly reliable interaction between cue and the shape of the to-be-detected target, $F(2,18) = 10.63$, $P < 0.0005$ (Fig. 5*B*). To obtain sufficient data from false-alarm trials for computing d', the data were collapsed into the bins shown in Fig. 5 before the signal-detection analysis. Cues had no effect on criterion, as measured by logβ, $F(2,18) = 1.20$, $P = 0.32$. An analysis of RTs revealed no evidence of a speed–accuracy tradeoff: slower responses correlated with poorer performance. In contrast to experiments 1 and 2, RTs were not affected by labeling, $M_{\text{noise}} = 1{,}340$ ms, $M_{\text{valid}} = 1{,}353$ ms, $M_{\text{invalid}} = 1{,}429$ ms, $F < 1$.

Using a parameterized shape continuum allowed us to test the hypothesis that the effects of language on visual awareness stem at least in part from the label activating shape information diagnostic of the category. The results showed that the effect of the verbal label on detection depended strongly on the match between shape information, which we hypothesize to be activated by the label (e.g., broad orientation tuning for circles) and the physical properties of the stimulus (see *SI Results and Discussion, Supplementary Analysis of Experiment 3 Detection Performance as a Function of Distribution of Orientations* and Figs. S1 and S2).

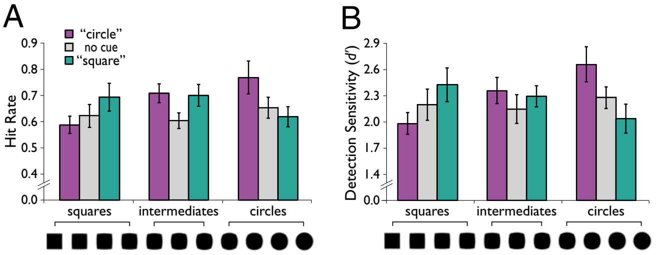

Fig. 5. (*A*) Hit rates and (*B*) detection sensitivity for detecting geometric shapes used in experiment 3 as a function of shape and presented label during a 6-s stimulus presentation. Error bars show ±1 SEM with between-subject variance removed (74).

Discussion

Our findings show that language affects detection of invisible objects. Hearing a valid verbal label helped participants become aware of the mere existence of objects that were clearly presented to one eye but suppressed from awareness via CFS. In contrast to valid labels, objects cued by invalid labels tended to be less likely to be detected. A possibility left open by experiment 1 was that labels encouraged stimulus-specific shifts of spatial attention. If true, increasing the variety of exemplars should have attenuated or eliminated the cuing effects, but experiment 2 showed that it did not. In experiment 3 we further demonstrated that the effects of labels on visual awareness are partially due to labels activating shape information diagnostic of the category denoted by the label. By parametrically varying the overlap in shape between the label and object, we showed that labels ("square" or "circle") aided detection as a function of the to-be-detected object's location on a square-to-circle continuum. Together these results suggest that labels affect performance not just on tasks requiring explicit identification, categorization, or discrimination of visual stimuli (e.g., refs. 22 and 25) but can boost into awareness the presence of an object through a hypothesized top-down activation of visual properties diagnostic of the object's category.

The mechanism by which verbal labels affect visual sensitivity for the named objects may involve a top-down retuning, similar to the now-familiar findings of top-down attentional and task effects throughout the visual hierarchy, from lateral geniculate nucleus onward (refs. 41 and 42 for review). Although the neural loci of the present effects are not known, one possibility is that processing an object name initiates feedback activity to object-selective regions of cortex such as inferotemporal cortex, producing a predictive signal or "head start" to the visual system (9, 43, 44). On several theories of attention (e.g., biased competition theory of ref. 45), these predictive signals would enable neurons that respond to the named object to gain a competitive advantage (see also refs. 46 and 47). Given feedback from object-selective cortical regions, winning objects can bias earlier regions of visual cortex. A weak bottom-up signal combined with the top-down signal produced by the label may be sufficient to propel the percept into awareness.

