

# The Evocative Power of Words: Activation of Concepts by Verbal and Nonverbal Means

Gary Lupyan  
University of Wisconsin—Madison

Sharon L. Thompson-Schill  
University of Pennsylvania

A major part of learning a language is learning to map spoken words onto objects in the environment. An open question is what are the consequences of this learning for cognition and perception? Here, we present a series of experiments that examine effects of verbal labels on the activation of conceptual information as measured through picture verification tasks. We find that verbal cues, such as the word “cat,” lead to faster and more accurate verification of congruent objects and rejection of incongruent objects than do either nonverbal cues, such as the sound of a cat meowing, or words that do not directly refer to the object, such as the word “meowing.” This label advantage does not arise from verbal labels being more familiar or easier to process than other cues, and it does extend to newly learned labels and sounds. Despite having equivalent facility in learning associations between novel objects and labels or sounds, conceptual information is activated more effectively through verbal means than through nonverbal means. Thus, rather than simply accessing nonverbal concepts, language activates aspects of a conceptual representation in a particularly effective way. We offer preliminary support that representations activated via verbal means are more categorical and show greater consistency between subjects. These results inform the understanding of how human cognition is shaped by language and hint at effects that different patterns of naming can have on conceptual structure.

*Keywords:* concepts, labels, words, representations, language and thought

*Supplemental materials:* <http://dx.doi.org/10.1037/a0024904.supp>

Two hallmarks of human development are developing conceptual categories—learning that things with feathers tend to fly, that animals possessing certain features are dogs, and that foods of a certain color and shape are edible (Carey, 1987; Keil, 1992; Rogers & McClelland, 2004)—and learning *names* for these categories (Waxman, 2004). Although many have commented on the transformative power of names (Clark, 1998; Dennett, 1996; Harnad, 2005; James, 1890; Vygotsky, 1962), it is only recently that the interplay between verbal labels and concepts is becoming a subject of systematic empirical study. Given the tight linkage between the representations of verbal

meanings and the larger conceptual system (Murphy, 2002), an important question is what effects language learning has on the activation and the organization of putatively nonverbal representations.

The learning of categories is, in principle, separable from the learning of their names. A child can have a conceptual category of “dog” without having a verbal label associated with the category. However, in practice, the two processes are intimately linked. Not only does conceptual development shape linguistic development (e.g., Snedeker & Gleitman, 2004), but linguistic development—and in particular, learning words—impacts conceptual development (e.g., Casasola, 2005; Gentner & Goldin-Meadow, 2003; Gumperz & Levinson, 1996; Levinson, 1997; Lupyan, Rakison, & McClelland, 2007; Spelke, 2003; Spelke & Tsivkin, 2001; Waxman & Markow, 1995; Yoshida & Smith, 2005). For example, Casasola (2005) found that 18-month old infants could form an abstract spatial category only when accompanied by a familiar word.

Words continue to impact category learning in adulthood. For example, Lupyan, Rakison, and McClelland (2007) showed that learning verbal labels for novel categories improved category learning, even though the labels were entirely redundant. Once a word is learned, it appears to exert influences on visual recognition memory (Lupyan, 2008b) as well as perceptual processing (Lupyan, 2008a; Winawer et al., 2007; see Gliga, Volein, & Csibra, 2010, for intriguing results with 1-year-old infants). For example, hearing a verbal label such as “chair” facilitates the visual processing of the named category, compared with trials on which participants know the relevant object category but do not actually hear its name (Lupyan, 2007a, 2007b, 2008a). Hearing a

---

This article was published Online First September 19, 2011.

*Editor's Note.* Gregory L. Murphy served as the action editor for this article.—IG

---

Gary Lupyan, Department of Psychology, University of Wisconsin—Madison; Sharon L. Thompson-Schill, Institute for Research in Cognitive Science, Center for Cognitive Neuroscience, Department of Psychology, University of Pennsylvania.

This work was supported by an Integrative Graduate Education and Research Traineeship training grant to Gary Lupyan and by National Institutes of Health Grant R01DC009209 and R01MH70850 to Sharon L. Thompson-Schill. We thank Nina Hsu for designing the stimuli used in Experiment 4 and Joyce Shin, Ali Shapiro, and Arber Tasimi for help with data collection.

Correspondence concerning this article should be addressed to Gary Lupyan, 419 Brogden Hall, Department of Psychology, University of Wisconsin, Madison, WI 53706. E-mail: [lupyan@wisc.edu](mailto:lupyan@wisc.edu)

label can even make an invisible object visible. Lupyán and Spivey (2010a) showed that hearing a spoken label increased visual sensitivity (i.e., increased the  $d'$ ) in a simple object detection task: Simply hearing a label enabled participants to detect the presence of briefly presented masked objects that were otherwise invisible (see also Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, Experiment 5).

Understanding a word requires activation of conceptual representations denoted by that word. Of course, activation of concepts occurs in nonlinguistic contexts as well. It therefore makes sense to ask: Are conceptual representations activated by words different in some way from those activated via nonverbal means? Do words simply offer a way to access a language-independent concept—a concept that can be accessed equivalently through other, nonverbal means, or do words activate conceptual representations in a special way?

Before proceeding, we provide a short definition of what we mean by the terms concept and category. For present purposes, we define a concept more narrowly, as the mental representation of a category. A category in turn is a collection of discriminable entities that are treated as equivalent in some way (Bruner, Austin, & Goodnow, 1956; Harnad, 2005). So, for example, the category of chairs forms a collection of discriminable entities that are equivalent in the certain contexts, such as finding something to sit on or something denoted by the word “chair.”

We focus here on the visual aspects of conceptual representations and compare the power of verbal and nonverbal cues to activate visual information of both familiar and novel categories—information we believe to be constitutive though clearly not exhaustive of the concept (Murphy & Medin, 1985).

### The Logic of the Present Studies: Activation of Visual Information by Verbal and Nonverbal Means

A response to a visual stimulus can be altered by a cue presented prior to the target stimulus. These cues can be nonverbal (Egly, Driver, & Rafal, 1994; Eriksen & Hoffman, 1972; Posner, Snyder, & Davidson, 1980) as well as verbal. For example, verbal cues in the form of words like “left” and “right” produce automatic shifts of attention just as reliably as nonverbal cues such as directional arrows, even when the words are entirely nonpredictive of the target’s location (e.g., Hommel, Pratt, Colzato, & Godijn, 2001). Words related to motion, for example, “float,” have been shown to affect visual motion processing, changing the sensitivity in detecting motion in random-dot kinematograms (Meteyard, Bahrami, & Vigliocco, 2007). A number of studies have also shown that visual object processing and attentional guidance can be altered by verbal cues (Pertzov, Zohary, & Avidan, 2009; Puri & Wojciulik, 2008; Schmidt & Zelinsky, 2009; Vickery, King, & Jiang, 2005; Walter & Dassonville, 2005; Yang & Zelinsky, 2009). Such effects of cues on visual processing have been linked to increases in category-specific cortical activity. For example, seeing the word *face* increases activity in the fusiform face area, an increase that correlates with an improvement in making a gender judgment of faces embedded in visual noise (Esterman & Yantis, 2010). In this article, we take a sensorimotor view of concepts: The neural activity that composes a concept is multimodal; that is, the visual aspects of the concept are represented by some of the same structures as those involved in sensory processing of the modality

(e.g., Allport, 1985; Barsalou, 2008; Thompson-Schill, Aguirre, D’Esposito, & Farah, 1999).

Is a verbal label merely a convenient method of communicating information, or is there something special in the effect of a verbal cue on the conceptual/perceptual information that is activated? To make this question more concrete: Are the representations activated when hearing the word “cow” different from those activated when hearing nonverbal cues (one that is similarly associated with the concept of cows, for example, a mooing sound)? Although both *cow* and the sound of a cow mooing are associated with cows, only the former is treated (in the normal course of things) as referring to a cow.

We present seven experiments in which we examined whether concepts evoked by verbal and nonverbal means are distinguishable. In particular, we focus on the visual aspects of concepts activated by verbal and nonverbal means. Experiments 1A–1C and Experiment 2 contrasted the effects of verbal and nonverbal cues on performance in picture-verification tasks. Experiments 3A–3B contrasted verbal and nonverbal cues in a visual discrimination task that requires minimal semantic processing of the target pictures. In Experiment 4, we controlled participants’ exposure to verbal and nonverbal cues by teaching them to associate novel labels and nonverbal sounds with novel object categories. This allowed us to test whether the results observed in Experiments 1 and 2 arose because participants were more familiar with verbal cues (e.g., “cow”) than the nonverbal cues (e.g., mooing sound) or whether verbal cues indeed produced a unique effect on conceptual activation visual processing, perhaps owing to their referential status.

### Experiments 1A–1C

A simple way to compare the relative efficacy with which verbal and nonverbal cues activated conceptual representations is through a verification task. In our implementation of this task, participants hear a cue that is either a word (e.g., “cow”) or a characteristic sound (e.g., a mooing sound) and then see a matching or mismatching picture, which remains on the screen until the participants respond “match” or “mismatch.” The more effective a cue is, the more quickly and/or accurately participants can respond to the target picture. If verbal and nonverbal cues both activate the very same concept cow (put into neural terms, the same assembly of neurons) and do so equally fast, then verification performance should be equivalent in the two cuing conditions.<sup>1</sup> A second possibility is that the two cues result in the same conceptual activation at the limit but that one cue leads to faster activation than the other. A third possibility is that verbal and nonverbal cues lead to qualitatively distinct patterns of activation. That is, rather than being two routes to activating the same concept, concepts activated via verbal means are different in some way from concepts activated via nonverbal means.

<sup>1</sup> Note that the converse is not necessarily true. On a distributed, sensorimotor theory of concepts (Allport, 1985; Martin, Ungerleider, & Haxby, 2000), the visual features are partly constitutive of the “conceptual” representation. A finding that different cues produce equal performance in picture verification may mean that the different cues activate nonoverlapping or partly overlapping representations that are both equally adequate for making a verification judgment.

One way to tease apart the second and third possibilities is by varying the delay between the cue and the target. If the difference between the two cues is just a difference in the speed of activation then it should diminish with longer delays, as the slower cue is allowed time to “catch up” with the faster cue. Varying the delay also allows us to test for the possibility that people may process one cue type more quickly than the other, for example, words may be processed more quickly because they are more familiar. Differences in verification performance for short delays may thus reflect an incomplete processing of the cue rather than a genuine difference in activation produced by the cue. For longer delays, however, verification time should reflect conceptual activation produced by the cue (e.g., Murphy & Brownell, 1985; Stadthagen-Gonzalez, Damian, Pérez, Bowers, & Marín, 2009; Yuval-Greenberg, & Deouell, 2009, for similar reasoning).

If conceptual representations activated by verbal and nonverbal cues are genuinely different, then difference in verification performance should persist, even for longer delays. The cue difference should be observed for both matching and mismatching trials because the information activated by the cue is useful both for accepting a match object and for rejecting a mismatch.

## Method

**Participants.** A total of 43 University of Pennsylvania undergraduates volunteered in the experiments in exchange for course credit: 18 in Experiment 1A, 15 in Experiment 1B, and 10 in Experiment 1C.

**Materials.** We selected 10 objects that were easily nameable and had characteristic sounds (cat, car, dog, frog, gun, motorcycle, rooster, train, cow, and whistle). Each category was instantiated by five images: normed color drawings (Rossion & Pourtois, 2004), three photographs obtained from online image collections, and one cartoon image (see Figures S1 and S2 in the supplemental materials). We used several instances of each category to introduce visual heterogeneity. Spoken labels were all basic-level names. The nonverbal cues were animal sounds for the animals in the set and characteristic sounds for the artifacts in the set (e.g., a gun firing, the sound of a whistle). These sounds were obtained from online environmental sound libraries and are available for download (see Appendix).

All auditory stimulus sounds were volume normalized. We also equated the length of the label and sound cues for each category (i.e., the barking sound and the word “dog” were of identical durations). Two of the nonverbal sounds—the sound of a starting car, and the sound of a train—were difficult to recognize when presented at durations that matched the words “car” and “train.” We therefore replaced the label cues for these categories with the longer (but less common) labels “automobile” and “locomotive,” respectively (in Experiment 2, we used revised car and train, enabling the words “car” and “train” to be matched for length, and obtained results similar to the present studies).

In order to ensure that the sounds were (a) easily recognizable and (b) of comparable predictive power of the target category, we conducted two norming experiments that are described in detail in the Appendix. These results indicated that (with the possible exception of one item) the sound cues were easily recognizable and were of comparable predictive power as the cues.

**Procedure.** On each trial, participants heard a cue—a verbal label or a nonverbal sound—followed by a picture. The picture matched the cue 50% of the time. On the nonmatching trials, the picture was randomly selected among the nonmatching images. Participants responded by pressing a *match* or *does not match* key on a keyboard. Immediately following their response, auditory feedback in the form of a buzz or bleep indicated whether the response was correct. All factors were within-subjects and each participant completed 400 verification trials: 10 Categories  $\times$  5 Category Exemplars  $\times$  2 Levels of Congruence  $\times$  2 Cue-Types (sound vs. label)  $\times$  2 Repeats.

Experiments 1A–1C differed in just one respect: In Experiment 1A, the delay between cue offset and target picture onset was 400 ms. We refer to this as the ISI (interstimulus interval) from here onward. Note that because of differences in cue length between categories, it is timed from cue *offset* to picture onset. In Experiment 1B, it was increased to 1 s—a common delay used in verification tasks (Stadthagen-Gonzalez et al., 2009). In Experiment 1C, the delay was increased further to 1.5 s. By repeating Experiment 1A with a longer delay, we could test whether any label advantage found in Experiment 1A arose from incomplete processing of the cue. By increasing the delay, we could ensure that the cue was sufficiently processed by the time the picture appeared. Thus, we could be sure that the verification reaction times (RTs) (the principal dependent measure) were determined by the time it took to recognize the picture rather than reflecting the residual processing of the label or sound cue (Murphy & Brownell, 1985; Stadthagen-Gonzalez et al., 2009).

## Results

Mean latencies for Experiments 1A–1C are shown in Figures 1A–1C; latency and accuracy means are also presented in Table 1. The data were analyzed with a 2 (label or sound)  $\times$  2 (match or mismatch) within-subjects analysis of variance (ANOVA). RTs less than 200 ms or greater than 1,500 ms were excluded. An analysis of correct RTs from Experiment 1A revealed a highly reliable matching advantage,  $F(1, 17) = 35.72, p < .0005$ , and a strong advantage for label trials,  $F(1, 17) = 24.77, p < .0005$ . In a subsequent analysis, we added category as a fixed factor and observed a highly reliable Cue-Type  $\times$  Picture Category interaction for RTs,  $F(9, 153) = 2.7, p = .005$ . That is, responses to all items were facilitated by the label, relative to nonverbal sound cue, but to different degrees. We explore this in more detail below.

The label advantage was also observed in accuracy,  $F(1, 17) = 6.38, p = .02$ . The label advantage was highly consistent. Of the 18 participants, 16 had shorter RTs on the label than on sound trials, and 11 had higher accuracy (of the remaining seven, three had equal performance in the two conditions). The label advantage was also reliable in an item-based analysis—RTs:  $F_2(1, 9) = 5.89, p = .038$ , accuracy:  $F_2(1, 9) = 8.93, p = .015$ .

Experiment 1B likewise revealed a match advantage for RTs,  $F(1, 14) = 20.80, p < .0005$ , and a strong label advantage,  $F(1, 14) = 26.80, p < .0005$ . Every participant showed this label advantage. The label advantage was also observed in accuracy,  $F(1, 14) = 13.11, p = .003$ . There were no significant Cue-Type  $\times$  Picture Category interactions for RTs or accuracy ( $F_s < 1$ ).

Increasing the delay further to 1.5 s—Experiment 1C—produced a similar pattern of results. There was again a match

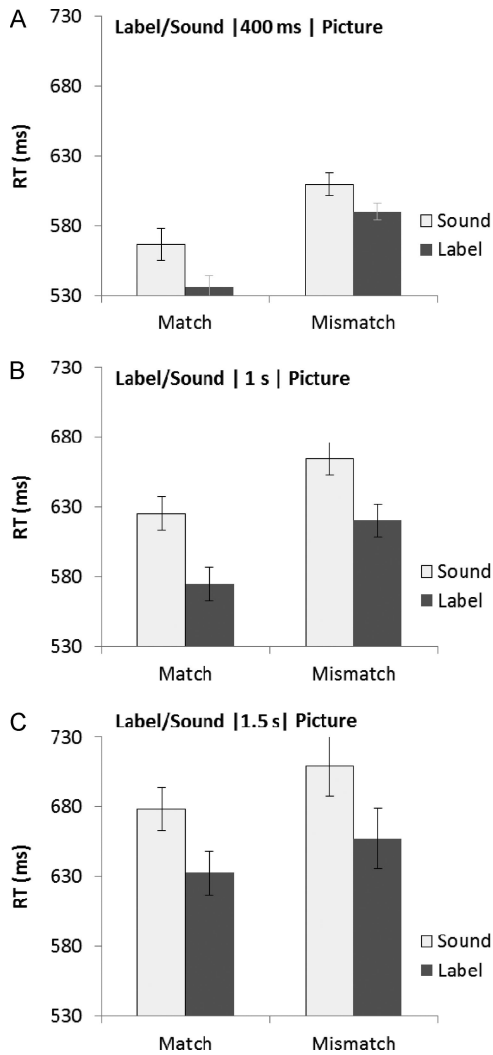


Figure 1. Verification times for the sound trials versus label trials. A: Experiment 1A. B: Experiment 1B. C: Experiment 1C. The auditory cue and the picture matched on match trials and mismatched on mismatch trials. Error bars show  $\pm 1$  standard error of the mean difference between label and sound conditions. RT = reaction time.

advantage in RTs,  $F(1, 9) = 7.56, p = .022$ , and a strong label advantage for both RTs,  $F(1, 9) = 7.66, p = .022$ , and accuracy,  $F(1, 9) = 62.61, p < .0005$ . Seven out of the 10 participants showed the label advantage in latency, and all 10, in accuracy. There was a marginal Cue-Type  $\times$  Picture Category interaction,  $F(9, 81) = 1.90, p = .064$ , in the same direction as in Experiment 1A (see Experiments 1A–1C, Discussion, for clarification). The label advantage remained highly significant in an item-based analysis—RTs:  $F_2(1, 9) = 39.26, p < .0005$ , accuracy:  $F_2(1, 9) = 25.49, p = .001$ . The RTs in this experiment were nonsignificantly longer than in Experiments 1A–1B, probably due to the greater uncertainty of target onset owing to the longer ISI.

The label advantage observed for the shortest ISI became larger for the longer ISIs, increasing from 25 ms to 47 ms and 49 ms for the 1 s and 1.5 s ISIs, respectively. There was no difference between the label advantages for the two longer ISIs ( $t < 1$ ). We

therefore pooled these data and compared the label advantage between the shortest ISI (Experiment 1A) with the two longer ISIs (Experiments 1B–1C). The advantage was significantly larger in Experiments 1B–1C, compared with Experiment 1A,  $t(37) = 2.37, p = .023$ .

It is conceivable that the advantage of labels is short-lived, owing its existence to the initial unfamiliarity of the sound cues. If so, the advantage should diminish or vanish with practice. We divided each participant’s data into four equal blocks and ran an analysis of covariance with block as a covariate. Although participants became faster and more accurate over time ( $F_s > 10$ ), there were no hints of an interaction between block and cue-type for either RT or accuracy in any of the three studies ( $F_s < 1$ ). This is surprising, and we do not have a full explanation for this negative finding. However, combined with the norming results (see Appendix) it supports the interpretation that the label advantage is not due to differences in familiarity insofar as it does not change, even as participants become more familiar with the sound cues during the experiment. In Experiment 4, we test this interpretation more directly by training participants on objects with which they have no prior experience.

We now return to the Cue-Type  $\times$  Picture-Category interaction found in Experiment 1A, the experiment with the shortest cue-target delay. To explore this effect further, we divided the pictures into two semantic categories, animals ( $n = 5$ ) and artifacts ( $n = 5$ ), and ran an ANOVA with matching, cue-type, and semantic category as within-subject factors. We found a reliable main effect of semantic category: Participants responded about 20 ms faster to animals than to artifacts,  $F(1, 17) = 4.77, p = .043$ . A separate analysis with accuracy as the dependent variable was congruent with the RT analysis, showing greater accuracy for the animal targets,  $F(1, 17) = 10.07, p = .006$ . There was also a reliable Cue-Type  $\times$  Semantic-Category interaction for RTs,  $F(1, 17) = 12.37, p = .003$ . Although label cues produced faster judgments to both animals and artifacts, the label advantage was larger for artifacts ( $M = 42$  ms) than for animals ( $M = 18$  ms).

## Discussion

Hearing a verbal label compared with a nonverbal sound afforded a faster and more accurate identification of a subsequent

Table 1  
Results From Experiments 1A–1C

Experiment	Trial type	Sound cue latency	Label cue latency
1A	Match	610 (95.6)	590 (97.2)
	Mismatch	567 (94.8)	536 (95.4)
	Mean	589 (95.2)	563 (96.3)
1B	Match	625 (95.4)	575 (97.6)
	Mismatch	664 (96.5)	620 (98.0)
	Mean	645 (96.0)	598 (97.8)
1C	Match	678 (89.9)	632 (96.0)
	Mismatch	709 (94.4)	657 (96.3)
	Mean	694 (92.2)	645 (96.2)

Note. Latency means (ms) are outside the parentheses, and accuracy means (percentage correct) are inside the parentheses.



picture. This label advantage is entirely unexpected, on the view that there is a single concept that is *accessed* by the verbal cues, the nonverbal cues, and the picture and that the match/no-match response is generated based on the activation of this common concept (e.g., Gleitman & Papafragou, 2005; Jackendoff, 2002; Li, Dunham, & Carey, 2009; Snedeker & Gleitman, 2004; Snodgrass, 1984; Vanderwart, 1984).