A possible alternative interpretation is that the labels affect performance at a postperceptual semantic/decision level. On that account, hearing "zebra" would activate an amodal semantic representation that is also accessed by the visually presented (but suppressed) image of a zebra, and the detection decision is made on the basis of this amodal activation. This alternative depends on the suppressed image receiving substantial semantic processing. However, existing evidence suggests that CFS (and interocular suppression more broadly) seems to interfere with visual processes occurring before semantic analysis of words (48, 49) and objects (50, 51); for example, interocularly suppressed words and objects cannot prime subsequent processing of related stimuli (52), and neurons in medial temporal cortex do not respond to images suppressed through CFS in humans (53) or suppressed through binocular rivalry in monkeys (54).

---

*This hit rate is substantially higher than the 50% QUEST threshold to which the staircasing procedure was set because participants' performance tends to improve during the course of the experiment. The mean hit-rate for no-label trials during the start of the experimental session (first 20 trials) was not significantly different from 50%, $t < 1$.

The sum of the evidence points to interocular suppression affecting temporally and anatomically early processing (53–57). Although several findings indicate that certain visual properties of stimuli suppressed from awareness can nevertheless guide spatial attention (58, 59), there is no clear evidence that stimuli suppressed through CFS are processed semantically. We discuss some caveats to this conclusion in the *SI Results and Discussion, Semantic Processing of Images Made Invisible by Continuous Flash Suppression*.

If the suppressed image does not reach semantic analysis, it is unclear how its detection can be facilitated at this high level. We therefore argue that the facilitation due to labels occurs at a lower, perceptual level (although our results leave open the question of where precisely in the visual system this interaction occurs). One possibility is that the verbal cue activates visual features typical/diagnostic of the cued category (ref. 13 for discussion). The combination of bottom-up and top-down information becomes sufficient to propel the sensory input into awareness. This mechanism can be thought of as a kind of category-based attention (23).[†]

In the case of simple geometric categories like squares and circle, the tuning may be to orientation profiles. In the case of categories like eyeglasses, the diagnostic features are more complex (ref. 61 for a somewhat similar proposal).

Although we believe the present results suggest that labels affect aspects of visual awareness by modulating perceptual processes, an alternative is that linguistic cues direct attention to a relatively late object representation that drives object-decision responses. On this account, differences in reportability—the currently accepted measure of explicit awareness (62)—may be driven by differences in high-level activation without a corresponding change in subjective awareness. We cannot rule this out, but if true, it is unclear how subjective awareness of perceptual states could be measured.

Putting aside the anatomical locus of CFS, the present results show that language and vision interact to affect behavior in a basic visual task—simple detection. If CFS disrupts visual processing at an anatomically or temporally early stage and prevents higher-level/semantic processing, then our finding that verbal labels can help overcome this disruption can be taken to indicate that verbal labels modulate low-level visual processing. If CFS disruptions leave initial processing unaffected, then our findings indicate that this early processing, although necessary, is insufficient for visual awareness of stimuli suppressed through CFS.

It may seem maladaptive for vision to be so sensitive to input from outside its "domain" (1, 11). On such modular formulations, the purpose of vision is to construct objective models of the world—a process that would be disturbed if it could be influenced by language (or other high-level factors, such as desires or expectations). However, if we consider that the real purpose of perceptual systems is to help guide behavior according to incomplete and underdetermined inputs, and that perception is at its core an inferential process (5, 9, 63–69), then perception needs all of the help it can get. If tuning the visual system can make it more sensitive to a class of stimuli or a perceptual dimension that is currently task-relevant, then having a highly permeable perceptual system that allows for influences outside vision, including language, can be viewed as highly adaptive. Indeed it is perhaps this power of language to modulate processing on demand—from perception onward—that makes it so effective in guiding behavior.