The label advantage not only held for a wide range of delays between the cue and the picture (ISIs) but actually increased with longer ISIs (compare Figures 1A–1C). This finding further supports our claim that the label advantage does not arise from incomplete processing of the nonverbal cue. Moreover, if the only difference between the cuing conditions was the speed of activation, one would expect that the label advantage would diminish or disappear with an ISI as long as 1.5 s. That it did not, suggests to us that verbal labels do not simply activate conceptual representations faster but that representations activated via verbal cues are *different* in some way from representations activated via nonverbal means. An examination of correlations between RTs for the two cuing conditions and typicality ratings (see Further Analyses of Experiments 1A–1C: Effects of Typicality) provides further support for this claim.

Might the advantage arise from the label cues being more specific than the sound cues?

For example, one would not be surprised if a superordinate cue, such as “animal,” led to slower verification RTs of pictures of dogs than a more specific cue, such as “dog” (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Murphy & Smith, 1982). Even allowing for the possibility that a sound of, for example, a barking dog is less uniquely associated with dogs than the word “dog”—an assumption not supported by the norming results—the task provided plenty of opportunities for associating the particular sound cue with a category. One would imagine that by the time one heard the identical barking sound for the 20th time, any doubt as to its referent would be eliminated (especially because participants received accuracy feedback for each trial). The argument that the difference in specificity/predictiveness was not the main driver of the label advantage is also supported by the finding that the label advantage was found even for the categories whose sound cues elicited, in a free response task, the target categories from almost all tested participants (see Appendix). The strong association between the sounds and the labels means that it was possible that the verbal labels were (perhaps automatically) activated in response to the sounds, for example, hearing a meowing sound, activated the word “cat.” This possibility does not detract from the results but does make interpretation more complex. We return to this point in Experiment 4.

Last, we observed in Experiment 1A (and somewhat in Experiment 1C) that the label advantage was stronger for artifacts than for animals. The present work was not aimed at investigating differences between semantic categories, but we can speculate as to why there was a Cue-Type  $\times$  Semantic-Category interaction. Thompson-Schill and colleagues (1999) found that making judgments involving visual features of animals or artifacts activated the left fusiform gyrus, a cortical region associated with retrieval of visual information (e.g., D’Esposito et al., 1997). Notably, the left fusiform gyrus was also activated by judgments involving nonvisual features for animals, but not artifacts. This suggests that as a whole, representations of living things are more grounded in visual

features than are representations of artifacts (which may be organized more according to their function), an idea supported by neuropsychological and computational evidence (Farah & McClelland, 1991; Warrington & Shallice, 1984). The present task was one of *visual* identification. A correct response could only be made if participants processed the visual features of the target picture; preactivation of visual features by the cue is hypothesized to speed the response. Because representations of animals appear to be more grounded in visual information than do representations of artifacts, it is conceivable that a cue such as the crowing of a rooster activates the visual features of the rooster to a greater extent than, for example, a motorcycle engine sound activates the visual properties of a motorcycle. This results in a smaller label advantage for animals than for artifacts. Interestingly, although the size of the label advantage did not diminish with increasing cue-target delays, the Cue-Type  $\times$  Semantic-Category interaction disappeared when a longer delay was used (Experiments 1B–1C). There are many ideas about the differences between living things and artifacts beyond differential grounding in visual features, such as greater coherent covariation between perceptual aspects of living things and artifacts (Rogers & McClelland, 2004). Whether such distinctions capture the interaction we observed here between Cue-Type  $\times$  Semantic Category should be the subject of future work.

## Experiment 2

The results from Experiments 1A–1C suggest that labels activate conceptual information more effectively than familiar sounds. As outlined in Table 2, the verbal cues and nonverbal sounds differ in a number of ways. Labels are words, labels are used to refer to object categories, and labels have phonological forms that can be easily reproduced by a person. Nonverbal sounds have none of these properties. In Experiments 1A–1C, these differences were all conflated. This makes it unclear whether advantage is a *referential label* advantage or a *word* advantage, or even possibly, a speech advantage. Experiment 2 teases these apart by introducing two new cue-types (see Table 2). Verbs referring to characteristic sounds are words but do not refer to the object’s category. Sound imitations (e.g., “arf-arf”) constitute speech but are not conventional words; the degree to which they “refer” to the object’s category is unclear. If the label advantage is simply an advantage of having a cue that is a word, we should observe equal performance in the label-noun and label-verb conditions. If the label advantage at least partially derives from a speech advantage, then sound-imitation cues should lead to faster RTs than do nonverbal sound cues. A finding that noun labels lead to better performance than all the other cues would support the hypothesis that it is the referential

Table 2  
*Cue Types, Word-Hood, and Referentiality*

Cue type	Is the cue a word?	Is the cue produced by a human?	Does the cue refer to the object’s category?
Noun label	Yes	Yes	Yes
Verb labels	Yes	Yes	No
Sound	No	No	No
Sound imitations	No	Yes	Somewhat

status of labels that is responsible for the effect (the idea of referentiality is discussed in more detail in the General Discussion).

## Method

**Participants.** A total of 20 University of Wisconsin—Madison undergraduates participated for course credit.

**Materials.** The materials partially overlapped with those used in Experiments 1A–1C but had to be altered to meet the requirements of this task, namely that each item needed to have not only a characteristic sound but also a characteristic verb that referred to the sound. In addition, each item had to have a sound that could be imitated by a person in a stereotypical way. We selected the following 10 items: car, cat, clock, cow, dog, frog, motorcycle, phone, rooster, and spring. Some example verbs and sound imitations were barking/arf-arf for a dog, revving/vroom-vroom for a motorcycle, and ticking/tick-tock for a clock. The full list of cues along with a download link is provided in the Appendix. The labels (both nouns and verbs) and sound imitations were produced by the same female native English speaker. For each item, the four cue-types were edited to have the exact same duration, and all sounds were volume normalized. To compensate for the addition of the new cue-types, we reduced the number of pictures per category from the five used in Experiments 1A–1C to two.

**Procedure.** Experiment 2 was identical to Experiment 1B, except for the addition of two additional cue-types (verbs and sound imitations). Participants were instructed that they should respond with “match” if the animal or object shown in the picture was associated with the word or sound preceding it. To familiarize the participants with the nature of the cues, the experiment began with 15 practice trials. Each participant completed 320 trials: 10 Categories  $\times$  2 Picture per Category  $\times$  2 Levels of Congruence  $\times$  4 Cue-Types  $\times$  2 Repeats.

## Results

Mean latencies are shown in Figure 2. The data were analyzed as in Experiments 1A–1C. We once again found a highly reliable matching advantage ( $M_{\text{match}} = 612$  ms,  $M_{\text{mismatch}} = 661$  ms),  $F(1, 39) = 32.82$ ,  $p < .0005$ , as well as an effect of cue-type,  $F(3, 39) = 7.97$ ,  $p < .0005$ . There was no reliable Matching  $\times$

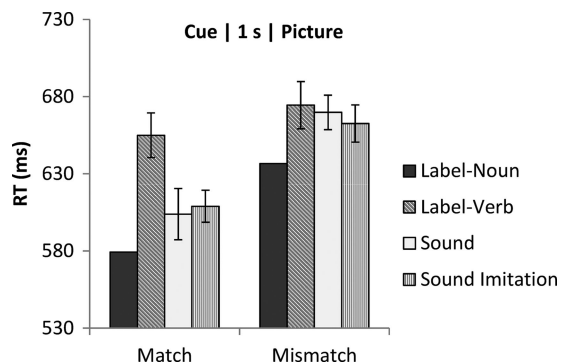


Figure 2. Verification times for each condition of Experiment 2. Error bars show  $\pm 1$  standard error of the difference between the label-noun condition and the condition of interest. RT = reaction time.

Cue-Type interaction,  $F(3, 39) = 2.23$ ,  $p = .10$ , and we collapsed matching and mismatching trials for the subsequent analyses. The label cues (606 ms) led to significantly faster RTs than did sound-imitation cues (635 ms),  $t(13) = 3.80$ ,  $p = .002$ , sound cues (636 ms),  $t(13) = 2.31$ ,  $p = .038$ , and verb cues (664 ms),  $t(13) = 4.77$ ,  $p < .0005$ . The effect of cue type was likewise significant in an item-based analysis,  $F(3, 27) = 14.45$ ,  $p < .0005$ , with the label-cue trials being faster than the other three pooled RTs from the other cuing conditions,  $F(1, 9) = 63.15$ ,  $p < .0005$ . It is unclear why the label-verb condition led to such high RTs for matching trials.

In an additional analysis, we examined whether the difference between cue-types changed between the first half and the second half of the experiment. The interaction was not significant,  $F(3, 39) = 1.73$ ,  $p = .18$ . A more targeted analysis comparing the differences between label cues and all other cues found no hint of the difference changing between the first half and the second half of the experiment,  $F(1, 13) = 0.06$ .

For nine subjects, hearing a label as a cue led to faster RTs than hearing any of the other cues ( $p = .002$ ; exact binomial test;  $H_0 =$  no effect of cue type). For comparison, the sound cues were the fastest condition for three participants, the verb cues were the fastest condition for one participant, and the sound-imitation cues were the fastest condition for one participant.

The average accuracy was 92.7%; there were no reliable differences in accuracy among the four cuing conditions ( $F < 1$ ) and no evidence of a speed–accuracy tradeoff for the label-cue condition.

## Discussion

Our aim in Experiment 2 was to examine whether the label advantage found in Experiments 1A–1C stemmed from a difference between the word status of label and sound cues or a difference in their referential status. We found that verbal category labels (e.g., “dog”) led to faster verification RTs than did words labeling the characteristic sound (e.g., “barking”) and nonword sound imitations (e.g., “arf-arf”).

The much higher mean RTs for the verb condition relative to the other conditions may have partly been due to some verb labels being somewhat ambiguous (e.g., “revving” and “bouncing”). However, other verb labels were quite unambiguous. For example, in a free-response task of the type described in the Appendix, eight of eight participants typed *dog* in response to hearing the word “barking” and typed *cat* in response to the word “meowing”; seven of eight typed *clock* in response to “ticking,” and six of eight typed *phone* in response to “ringing” (recall that in the experiment, participants received extensive exposure and accuracy feedback). In a repetition of the task with the sound imitation cues, the target category was the modal response for all the categories except “motorcycle.”

Results from Experiment 2 suggest that the label advantage does not stem simply from a difference in word status of the cues. Nouns (words referring to the category of the pictured objects), but not verbs (words referring to a property of the pictured object) or sound imitations (quasireferential expressive terms), facilitated picture recognition relative to nonverbal cues.