Methods

Stimuli. For experiment 1, we selected eight gray-scale line-drawings from Rossion and Pourtois (70, see Table S1) on the basis of their high imageability (Table S1 lists stimuli from experiments 1 and 2). CFS was implemented using anaglyph images (e.g., ref. 71): participants wore red/cyan glasses and viewed stereograms containing a high-contrast red mask (~9° × 9°) and—on object-present trials—a superimposed lower-contrast cyan object (Fig. 1*A*). Only the object was visible to the right eye and only the mask to the left. The dynamic mask comprised curved line segments, with frames randomly alternating at 10 Hz. Because similarity in spatial properties between stimuli and masks is important for effective suppression of stimuli (72), line segments were used to better mask the curvilinear character of the objects. Compared with the more commonly used Mondrian rectangle masks, our line-segment masks were more similar to the objects in terms of image directionality profiles (computed using Fourier component analysis using Fiji). The root mean squared (RMS) difference in orientation profiles was lower between the line-segment masks and the line-drawn stimuli, RMS = 0.033, than between rectangle masks and stimuli, RMS = 0.044, $t(7) = 7.20$, $P < 0.001$, indicating that the line-segment masks were a better match to the images than the rectangular Mondrians [additional manipulation checks described in *SI Results and Discussion, Additional Experiments (Manipulation Check)*]. Auditory cues were volume-normalized object names recorded by a native English-speaker, with average duration ~650 ms. A white-noise cue with the same duration replaced a verbal cue for the baseline condition (e.g., ref. 22).

Experiment 2 was identical to experiment 1 except as noted. Images were grayscale photographs from 40 object categories [e.g., "eyeglasses," "mailbox," "tractor" (73) (see Table S1)]. There were four exemplars per category (160 separate images), and each image appeared only once. Compared with rectangular Mondrian masks, the line-segment masks were again a better match to the stimuli: $\text{RMS}_{\text{line-segment masks}} = 0.024$, $\text{RMS}_{\text{rectangle masks}} = 0.027$, $t(159) = 4.71$, $P < 0.001$. The mask/image combinations were now presented for 6 s, rather than 1.5 s.

In experiment 3, rather than using pictures of objects and animals, the to-be-detected stimuli comprised parametrically varying geometric shapes on a continuum from a square to a circle. We used 11 shapes from the continuum, spaced in 10% increments with the endpoints comprising a perfect square and circle, respectively (Fig. 5). All shapes had equal areas and subtended ~1.8° × 1.8°. The auditory labels were the words "square," "circle," and a segment of white-noise. All sounds were length- and volume-normalized.

Procedure. ***Experiments 1 and 2.*** For experiment 1, 20 participants with normal or corrected-to-normal vision were recruited from University of Pennsylvania in exchange for course credit or payment. Twenty different students from University of Pennsylvania and University of Wisconsin-Madison (UW Madison) participated in experiment 2.

We used a 2 (object presence) × 3 (cue type) within-subject design. Suppressed images were presented on a random 50% of the trials (object-present trials); the remaining 50% contained only the masks (object-absent trials). Half of all trials were preceded by verbal labels, which matched the object on object-present trials 80% of the time. This design ensured that labels were not predictive of object presence. Each participant completed 320 trials.

Each trial began with a 500-ms fixation cross followed by a cue (label or noise), and 450 ms later the stimulus (mask or object + mask) was displayed for up to 1.5 s in experiment 1 and 6 s in experiment 2, although participants were instructed to respond as soon as they detected an any object within the mask or were sure that an object was absent. If no response was received within the time allotted, an object-presence prompt was shown until a response was received (Fig. 1*B*). If participants thought an object had been present, a follow-up question asked what object they thought they saw (e.g., "Was the object you saw a pumpkin?"). The label in the validity prompt always matched the cue (if there had been a label) but was uncorrelated with whether any object was actually presented. These validity prompts are not the focus of the analysis but do allow us to look at the effects of labels on recognition bias.