### Experiments 3A–3B

A limitation of using picture verification as a measure of conceptual processing is that making a verification response requires, by design, participants to explicitly compare the cue with the semantic category of the target. Thus, one might obtain faster verification responses due either to faster activation of the concept (as, for example, caused by a top-down activation of visual features by the label) or to a facilitated comparison process. That is, it may be easier to compare a picture and a verbal cue relative to a picture and a sound cue because the comparison process in both cases is mediated by the picture name. One way to tease apart these explanations, at least to some degree, is to use a task that greatly minimizes conceptual processing of the target images and a task in which the cue is entirely incidental to the task.

Experiments 3A–3B provide just such a task. Participants were asked to discriminate an upright image from an upside-down one. The task is similar to one used by Puri and Wojciulik (2008) to examine effects of general and specific cues on visual processing. Participants heard sound or label cues, as in Experiments 1A–1C, and were then presented with two side-by-side pictures of an identical object. One of these pictures was upside down. Participants had to report the side (left or right) of the upright object. The cues were either valid or invalid. Because responses were now entirely independent of the cue, we could include no-cue trials to serve as a baseline. This allowed us to measure the potential benefits of valid cues as well as potential costs of invalid cues. Note that the function of the cue here is somewhat different from its function in Experiments 1–2, in that it is now entirely incidental to the task.

### Method

**Participants.** A total of 43 University of Pennsylvania undergraduates volunteered in the experiments in exchange for course credit: 18 in Experiment 3A and 25 in Experiment 3B.

**Materials.** The verbal and nonverbal sounds were identical to those in Experiments 1A–1C, except that one item (the motorcycle) was omitted.<sup>2</sup> All auditory cues were normalized to the same volume. In addition to the verbal and nonverbal cue, we created an uninformative cue consisting of white noise with a duration of the average sound-label cue. For the picture stimuli in Experiments 3A and 3B, we used a subset of the pictures in Experiments 1A–1C, with each category instantiated by a single picture from a set of normed colored pictures (Rossion & Pourtois, 2004; see Figure S2 in supplemental materials).

**Procedure.** On each trial, participants saw for 200 ms two pictures presented simultaneously to the left and the right of a fixation cross. These pictures were identical except that one was upside down (flipped about the *x*-axis). The participants' task was simply to indicate which side of the screen contained the upright picture by pressing the *Z* key with their left index finger if it was the picture on the left and the */* key (the *slash* key) with their right index finger if it was the picture on the right. It was stressed that it did not matter what object was shown in the picture. The pictures were preceded by an auditory cue. The trials were evenly divided into label cues, sound cues, and uninformative noise cues. The label and sound cues validly cued the upcoming picture on 80% of the trials. On the remaining 20%, the cue was invalid, for example,

participants would hear “cow” or hear a mooing sound but then see a car. This allowed us to measure the advantage of a valid cue relative to a noise cue (Are people faster to locate the upright cow after hearing “cow”/a mooing sound?), the cost of an invalid cue relative to a noise cue baseline, and, critically, a comparison of these benefits and costs for label versus sound cues. Unlike Experiment 1, in which the participant could not respond without attending to the cue, in the present experiment, one could achieve 100% accuracy while completely ignoring the cues. Setting the validity proportion to 80% provided an implicit signal that the cues should be attended.

Experiments 3A and 3B were identical except for the delay between the offset of the cue and the onset of the pictures. In Experiment 3A, the delay was 400 ms. In Experiment 3B, it was lengthened to 1 s to determine whether the results observed in Experiment 3A were due to insufficient time to process the nonverbal sound. There were 20 practice and 300 experimental trials in each experiment.

### Results

Mean latencies are shown in Figure 3. Latencies were analyzed with a repeated-measures ANOVA followed by planned comparisons. The first analysis included validity and cue-type (sound vs. label) as within-subject fixed factors (validity is undefined for noise cue trials). We found a highly reliable effect of validity, with valid trials being reliably faster than invalid trials,  $F(1, 17) = 39.72, p < .0005$ . We also found a significant Validity  $\times$  Cue-Type interaction, with label cues showing a larger cuing effect than sound cues,  $F(1, 17) = 8.23, p = .011$ . Relative to the no-cue baseline, valid sound cues improved performance,  $t(17) = 2.84, p = .03$ . Label cues also improved performance,  $t(17) = 5.01, p < .0005$ , but importantly, this improvement was significantly greater than the improvement due to sounds,  $t(17) = 2.93, p = .009$ . Relative to the no-cue baseline, invalid label cues significantly slowed responses,  $t(17) = 4.38, p < .0005$ ; sounds cues did not,  $t(17) = 1.19, p > .2$ . The effect of invalid cues differed reliably between cuing conditions: invalid labels hurt performance more than invalid sounds did,  $t(17) = 2.12, p = .048$ . Accuracy was very high ( $M = 97.8\%$ ) and did not vary reliably between any of the conditions ( $ps > .5$ ).

Did the label advantage result from a lack of time to process the sound cue? This was unlikely, given the results of Experiments 1B and 1C, but nevertheless, we conducted a replication of Experiment 3A with a longer (1 s) delay between cue offset and picture onset. There was, once again, a highly reliable validity effect,  $F(1, 24) = 8.41, p = .008$ . The Cue-Type  $\times$  Validity interaction was marginally significant,  $F(1, 24) = 4.26, p = .05$ . As shown in Figure 3B, valid labels led to reliably faster RTs relative to baseline,  $t(24) = 2.45, p = .022$ , whereas sounds did not,  $t(24) = 1.13, p > .2$ , though valid labels and sounds lowered RTs by a similar degree. As in Experiment 3A, invalid cues resulted in slower responses relative to the baseline, though with the longer

<sup>2</sup> This procedural deviance was due to the fact that Experiment 3 was conducted chronologically prior to Experiment 1. The motorcycle item was added to balance the number of pictures representing artifact and animal categories.

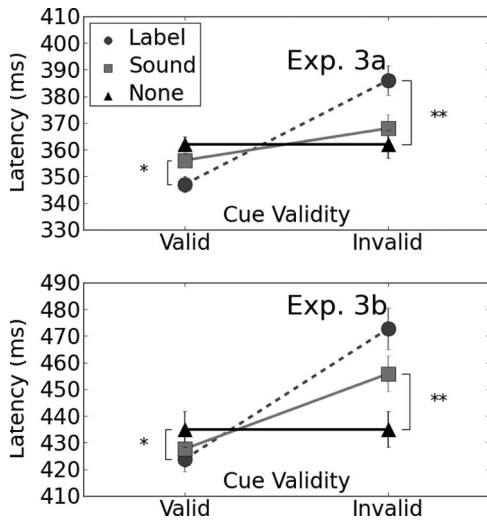


Figure 3. Results of Experiments 3A–3B. Error bars show  $\pm 1$  standard error of the difference between the no-cue (noise cue) condition and the condition closest to its mean. The mean of the noise cue trials is plotted twice for ease of comparison. Exp. = experiment. \*  $p < .05$  (condition difference). \*\*  $p < .001$  (condition difference).

ISI, both label cues,  $t(24) = 4.90$ ,  $p < .0005$ , and sound cues,  $t(24) = 3.14$ ,  $p = .004$ , increased the RTs. Importantly, the RT cost of invalid label cues was greater than that for invalid sound cues,  $t(24) = 2.22$ ,  $p = .036$ . Accuracy was very high ( $M = 98.0\%$ ) and did not vary reliably between any of the conditions,  $ps > .5$ . In sum, in both Experiments 3A and 3B, labels continued to function as more effective cues than nonverbal sounds.

## Discussion

Experiments 3A–3B showed that auditory cues facilitate judgments of images that are congruent with the cues and slow down judgments of images that are incongruent with the cues. Critically, verbal cues produce substantially greater validity/invalidity effects relative to uninformative cues than do nonverbal cues. Because in these experiments the cue was entirely incidental to the task—that is, the response was independent of the cue—the observed pattern is consistent with our claim that verbal cues are particularly effective in activating at least the visual components of the conceptual representation.

Despite the qualitative similarity of the results in Experiments 3A and 3B—a greater effect of label cues relative to sound cues—there was suggestive evidence that increasing the ISI from 400 ms to 1 s had some effect. The longer ISI increased the effectiveness of invalid cues, with sound cues now having a significant negative impact on RTs relative to the no-cue baseline (though still a significantly smaller impact than invalid label cues). An additional departure from Experiment 3A was that although valid label cues significantly decreased RTs and valid sound cues did not, the two valid conditions did not differ from each other. The effect of increasing the ISI suggests that cues processed for a longer time are more effective (though note that they do not lead to faster RTs overall). Having a longer time in which to “commit” to a particular category appears to more significantly impair per-

ceptual judgments involving nonmatching categories. It is possible that an increase in the effectiveness of sound cues was a form of verbal mediation, with the longer ISI allowing sufficient time for the activation of the verbal label in response to the sound cue (cf. Experiment 4).

Although the present task did not require participants to identify the target image, it is likely that participants identified the target images (i.e., they knew that the image on the left was of an upside-down cow and not just something that was upside down). This does not detract from the present finding because the response the participants needed to make was completely independent of the category. Participants did not *need* to hear a cue of any kind to make a response (indeed, accuracy was no lower in the uninformative cue condition than in other conditions).

The results from Experiments 3A–3B are consistent with our claim that labels activate visual information more effectively than do nonverbal cues and speak to the broader claim that the representations activated by verbal means (via a noun label) are not identical to representations activated by nonverbal means.

## Experiment 4

Experiments 1–3 have examined the effects of words and sounds on the visual processing of objects with which participants have had extensive prior experience. We had no control over this prior experience, making it difficult to know, for instance, whether the label advantage derives simply from differences in the quantity of experience with nouns compared with other cues—verbal and nonverbal. If true, the effects we are seeing may be different manifestations of familiarity making the rather mundane point that more familiar cues are more effective cues. (see Lupyan & Spivey, 2010b, for an argument that many effects of familiarity are actually effects of meaningfulness). Additionally, although the norming data leads one to argue against the possibility that the label advantage occurs because sound cues are systematically less specific or predictive of the target category than are the label cues, it is difficult to fully eliminate this possibility with norming.

This experiment tests the hypothesis that verbal cues activate conceptual information differently from nonverbal cues even when (a) the concepts are newly acquired and when (b) the experience with verbal and nonverbal cues is fully equated. Our goal in Experiment 4 was to test the hypothesis that verbal cues will activate conceptual information more effectively than will nonverbal cues when both conditions above are met. In this experiment, participants learned either verbal labels or nonverbal sounds with six novel categories. Example items are shown in Figure 4. This design had the advantage of equating the learning opportunities between the cues and their visual referents and measuring the degree to which participants are able to learn the picture-label versus picture-sound associations. Because individuals in the sound group never learned the corresponding labels (i.e., cue-type was now a between-subjects factor), there was no opportunity for the sound group participants to label the sound with its corresponding name, as was possible in Experiment 3A (as well as in Experiments 1–2). That is, with familiar categories, hearing a meowing sound can activate the verbal label (“cat”). This is not possible in the present study because participants in the sound condition have



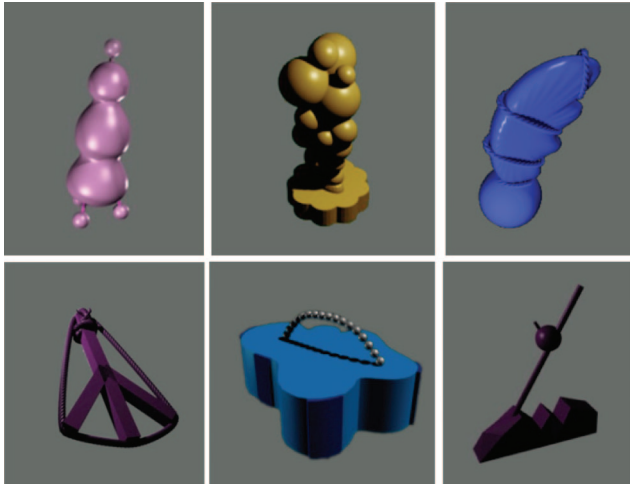


Figure 4. Materials used in the learning task for Experiment 3. Each category had two additional variants.

no names for the objects. A finding of a label advantage in this context would further strengthen the conclusion that words—even those just learned—have a special power to activate visual features (partly) constitutive of the object concept.

After being trained to associate these objects with either novel auditory labels or novel sounds, participants performed a speeded orientation judgment task identical to Experiment 3B.

## Method

**Participants.** A total of 20 University of Pennsylvania undergraduates volunteered in the experiment in exchange for course credit.

**Materials.** The training set consisted of six categories of novel three-dimensional objects shown on a computer screen (see Figure 4). There were three variants of each object to increase visual heterogeneity. These variants involved changes in viewpoint and slight changes in feature configuration. Each category was paired with a novel label (shonk, whelp, scaif, crech, foove, and streil). Each of these nonce words was designed to have approximately equal bigram and trigram statistics and similar real-world lexical neighborhoods. We also created six nonverbal sounds, one for each category. These were created by splicing and editing environmental and animal sounds to create six sounds that were not readily nameable, as judged by pilot testing. The sounds may be downloaded from <http://sapir.psych.wisc.edu/stimuli/labelsSoundsExp4.zip>

**Procedure.** Participants were randomly assigned to either the *label* group or the *sound* group to form two groups of equal size ( $n = 10$ ). The experimenter told participants a cover story about the task, explaining that they would see some alien musical instruments and animals and would be asked to learn what sounds they make (sound group) or what they are called (label group).<sup>3</sup> The experiment had three parts, presented one after the other. In the first part (*pretraining*), participants passively viewed all the 12 trials during which all three exemplars of each category were presented together with a recording, e.g., “These are all shonks” (for the label condition) or “These all make the sound \_\_\_” (for the

sound condition). Part 2 consisted of a verification-with-feedback task. Participants saw two exemplars from different categories followed by a prompt, for example, “Which one’s the streil?” or “Which one makes the sound \_\_\_” and had to select whether the left or right stimulus matched. There were 180 verification-with-feedback trials. All verbal and sound cues were presented only auditorily.

The last part of Experiment 4 was a replication of Experiment 3B, but now with the newly learned novel stimuli. As in Experiment 3B, participants had to judge whether the left or right picture was upright (i.e., in the familiar orientation) after hearing one of the newly learned nonverbal sounds or labels. The images were presented for 200 ms after a 1 s delay following the offset of the cue.

## Results

Participants were highly adept at learning the six categories. After pretraining—just two exposures to each category—participants performed the verification-with-feedback task with ~95% accuracy. The label group was slightly less accurate and slower than the sound group ( $ps = .08$ ; ANOVA with condition as a between-subjects variable); there were no Reliable Condition  $\times$  Block interactions. By Block 5, both groups were performing at 99% (see Figure S3 in the supplemental materials), demonstrating that participants were able to learn both labels and nonverbal cue-to-picture, equally quickly, with a slight advantage for learning to associate novel objects with nonverbal sounds.

The critical part of the experiment was the subsequent orientation judgment task. Having ruled out differences in familiarity and association strength between labels and sounds, would labels continue to evoke visual activations in a more robust way than sounds? Indeed, that is what we found. As shown in Figure 5, there was a significant validity advantage ( $M_{\text{valid}} = 400$  ms,  $M_{\text{invalid}} = 443$  ms,  $F(1, 18) = 49.55$ ,  $p < .0005$ , and this validity advantage was significantly larger for labels ( $M = 59$  ms) than sound cues (28 ms),  $F(1, 18) = 6.14$ ,  $p = .023$ . The valid cue also reduced RTs relative to the uninformative noise cue,  $F(1, 18) = 38.34$ ,  $p < .0005$ , and this benefit also was larger for the label ( $M = 51$  ms) than for the sound trials ( $M = 22$  ms),  $F(1, 18) = 10.73$ ,  $p = .004$ . Finally, there was a significant cost of hearing an invalid cue relative to not hearing a cue,  $F(1, 18) = 8.08$ ,  $p = .011$ . The pattern was a qualitative match to that observed in Experiment 3, but in the present case, the cost due to invalid labels was not reliably greater than the cost due to invalid sounds ( $F < 1$ ). Examining the effects of invalid versus no-cue trials separately for sound and label groups revealed marginally slower RTs in the invalid-cue condition, labels:  $t(9) = 2.25$ ,  $p = .051$ ; sounds,  $t(9) = 2.00$ ,  $p = .076$  (see Figure 5).

An identical pattern of results was found when we repeated the analyses using RT proportions (e.g.,  $RT_{\text{invalid}}/RT_{\text{valid}}$ ) instead of

<sup>3</sup> Participants were not told which objects were animals and which were instruments but were told that they would never need to know which ones were which. This distinction was added to the cover story to impose some control over what participants thought the novel nonverbal sounds were. Some of these sounds had characteristics of artifacts, whereas others sounded like they could be made by an animate entity.

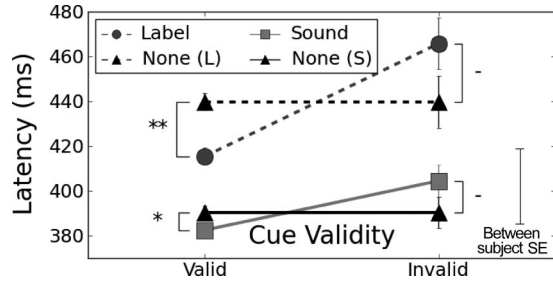


Figure 5. Results of Experiment 4. Error bars show  $\pm 1$  standard error (SE) of the difference from the cueing to the no-cue condition from the label (L) or sound (s) group. Mean differences marked “-” are marginally significant ( $.08 > p > .05$ ).

\*  $p < .05$  (condition difference). \*\*  $p < .001$  (condition difference).

RT differences. There were no reliable effects of cues on accuracy ( $F < 1$ ).

The two groups did not differ in overall response times,  $M_{\text{label}} = 433$  ms,  $M_{\text{sound}} = 389$  ms,  $F(1, 18) = 1.70$ ,  $p > .2$ , or accuracy,  $M_{\text{label}} = 96.4\%$ ,  $M_{\text{sound}} = 95.7\%$ ,  $F < 1$ . There were also no reliable effects of validity on accuracy,  $F(1, 18) = 2.04$ ,  $p = .17$ .

## Discussion

In Experiment 4, we had complete control over participants’ exposure to all the materials. We could thereby ensure that participants were equally familiar with verbal and nonverbal cues and had equivalent experience associating the two with their referents. Indeed, participants were equally proficient in learning to associate the novel categories with labels and sounds. Given this equivalence, we could then ask whether labels still had an advantage in activating visual information. A positive finding would indicate that one cannot account for the label advantage simply through differences in cue strength/familiarity.

After only about 10 min of training (pretraining plus the category verification-with-feedback task), hearing a label or sound facilitated a visual judgment, as revealed by an RT advantage on valid trials and an RT cost on invalid trials. This in itself is quite remarkable. Critically for our thesis, the label cues were more reliable in activating the concept than were the sound cues, as measured by stronger cuing effects of labels relative to sounds. This result confirms that even when familiarity and type of experience with verbal and nonverbal associates are strictly equated, verbal cues activate the associated concept more effectively than nonverbal cues. How is this possible? One general answer is that participants in our studies brought with them a lifetime of experience with language. This experience is not limited to already familiar words and includes knowledge about the *type* of relation words have with the objects/categories they denote. This prior experience is brought to bear on learning novel categories (this idea is explored computationally by Rakison & Lupyan, 2008).

### Further Analyses of Experiments 1A–1C: Effects of Typicality

Our original goal in the present work was to ask the basic question of whether concepts activated via verbal means differ in some way from those activated via nonverbal means. The reliable

difference in accuracy and RTs provide one type of evidence for difference between conceptual representations cued by verbal and nonverbal means. In this section, we provide additional evidence that the differences between the label and the sound conditions in Experiments 1A–1C are not simply differences in speed in activating a putatively common representation but rather are that the two cues produce representations that actually diverge over time.

In assembling the pictorial materials of Experiments 1A–1C, we collected typicality ratings for each picture by asking a separate group of 15 participants from the university participant pool to rate the typicality of each picture relative to the category referred to by the label. Participants saw each of the 50 pictures used in Experiments 1A–1C and responded to the prompt, “How typical is this picture of the text label below it?” using a Likert scale (1 = *very atypical*; 5 = *very typical*).

Not surprisingly, picture identification times were correlated with the rated typicality of the pictures: people were faster to recognize more typical pictures. The correlation coefficients for Experiments 1A–1C were, respectively,  $r = -.32$ ,  $p = .024$ ;  $r = -.14$ ,  $p = .32$ , and  $r = -.44$ ,  $p = .002$ . The corresponding  $b$  values were  $-15$  ms,  $-10$  ms, and  $-31$  ms.<sup>4</sup>

Importantly, as shown in Figure 6, the relation between typicality and recognition times appeared to depend on how the concept was cued. Across all three studies, the relation between typicality and RTs was numerically stronger (i.e., had a more negative slope) for label-cue trials compared with sound-cue trials.

Collapsing across the three experiments (by averaging the mean RTs for each picture across the three experiments, cued by either the sound or the label), we found that the label advantage was reliably stronger for the more typical items (i.e., hearing the labels induced a steeper typicality gradient),  $F(1, 47) = 4.51$ ,  $p = .039$ .

To test the hypothesis that the association between label advantage and typicality varied as a function of ISI, we ran a general linear model predicting the difference between label and sound conditions from the ISI and typicality ratings. The model showed a reliable interaction between typicality and ISI ( $b = .047$ ,  $t = 2.2$ ,  $p = .028$ ; outliers, identified as items having standardized Cook’s distances  $> 3.5$  SDs, were excluded from the analysis): The effect of typicality on the label advantage increased with increasing ISI.