Experiment 3. Ten undergraduate psychology students from UW Madison participated for course credit. One participant was excluded for having an excessively high false alarm rate (2.75 SDs above the mean) and was replaced. The design was analogous to that of experiments 1 and 2, with several key differences. Instead of using anaglyph glasses, we used liquid crystal display active shutter glasses (Nvidia 3D vision) such that the masks were shown only to the left eye and the to-be-detected shapes to the right eye. Rather than setting stimuli to the same level of contrast for each participant, the contrast was estimated using a short (66-trial) QUEST staircasing procedure for each participant at the start of the experiment. Each hit resulted in the contrast of the to-be-detected shape decreasing slightly; misses resulted in a slight increase to the contrast such that the hit-rate converged on 50% (Weibull function, $\delta = 0.01$, $\gamma = 0.03$, linear steps). Each trial began with a fixation cross for 800–1,000 ms, followed by flashing masks, and on target-present trials, a random shape from the square-to-circle continuum. The stimulus/mask display was present up to 6 s, as in experiment 2. Because we were primarily interested in the target-present trials (recall that there were no theoretically

---

[†] Although Pylyshyn (1) explicitly excludes attentional effects as instance of cognitive penetration of perception, contemporary theories of attention—both spatial and feature-based—point to attention itself as modulation of perceptual processes (60).

interesting effects on target-absent trials), the proportion of target-present trials was increased to 67%. Unlike experiments 1 and 2, there was no follow-up question after "present" responses asking to confirm what the subjects saw.

After this staircasing procedure, participants were informed that they would hear words or other sounds at the start of each trial. The experimenter emphasized that the words/sounds were not predictive of whether there is a shape presented in the subsequent display. Stimulus contrast for all stimuli in the experiment was set to the final estimate produced by the staircasing procedure. Labels were presented on 50% of the trials and were evenly split between the words "circle" and "square," resulting in a perfectly counterbalanced design. As in experiments 1 and 2, the word cues were completely uncorrelated with stimulus-presence. The words were now also completely uncorrelated with stimulus type: participants were equally likely to hear the words "square," "circle," or a noise sound before any of the shapes along the square-to-circle continuum. Each participant completed 264 experimental trials.