Analyzing the three experiments separately, this effect did not reach significance for Experiments 1A and 1B. With the longer ISI of Experiment 1C, the difference between effects of typicality on RTs between sound and label trials became reliable,  $F(1, 47) = 8.79$ ,  $p = .005$ . This difference also means that the label advantage was significantly correlated with typicality ( $r = .40$ ,  $p = .005$ ): The label advantage was largest for the more typical, compared with less typical, exemplars. Thus, at least with a longer ISI, label cues induced a steeper typicality gradient than did sound cues (see Figure 6, right). The analysis above supports the idea that labels activate more typical representations of the cued category than do nonverbal cues and is consistent with a number of other reports showing that effects of verbal labels on memory and basic visual processing is strongly modulated by typicality (Lupyan, 2007b, 2008a, 2008b).

<sup>4</sup>One item, the cartoon car image, was identified as an outlier—standardized Cook’s distance  $> 6$  SDs—and was removed. Inclusion of this item artificially inflated all correlation coefficients.

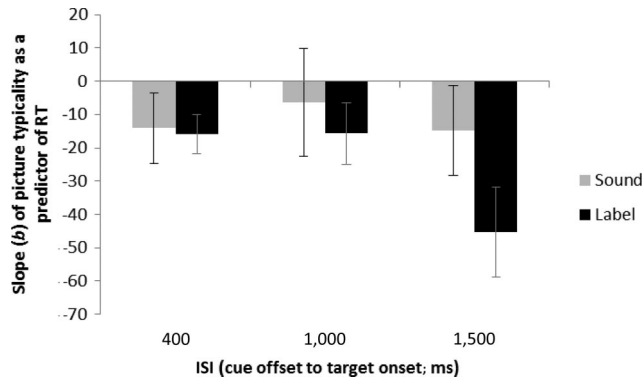


Figure 6. Effects of picture typicality on mean RT. For example, a coefficient of  $-10$  indicates that for each 1 point increase in rated typicality, there is a 10 ms increase in the speed of correct identification. Error bars show  $\pm 1$  standard error of the coefficient estimate. RT = reaction time; ISI = interstimulus interval.

### General Discussion

In acquiring language, humans master an elaborate system of conventions, learning that certain sequences of sounds that denote categories of objects, relations, types of motion, emotions, and so on. With this system of associations in place, a concept can be activated simply by hearing the appropriate word—arguably the dominant function of language. In this work we asked whether conceptual representations are activated differently via verbal means and nonverbal means. An affirmative finding has important consequences for not only the understanding of the relation between language and other cognitive processes but also for understanding how conceptual information is brought to bear on a given situation.

Experiments 1A–1C comprised picture identification tasks in which participants heard a word (e.g., “dog”) or a characteristic sound (e.g., barking) and made a match/no-match response to a picture appearing after a varying delay. Verbal labels were more effective at activating the concept, as evidenced by consistently shorter RTs on the label, compared with sound trials. A subsequent analysis of the data provided preliminary evidence that labels instantiated conceptual representations corresponding to more typical category exemplars and representations that were more similar across participants. Experiment 2 contrasted labels referring to the objects/animals with other types of verbal cues (see Table 2). We found that referential words resulted in faster identification times than did all other cues (e.g., faster than speech utterances associated with the category). Experiments 3A–3B extended these findings to a simple, visual minimally semantic task in which participants had to discriminate a normal exemplar from an upside-down one. Relative to baseline and to sound cues, valid label cues facilitated performance; invalid label cues hurt performance. Experiment 4 showed that the label advantage found in Experiment 3 holds not only for familiar stimuli but, in fact, emerges quickly with novel stimuli, the visual forms of which are activated more quickly by newly learned verbal labels than are equally well-learned sounds.

The finding that both highly familiar categories (e.g., dogs, cats, and cars) and newly learned categories can be activated more

effectively by labels than by sounds, even a full 1.5 s after cue offset, hints at the powerful effects of language on the activation of at least the visual components of the conceptual representation and shows that at least temporarily, the representation of “dog” that is activated by hearing the category name is not the same as that activated by hearing a barking sound. These findings contradict the idea that language simply *accesses nonverbal concepts* (e.g., Gleitman & Papafragou, 2005; Jackendoff, 2002; Li et al., 2009; Snedeker & Gleitman, 2004; Snodgrass, 1984; Vanderwart, 1984) because presumably such concepts should have been accessed in the same way by equally informative and predictive nonverbal cues.

The analysis of typicality in Experiments 1A–1C suggest that not only are conceptual representations activated more quickly by verbal cues but that they are *different*, diverging over longer delays between the cue and the target, with representations activated by labels providing a better match to typical category exemplars.

### Relevance of the Present Work to Understanding the Relation Between Language and Thought

Much has been written on the subject of how learning and using language might supplement or augment our cognition and perception (see Boroditsky, in press; Casasanto, 2008; Wolff & Holmes, 2010, for recent reviews). In most work investigating the relations among language, cognition, and perception, it has been assumed that verbal and nonverbal representations are fundamentally distinct and that the goal of the “language and thought” research program is to understand whether and how linguistic representations affect nonlinguistic representations (Wolff & Holmes, 2011). This assumption is problematic for a number of reasons, with the primary one being that it is impossible to determine a priori whether a particular representation is verbal.

Despite the inherent difficulty with distinguishing verbal and nonverbal representations, practically all language-and-thought debate has taken place under the assumption that information communicated or encoded via language comprises what is essentially a “verbal” modality—the notion at the core of dual coding theory (Paivio, 1986). On this account, the reason why a verbal cue and a nonverbal cue can lead to different representations is that verbal and nonverbal representations are distinct and processed separately, being combined at some higher level (e.g., Dessalegn & Landau, 2008; Mitterer, Horschig, Musseler, & Majid, 2009; Pilling, Wiggett, Ozgen, & Davies, 2003; Roberson & Davidoff, 2000). Indeed, the very use of terms such as “verbal” and “nonverbal” representations (e.g., Wolff & Holmes, 2010, for review and discussion) presupposes that they are separable.

On a sensorimotor view of concepts to which we subscribe, visual representations used to make the decisions in the present tasks are partly constitutive of the concept to which the label refers. The label does not constitute a verbal code for the concept. Rather, it is a cue (Elman, 2004) that modulates, among other things, ongoing perceptual processing. Thus, although it is useful to distinguish between a verbal and a nonverbal *stimulus*, the distinction between a verbal and a nonverbal representation may be moot.

A framework that naturally accommodates this account and the present findings is what Lupyan has called the *label feedback hypothesis* (Lupyan, 2007a, 2008b). According to this view, verbal

labels, once activated, modulate ongoing perceptual processing. As a result, the dog that you see after hearing the word “dog” is, in a sense, not the same dog as the dog you see after hearing a barking sound or just thinking about dogs. This is not language voodoo: It is simply the consequence of the fact that visual representations involved in making even simple visual decisions are subject to substantial top-down modulations (e.g., Foxe & Simpson, 2002; Lamme & Roelfsema, 2000). Language is one form of such top-down modulation. Because the present experiments require visual decisions, modulation of visual representations induced by language would affect our dependent variables.

According to the label-feedback hypothesis, rather than trying to decide whether representation comprises a verbal or visual “code” (Dessaegh & Landau, 2008), one should aim to classify *behavior* on specific tasks as either being modulated by language, being modulated differently by different languages, or being truly independent of any experimental manipulations that can be termed linguistic. The present findings suggest that a concept activated via different means, for example, via an auditory verbal label or via nonverbal auditory information, is detectably different. Specifically, the visual aspects of a category (e.g., the shape of a dog) are more effectively activated after hearing a word rather than a nonverbal sound or a nonreferential word associated with the category.

Obviously, the concept of a dog is more than just its visual information. The present studies are limited to testing visual aspects of concepts (which we believe to be partly constitutive of the dog “concept”). Verbal cues can be used to elicit all sorts of representations including smells (Herz, 2000), motor actions (Willems, Toni, Hagoort, & Casasanto, 2009), and emotions (Lindquist, Barrett, Bliss-Moreau, & Russell, 2006). It remains an open question whether the verbal modality provides an advantage (or a disadvantage) in eliciting nonvisual information, as well as what specific aspects of visual representations are most strongly evoked by verbal means (e.g., texture, shape, size, color).

It is informative that learning nouns and attention to shape are, at least in English, closely linked during language acquisition (Colunga & Smith, 2005; Woodward, Markman, & Fitzsimmons, 1994), suggesting that the label advantage observed in the present experiments may have to do with nouns selectively activating shape information. Although such a shape bias may contribute to the present results, we believe that it is overly limiting to focus on shape as a dimension selectively affected by verbal cues.

Lexicalization patterns differ dramatically between languages (e.g., Bowerman & Choi, 2001; Evans & Levinson, 2009; Lucy & Gaskins, 2001; Majid, Gullberg, van Staden, & Bowerman, 2007). The present results suggest that the consequence of activating a conceptual representation via verbal means may differ cross-linguistically according to the patterns of learned associations between words and their referents (e.g., Huettig, Chen, Bowerman, & Majid, 2010). Words may matter far more for conceptual representations than previously considered, in that some concepts may only attain sufficient “coherence” when activated by verbal means.

### Further Considerations

It has been long known that concepts activated in different contexts are different. For example the piano concept that is

activated by reading about moving pianos is different from the concept activated by reading about playing pianos (Tabossi & Johnson-Laird, 1980). Similarly, findings that instantiations of eagles differ when reading about flying versus sitting eagles (Zwaan, Stanfield, & Yaxley, 2002) also speak to the flexibility with which we deploy our semantic knowledge. The label advantage observed in the present studies was obtained without explicitly manipulating context or experimental task. In the present studies, although different aspects of the concept’s meaning may have been activated by the verbal and nonverbal cues (that is partially our point), the task and the participants’ goal was exactly the same on all trials. For example, in Experiments 3–4, participants were performing one extremely simple task: judging which side of the screen showed the upright of a picture.

A further question concerns the specificity of information evoked by verbal versus nonverbal means. On the one hand, language can be used to specify visual properties with a great degree of specificity. A sentence such as “The ranger saw the eagle in the sky” activates a representation of not just an eagle but an eagle with outstretched wings (Zwaan et al., 2002). The hypothesis that sentence context actually modulates visual processing has recently received more direct support through a MEG study by Hirschfeld and colleagues (Hirschfeld, Zwitserlood, & Dobel, 2010).

On the other hand, in the absence of such rich context, concepts evoked through verbal means (of which perceptual images are a part) may be more categorical than concepts evoked through nonverbal means. For example, the concept evoked by the word “dog” may, in the absence of other information, correspond to a more prototypical exemplar of the category than the more idiosyncratic representation activated by other cues. The section on typicality effects provides some support for this claim: We found a significantly steeper typicality gradient in verification RTs following label cues, compared with sound cues. The difference in typicality gradients between label and sound conditions appeared to increase with longer ISIs, reaching significance for the longest (1,500 ms) ISI (Experiment 1C). This finding is consistent with the notion of verbal labels aligning category representations between individuals.