1. Pylyshyn Z (1999) Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav Brain Sci* 22(3):341–365, discussion 366–423.
2. Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14(1):147–152.
3. Watanabe K, Shimojo S (2001) When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychol Sci* 12(2):109–116.
4. Freeman WJ (2001) *How Brains Make Up Their Minds* (Columbia Univ Press, New York), 1st Ed.
5. Mesulam MM (1998) From sensation to cognition. *Brain* 121(Pt 6):1013–1052.
6. McClelland JL, Rumelhart DE (1981) An interactive activation model of context effects in letter perception. 1. An account of basic findings. *Psychol Rev* 88(5):375–407.
7. Churchland PS, Ramachandran V, Sejnowski TJ (1994) *Large-Scale Neuronal Theories of the Brain*, eds Koch C, Davis JL (MIT Press, Cambridge, MA), pp 23–60.
8. Cox D, Meyers E, Sinha P (2004) Contextually evoked object-specific responses in human visual cortex. *Science* 304(5667):115–117.
9. Kveraga K, Ghuman AS, Bar M (2007) Top-down predictions in the cognitive brain. *Brain Cogn* 65(2):145–168.
10. McClelland JL, Mirman D, Holt LL (2006) Are there interactive processes in speech perception? *Trends Cogn Sci* 10(8):363–369.
11. Fodor JA (1983) *The Modularity of Mind* (MIT Press, Cambridge, MA).
12. Anderson SE, Chiu E, Huette S, Spivey MJ (2011) On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychol (Amst)* 137(2):181–189.
13. Lupyan G (2012) Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front Pyschol* 3:54.
14. Shams L, Kim R (2010) Crossmodal influences on visual perception. *Phys Life Rev* 7(3): 269–284.
15. Lupyan G, Thompson-Schill SL, Swingley D (2010) Conceptual penetration of visual processing. *Psychol Sci* 21(5):682–691.
16. Macpherson F (2012) Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philos Phenomenol Res* 84(1):24–62.
17. Stokes D (2011) Perceiving and desiring: A new look at the cognitive penetrability of experience. *Philos Stud* 158(3):479–492.
18. Siegel S (2012) Cognitive penetrability and perceptual justification. *Noûs* 46(2):201–222.
19. Winawer J, et al. (2007) Russian blues reveal effects of language on color discrimination. *Proc Natl Acad Sci USA* 104(19):7780–7785.
20. Regier T, Kay P (2009) Language, thought, and color: Whorf was half right. *Trends Cogn Sci* 13(10):439–446.
21. Athanasopoulos P, Wiggett A, Dering B, Kuipers JR, Thierry G (2009) The Whorfian mind: Electrophysiological evidence that language shapes perception. *Commun Integr Biol* 2(4):332–334.
22. Lupyan G (2008) The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition* 108(2):566–577.
23. Lupyan G, Spivey MJ (2010) Redundant spoken labels facilitate perception of multiple items. *Atten Percept Psychophys* 72(8):2236–2253.
24. Lupyan G, Swingley D (2012) Self-directed speech affects visual search performance. *Q J Exp Psychol (Hove)* 65(6):1068–1085.
25. Meteyard L, Bahrami B, Vigliocco G (2007) Motion detection and motion verbs: Language affects low-level visual perception. *Psychol Sci* 18(11):1007–1013.
26. Dils AT, Boroditsky L (2010) Visual motion aftereffect from understanding motion language. *Proc Natl Acad Sci USA* 107(37):16396–16400.
27. Meteyard L, Zokaei N, Bahrami B, Vigliocco G (2008) Visual motion interferes with lexical decision on motion words. *Curr Biol* 18(17):R732–R733.
28. Pelekanos V, Moutoussis K (2011) The effect of language on visual contrast sensitivity. *Perception* 40(12):1402–1412.
29. Landau AN, Aziz-Zadeh L, Ivry RB (2010) The influence of language on perception: Listening to sentences about faces affects the perception of faces. *J Neurosci* 30(45): 15254–15261.
30. Raftopoulos A (2005) *Cognitive Penetrability Of Perception: Attention, Action, Strategies, And Bottom-up Constraints* (Nova, Hauppauge, NY).
31. Fodor JA (1988) A reply to Churchland's 'Perceptual plasticity and theoretical neutrality' *Philos Sci* 55(2):188–198.
32. Raftopoulos A (2011) Late vision: Processes and epistemic status. *Front Psychol* 2:382.
33. Lupyan G, Spivey MJ (2010) Making the invisible visible: Auditory cues facilitate visual object detection. *PLoS ONE* 5:e11452.
34. Bar M, et al. (2006) Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA* 103(2):449–454.
35. Fahrenfort JJ, Scholte HS, Lamme VAF (2007) Masking disrupts reentrant processing in human visual cortex. *J Cogn Neurosci* 19(9):1488–1497.
36. Frenck-Mestre C, Grainger J (1998) Masked priming by translation equivalents in proficient bilinguals. *Lang Cogn Process* 13(6):601–623.
37. Kouider S, Dehaene S (2007) Levels of processing during non-conscious perception: A critical review of visual masking. *Philos Trans R Soc Lond B Biol Sci* 362(1481):857–875.
38. Blake R (2001) A primer on binocular rivalry, including current controversies. *Brain Mind* 2(1):5–38.
39. Tsuchiya N, Koch C (2005) Continuous flash suppression reduces negative afterimages. *Nat Neurosci* 8(8):1096–1101.
40. Weiskrantz L (1999) *Consciousness Lost and Found: A Neuropsychological Exploration* (Oxford Univ Press, New York), 1st Ed.
41. O'Connor DH, Fukui MM, Pinsk MA, Kastner S (2002) Attention modulates responses in the human lateral geniculate nucleus. *Nat Neurosci* 5(11):1203–1209.
42. Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61(2): 168–185.
43. Puri AM, Wojciulik E (2008) Expectation both helps and hinders object perception. *Vision Res* 48(4):589–597.
44. Esterman M, Yantis S (2010) Perceptual expectation evokes category-selective cortical activity. *Cereb Cortex* 20(5):1245–1253.
45. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222.
46. Deco G, Lee TS (2002) A unified model of spatial and object attention based on intercortical biased competition. *Neurocomputing* 44:775–781.
47. Vecera SP, Farah MJ (1994) Does visual attention select objects or locations? *J Exp Psychol Gen* 123(2):146–160.
48. Kang MS, Blake R, Woodman GF (2011) Semantic analysis does not occur in the absence of awareness induced by interocular suppression. *J Neurosci* 31(38):13535–13545.
49. Zimba LD, Blake R (1983) Binocular rivalry and semantic processing: Out of sight, out of mind. *J Exp Psychol Hum Percept Perform* 9(5):807–815.
50. Cave CB, Blake R, McNamara TP (1998) Binocular rivalry disrupts visual priming. *Psychol Sci* 9(4):299–302.
51. Moradi F, Koch C, Shimojo S (2005) Face adaptation depends on seeing the face. *Neuron* 45(1):169–175.
52. Lin Z, He S (2009) Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Prog Neurobiol* 87(4):195–211.
53. Kreiman G, Fried I, Koch C (2002) Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc Natl Acad Sci USA* 99(12):8378–8383.
54. Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci USA* 94(7):3408–3413.
55. Blake R, Logothetis NK (2002) Visual competition. *Nat Rev Neurosci* 3(1):13–21.
56. Pasley BN, Mayes LC, Schultz RT (2004) Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* 42(1):163–172.
57. Fang F, He S (2005) Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat Neurosci* 8(10):1380–1385.
58. Hsieh PJ, Colas JT, Kanwisher N (2011) Pop-out without awareness: Unseen feature singletons capture attention only when top-down attention is available. *Psychol Sci* 22(9):1220–1226.
59. Chou WL, Yeh SL (2012) Object-based attention occurs regardless of object awareness. *Psychon Bull Rev* 19(2):225–231.
60. Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14(9):744–751.
61. Schyns PG, Oliva A (1999) Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* 69(3):243–265.
62. Dehaene S, Changeux JP, Naccache L, Sackur J, Sergent C (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn Sci* 10(5): 204–211.
63. Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36(3):181–204.
64. Geisler WS, Kersten D (2002) Illusions, perception and Bayes. *Nat Neurosci* 5(6): 508–510.
65. Howe CQ, Purves D (2002) Range image statistics can explain the anomalous perception of length. *Proc Natl Acad Sci USA* 99(20):13184–13188.
66. Koivisto M, Railo H, Revonsuo A, Vanni S, Salminen-Vaparanta N (2011) Recurrent processing in V1/V2 contributes to categorization of natural scenes. *J Neurosci* 31(7): 2488–2492.
67. Lamme VAF, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 23(11):571–579.
68. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2(1):79–87.
69. Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647.
70. 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(2):217–236.
71. Almeida J, Mahon BZ, Nakayama K, Caramazza A (2008) Unconscious processing dissociates along categorical lines. *Proc Natl Acad Sci USA* 105(39):15214–15218.
72. Yang E, Blake R (2012) Deconstructing continuous flash suppression. *J Vis* 12(3):8.
73. Brady TF, Konkle T, Alvarez GA, Oliva A (2008) Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci USA* 105(38):14325–14329.
74. Morey RD (2008) Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials Quantitative Methods Psychol* 4(2):61–64.