Whether an increase in the degree to which representations are categorical and/or typical is good depends on the task. Evoking a concept through verbal means may facilitate categorization (Casasola, 2005; Lupyan et al., 2007; Nazzi & Gopnik, 2001; Plunkett, Hu, & Cohen, 2008), enhance inference (e.g., Yamauchi & Markman, 2000), and enhance perceptual discriminability (Kikutani, Roberson, & Hanley, 2008; Lupyan, 2008a). However, when high-fidelity analogical representations are called for, as is the case in a *within-category* visual recognition task, representations modulated by verbal labels may prove detrimental (Lupyan, 2008b). Likewise, the finding that labels facilitated the recognition of artifacts more than the recognition of animals in Experiment 1A might hinge on the task involving visual recognition. The effect of labels on animal versus artifact concepts may well be reversed in a task calling on attributes from other modalities.

The consequences of activating categorical representations may go far beyond visual tasks requiring recognizing or discrimination of category instances. For example, activation of concepts through verbal means may be critical in reasoning about generics, for example, dogs bark; he’s a carrot eater (cf. he eats carrots; Gelman



& Coley, 1991). Categorical representations may also be crucial to reasoning about spatial relations such as “above” (Kosslyn et al., 1989) and for performing reasoning about relations: For instance, although naïve chimpanzees fail to group together instances of *same* versus *different* object pairs, they succeed after being taught to associate arbitrary tokens with instances of *same* and *different* relations—an experience that may facilitate forming categorical representations for instances of *same* versus *different* relations (Thompson, Oden, & Boysen, 1997).

### The Special Status of Words: From Reference Versus Associations to Reference Via Associations

The present work suggests that words are not simply a “pointer” or a means to access a nonverbal concept. Rather, they provide a special way to activate the multimodal representation that constitutes the concept. We have argued that verbal labels activate conceptual information—the visual components, at least—(a) more quickly and accurately and (b) in a less idiosyncratic way. This finding is well captured by the label feedback hypothesis briefly described above (Lupyan, 2007a, 2008b), which proposes that labels are cues that modulate the trajectories of perceptuo-conceptual representations. In this final section, we attempt to relate this way of thinking about the nature of verbal labels to the currently ongoing debate in the literature regarding the special status of words as referential entities.

A common claim in developmental psychology is that words are special because they are referential (Waxman & Gelman, 2009; Xu, 2002). A considerable amount of work within the field has been done to investigate the degree to which infants expect words to refer to object kinds, object properties, and so on (e.g., Colunga & Smith, 2005; Dewar & Xu, 2009; Waxman, 1999; Xu, 2003) and the consequences that learning and using words has on the categorization and inference process (Nazzi & Gopnik, 2001; Plunkett et al., 2008; Sloutsky & Fisher, 2001, 2004; Waxman & Hall, 1993; Waxman & Markow, 1995; Yoshida & Smith, 2003). This topic has been the subject of an ongoing debate between an associationist account in which words are features of the associated stimulus and an account in which words are special because they have referential powers (e.g., see Waxman & Gelman, 2009, for review).

We believe this dichotomy is a false one. There are indeed numerous reasons to reject simple associationist accounts of word learning and effects of words on concepts. For example, relying solely on the cue-to-picture associations learned during Experiment 4 fails to explain why labels relative to nonverbal sounds, despite being equally well-learned, are more effective cues. Clearly, accounting for the results of Experiment 4 requires that participants bring to bear on the task prior expectations about words.

The statement that words derive their powers from “reference” is an odd one. The fact that words refer is a property of language, not a mechanism for understanding interactions between language and thought. The idea that words are features of the entities to which they refer is also an odd one. It is useful descriptive shorthand to talk about bananas having features such as *is edible*, *is yellow*, and *is curved*. What makes these features useful is that objects that are not bananas possess some of them, and we can talk about two objects from different categories sharing a feature such as *is yellow*. But only bananas are “bananas.” This makes direct

comparisons of perceptual features and labels problematic (cf. Sloutsky & Fisher, 2004).

An effect of word learning and word usage on concepts means that concepts activated in response to a word are systematically different than in response to various nonverbal or nonreferential cues. It is this phenomenon that we should try to understand. Word learning requires associating an arbitrary token (the word) with external entities (objects, object attributes, relations, types of motion, etc.). The learning of these word-to-world associations may make it possible to activate the referent in a more categorical way. That is, the representations become more invariant to within-category differences and more sensitive to between-categories differences. In the present experiments this has the effect of facilitating visual recognition (Experiments 1–2) and facilitating locating the canonical (upright) object (Experiments 3–4). We believe this is accomplished through associationist mechanisms: Words are cues that activate perceptuo-conceptual representations. Our experience of treating words as referential, that is, “standing in” for real-world objects, enables verbal cues to activate these representations in a “special”—perhaps more categorical—way than is possible by nonverbal means. How this is accomplished vis-à-vis neural mechanisms is still a mystery.

### References

- Allport, D. A. (1985). Distributed memory, modular subsystems, and dysphasia. In S. K. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 32–60). New York, NY: Churchill Livingstone.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639
- Boroditsky, L. (in press). How the languages we speak shape the ways we think: The FAQs. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics*. Cambridge, England: Cambridge University Press.
- Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge, England: Cambridge University Press.
- Bruner, J. S., Austin, G. A., & Goodnow, J. J. (1956). *Study of thinking*. New York, NY: Wiley.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s mechanical Turk. *Perspectives on Psychological Science*, 6, 3–5. doi:10.1177/1745691610393980
- Carey, S. (1987). *Conceptual change in childhood*. Cambridge, MA: The MIT Press.
- Casasanto, D. (2008). Who’s afraid of the big bad whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, 58, 63–79. doi:10.1111/j.1467-9922.2008.00462.x
- Casasola, M. (2005). Can language do the driving? The effect of linguistic input on infants’ categorization of support spatial relations. *Developmental Psychology*, 41, 183–192. doi:10.1037/0012-1649.41.1.183
- Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 162–183). Cambridge, England: Cambridge University Press.
- Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112, 347–382. doi:10.1037/0033-295X.112.2.347
- Dennett, D. C. (1996). *The role of language in intelligence. What is intelligence? The Darwin College Lectures*. Cambridge, England: Cambridge University Press.

- D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., & Farah, M. J. (1997). A functional MRI study of mental image generation. *Neuropsychologia*, *35*, 725–730. doi:10.1016/S0028-3932(96)00121-2
- Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science*, *19*, 189–195. doi:PSCI2066
- Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*, *20*, 252–257. doi:10.1111/j.1467-9280.2009.02278.x
- Egley, R., Driver, J., & Rafal, R. D. (1994). Shifting visual-attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161–177. doi:10.1037/0096-3445.123.2.161
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*, 301–306. doi:10.1016/j.tics.2004.05.003
- Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*, 201–204. doi:10.3758/BF03212870
- Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253. doi:10.1093/cercor/bhp188
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–448. doi:10.1017/S0140525X0999094X
- Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*, 339–357. doi:10.1037/0096-3445.120.4.339
- Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining “early” visual processing. *Experimental Brain Research*, *142*, 139–150. doi:10.1007/s00221-001-0906-7
- Gelman, S. A., & Coley, J. D. (1991). Language and categorization: The acquisition of natural kind terms. In J. D. Byrnes & S. A. Gelman (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146–196). Cambridge, England: Cambridge University Press.
- Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.
- Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633–661). Cambridge, England: Cambridge University Press.
- Gliga, T., Volein, A., & Csibra, G. (2010). Verbal labels modulate perceptual object processing in 1-year-old infants. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn.2010.21427
- Gumperz, J. J., & Levinson, S. C. (1996). *Rethinking linguistic relativity*. Cambridge, England: Cambridge University Press.
- Harnad, S. (2005). Cognition is categorization. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 19–43). San Diego, CA: Elsevier. doi:10.1016/B978-008044612-7/50056-1
- Herz, R. S. (2000). Verbal coding in olfactory versus nonolfactory cognition. *Memory & Cognition*, *28*, 957–964. doi:10.3758/BF03209343
- Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2010). Effects of language comprehension on visual processing: MEG dissociates early perceptual and late N400 effects. *Brain and Language*. Advance online publication. doi:10.1016/j.bandl.2010.07.002
- Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, *12*, 360–365. doi:10.1111/1467-9280.00367
- Huetting, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, *10*, 39–58. doi:10.1163/156853710X497167
- Jackendoff, R. S. (2002). *Foundations of language: Brain, meaning, grammar, and evolution*. Oxford, England: Oxford University Press.
- James, W. (1890). *Principles of psychology* (Vol. 1). New York, NY: Holt. doi:10.1037/10538-000
- Keil, F. C. (1992). *Concepts, kinds, and cognitive development*. Cambridge, MA: The MIT Press.
- Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review*, *15*, 787–794. doi:10.3758/PBR.15.4.787
- Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., & Gabrieli, J. D. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 723–735. doi:10.1037/0096-1523.15.4.723
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*, 571–579. doi:10.1016/S0166-2236(00)01657-X
- Levinson, S. C. (1997). From outer to inner space: Linguistic categories and non-linguistic thinking. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 13–45). Cambridge, England: Cambridge University Press.
- Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology*, *58*, 487–524. doi:10.1016/j.cogpsych.2008.12.001
- Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, *6*, 125–138. doi:10.1037/1528-3542.6.1.125
- Lucy, J. A., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge, England: Cambridge University Press.
- Lupyan, G. (2007a). *The label feedback hypothesis: Linguistic influences on visual processing* (Doctoral dissertation). Carnegie Mellon University, Pittsburgh, PA.
- Lupyan, G. (2007b). Reuniting categories, language, and perception. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1247–1252). Austin, TX: Cognitive Science Society.
- Lupyan, G. (2008a). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, *108*, 566–577. doi:10.1016/j.cognition.2008.03.009
- Lupyan, G. (2008b). From chair to “chair”: A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*, 348–369. doi:10.1037/0096-3445.137.2.348
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science*, *18*, 1077–1083. doi:10.1111/j.1467-9280.2007.02028.x
- Lupyan, G., & Spivey, M. J. (2010a). Making the invisible visible: Auditory cues facilitate visual object detection. *PLoS ONE*, *5*, e11452. doi:10.1371/journal.pone.0011452
- Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*, 2236–2253. doi:10.3758/APP.72.8.2236
- Majid, A., Gullberg, M., van Staden, M., & Bowerman, M. (2007, September). *How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages*. Retrieved from <http://www.reference-global.com/doi/abs/10.1515/COG.2007.007>
- Martin, A., Ungerleider, L. G., & Haxby, J. V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of

- objects. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1023–1036). Cambridge, MA: MIT Press.
- Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science, 18*, 1007–1013. doi:10.1111/j.1467-9280.2007.02016.x
- Mitterer, H., Horschig, J. M., Musseler, J., & Majid, A. (2009). The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1557–1562. doi:10.1037/a0017019
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: The MIT Press.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 70–84. doi:10.1037/0278-7393.11.1.70
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316. doi:10.1037/0033-295X.92.3.289
- Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior, 21*, 1–20. doi:10.1016/S0022-5371(82)90412-1
- Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition, 80*, B11–B20. doi:10.1016/S0010-0277(01)00112-3
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.
- Pertsov, Y., Zohary, E., & Avidan, G. (2009). Implicitly perceived objects attract gaze during later free viewing. *Journal of Vision, 9*, 1–12. doi:10.1167/9.6.6
- Pilling, M., Wiggett, A., Ozgen, E., & Davies, I. R. L. (2003). Is color “categorical perception” really perceptual? *Memory & Cognition, 31*, 538–551. doi:10.3758/BF03196095
- Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*, 665–681. doi:10.1016/S0010-0277(07)00108-4
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160–174. doi:10.1037/0096-3445.109.2.160
- Puri, A. M., & Wojciulik, E. (2008). Expectation both helps and hinders object perception. *Vision Research, 48*, 589–597. doi:10.1016/j.visres.2007.11.017
- Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development, 73*(1).
- Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*, 977–986. doi:10.3758/BF03209345
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: Bradford Book.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439. doi:10.1016/0010-0285(76)90013-X
- 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. doi:10.1068/p5117
- Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology, 62*, 1904–1914. doi:10.1080/17470210902853530
- Sloutsky, V. M., & Fisher, A. V. (2001). Effects of linguistic and perceptual information on categorization in young children. In J. Moore & K. Stenning (Eds.), *Proceedings of the XXIII Annual Conference of the Cognitive Science Society* (pp. 946–951). Mahwah, NJ: Erlbaum.
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166–188. doi:10.1037/0096-3445.133.2.166
- Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *From many strands: Weaving a lexicon* (illustrated ed., pp. 257–294). Cambridge, MA: MIT Press.
- Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior, 23*, 3–22. doi:10.1016/S0022-5371(84)90479-1
- Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 277–311). Cambridge, MA: MIT Press.
- Spelke, E. S., & Tsivkin, S. (2001). Initial knowledge and conceptual change: Space and number. *Language acquisition and conceptual development* (pp. 475–511). Cambridge, England: Cambridge University Press.
- Stadthagen-Gonzalez, H., Damian, M. F., Pérez, M. A., Bowers, J. S., & Marin, J. (2009). Name–picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *The Quarterly Journal of Experimental Psychology, 62*, 1581–1597. doi:10.1080/17470210802511139
- Tabossi, P., & Johnson-Laird, P. N. (1980). Linguistic context and the priming of semantic information. *The Quarterly Journal of Experimental Psychology, 32*, 595–603. doi:10.1080/14640748008401848
- Thompson, R. K., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 31–43. doi:10.1037/0097-7403.23.1.31
- Thompson-Schill, S. L., Aguirre, G. K., D'Esposito, M., & Farah, M. J. (1999). A neural basis for category and modality specificity of semantic knowledge. *Neuropsychologia, 37*, 671–676. doi:10.1016/S0028-3932(98)00126-2
- Vanderwart, M. (1984). Priming by pictures in lexical decision. *Journal of Verbal Learning and Verbal Behavior, 23*, 67–83. doi:10.1016/S0022-5371(84)90509-7
- Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5*, 81–92. doi:10.1167/5.1.8
- Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press. doi:10.1037/11193-000
- Walter, E., & Dassonville, P. (2005). Semantic guidance of attention within natural scenes. *Visual Cognition, 12*, 1124. doi:10.1080/13506280444000670
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*, 829–853. doi:10.1093/brain/107.3.829
- Waxman, S. R. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233–284). Cambridge, MA: MIT Press.
- Waxman, S. R. (2004). Everything had a name, and each name gave birth to a new thought: Links between early word-learning and conceptual organization. Why is it hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *From many strands: Weaving a lexicon* (illustrated ed., pp. 295–335). Cambridge, MA: MIT Press.
- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13*, 258–263. doi:10.1016/j.tics.2009.03.006
- Waxman, S. R., & Hall, D. G. (1993). The development of a linkage between count nouns and object categories: Evidence from 15-month-old to 21-month-old infants. *Child Development, 64*, 1224–1241. doi:10.2307/1131336
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257–302. doi:10.1006/cogp.1995.1016
- Willems, R. M., Toni, I., Hagoort, P., & Casasanto, D. (2009). Body-specific motor imagery of hand actions: Neural evidence from right- and



left-handers [Abstract]. *Frontiers in Human Neuroscience*, 3, 39. doi: 10.3389/neuro.09.039.2009

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 7780–7785. doi:10.1073/pnas.0701644104

Wolff, P., & Holmes, K. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2, 253–265. doi:10.1002/wcs.104

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30, 553–566. doi:10.1037/0012-1649.30.4.553

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, 85, 223–250. doi:10.1016/S0010-0277(02)00109-9

Xu, F. (2003). Categories, kinds, and object individuation in infancy. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 63–89). Oxford, England: Oxford University Press.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 776–795. doi:10.1037/0278-7393.26.3.776

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, 49, 2095–2103. doi: 10.1016/j.visres.2009.05.017

Yoshida, H., & Smith, L. B. (2003). Known and novel noun extensions: Attention at two levels of abstraction. *Child Development*, 74, 564–577. doi:10.1111/1467-8624.7402016

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science*, 16, 90–95. doi:10.1111/j.0956-7976.2005.00787.x

Yuval-Greenberg, S., & Deouell, L. (2009). The dog’s meow: Asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research*, 193, 603–614. doi:10.1007/s00221-008-1664-6

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13, 168–171. doi:10.1111/1467-9280.00430

Appendix

Quantifying the Familiarity and Predictiveness of Verbal and Nonverbal Cues

To quantify whether the nonverbal cues used in Experiments 1 and 3 (see Table A1) had high predictive power, we recruited 17 participants from the online service Mechanical Turk (www.mturk.com; see Buhrmester, Kwang, & Gosling, 2011, for a detailed assessment) and presented them with the same nonverbal sounds used in the studies. Participants were asked to respond to the following prompt: “What object or animal typically produces this sound? Please respond using a single word in the singular.”

Table A1  
Stimuli Used in Experiments 1 and 3

Verbal label	Percentage producing target label in response to the nonverbal cue	Imagery concordance <sup>a</sup>	
		Label cues	Sound cues
<i>n</i>	17	13	13
Car (automobile)	94	3.4	3.8
Cat	100	3.6	4.2
Cow	100	3.9	4.4
Dog	100	3.6	4.5
Frog	100	4.1	4.2
Gun	59	4.4	2.3
Motorcycle <sup>b</sup>	65	3.2	3.8
Rooster	94	3.5	4.3
Train (locomotive)	88	3.2	4.2
Whistle	88	3.3	4.3

Note. All sounds can be downloaded from <http://sapir.psych.wisc.edu/stimuli/labelsSoundsExp1.zip>

<sup>a</sup> Participants answered the question “How well did the picture match the image you thought of?” using a 5-point Likert scale (1 = *did not match at all*; 5 = *matched perfectly*). <sup>b</sup> The motorcycle item was omitted from Experiments 3A–3B.

The target label was the modal response for all 10 categories. Of the nontarget responses, only one was provided by more than a single person (two people wrote “jackhammer” for motorcycle). Across all categories, an average of 15.1 out of 17 participants (89%) responded with the target category after hearing the sound (we accepted within-category answers such as shotgun for gun, “scooter” for motorcycle, “train whistle” for train, and “calf” for cow). Marking as errors responses that deviated in any way from the target labels, the mean number of participants providing the target label in response to the sound decreased, but only slightly: 14.2 out of 17 (84%).

The label advantage reported in Experiments 1A–1C was not driven by items with lower percentages of participants producing the target label in response to the sound cue. The item analysis in the text showing faster RTs following the label relative to sound cues held even for the six items, with agreement scores above 94%: Experiment 1,  $t(5) = 3.81, p = .013$ ; Experiment 1B,  $t(5) = 3.11, p = .027$ , and marginal for Experiment 1C,  $t(5) = 1.98, p = .104$ , though the label advantage for the four items with 100% agreement was reliable,  $t(3) = 4.19, p = .025$ . For these same six items, in Experiment 3A, valid labels led to significantly faster responses than did valid sounds,  $t(5) = 3.61, p = .015$  (all *ps* were two tailed). We did not conduct a parallel analysis for Experiment 3B because there was no reliable difference between valid label and valid sound cues.

The analysis above shows that the sound cues were highly informative and specific, but we sought to directly compare the informativeness of sound and label cues using a common task (it obviously does not make sense to repeat the task above with

(Appendix continues)



auditory labels). The common task we used was imagery concordance, following the procedure used by Rossion and Pourtois (2004). Participants were told that they would hear sounds (labels) produced by (referring to) common objects and animals and that they should form an image.

Separate groups of participants heard the sound or label cues and, 3 s after the offset, were shown one of the photographic exemplars of each category used in Experiments 1A–1C. The picture was displayed for 3 s. On each of the 10 trials (1 per category), participants responded to the question “How well did the picture match the image you thought of?” using a 5-point Likert scale (1 = *did not match at all*; 5 = *matched perfectly*). Before starting the task, participants were asked to imagine in their mind’s eye the object or animal making the sound or named by the label and to enter a rating depending on how well the picture they imagined matched the picture shown.

We recruited 26 new participants from Mechanical Turk and assigned them to a label-concordance or sound-concordance condition (see Table A2). There were no overall differences in concordance ratings in either a subject-based,  $t(24) = 1.23, p = .23$ , or item-based analysis,  $t(9) = 1.28, p = .23$ . The item with the greatest difference between label and sound condition ( $z = 2.7$ ) was “gun” ( $M_{\text{label concordance}} = 4.38, M_{\text{sound concordance}} = 2.31$ ). With this item excluded, the concordance between sounds and pictures was actually reliably greater than the concordance between the labels and the same pictures,  $t(8) = 6.40, p < .0005$ . This difference was marginal in a subject-based analysis,  $t(24) = 1.82, p = .08$ .

Table A2  
*Stimuli Used in Experiment 2*

Verbal label	Sound imitation cues	Verb label cues
Car	beep beep	honking
Cat	meow	meowing
Clock	tick tock	ticking
Cow	moo	mooing
Dog	arf arf	barking
Frog	ribbit ribbit	croaking
Motorcycle	vroom vroom	revving
Phone	brring brring	ringing
Rooster	cock-a-doodle-doo	crowing
Spring	boing	bouncing

*Note.* All sounds can be downloaded from <http://sapir.psych.wisc.edu/stimuli/labelsSoundsExp2.zip>

These results provide evidence that at least in an explicit (and untimed) rating task, the sounds cues were no less predictive of the category than the labels. For 9 of the 10 pictures, participants in the sound condition actually provided higher matching ratings than did participants in the label condition.

Received June 15, 2011  
Revision received June 15, 2011  
Accepted June 15, 2011 ■