Cognition 143 (2015) 93–100

Contents lists available at ScienceDirect

Cognition

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

What makes words special? Words as unmotivated cues

Pierce Edmiston *, Gary Lupyan

Department of Psychology, University of Wisconsin—Madison, United States

ARTICLE INFO

Article history:
Received 15 May 2014
Revised 28 May 2015
Accepted 16 June 2015

Keywords:
Concepts
Language
Categories
Labels
Environmental sounds

ABSTRACT

Verbal labels, such as the words "dog" and "guitar," activate conceptual knowledge more effectively than corresponding environmental sounds, such as a dog bark or a guitar strum, even though both are unambiguous cues to the categories of dogs and guitars (Lupyan & Thompson-Schill, 2012). We hypothesize that this advantage of labels emerges because word-forms, unlike other cues, do not vary in a *motivated* way with their referent. The sound of a guitar cannot help but inform a listener to the type of guitar making it (electric, acoustic, etc.). The word "guitar" on the other hand, can leave the type of guitar unspecified. We argue that as a result, labels gain the ability to cue a more abstract mental representation, promoting efficient processing of category members. In contrast, environmental sounds activate representations that are more tightly linked to the specific cause of the sound. Our results show that upon hearing environmental sounds such as a dog bark or guitar strum, people cannot help but activate a particular instance of a category, in a particular state, at a particular time, as measured by patterns of response times on cue-picture matching tasks (Exps. 1–2) and eye-movements in a task where the cues are task-irrelevant (Exp. 3). In comparison, labels activate concepts in a more abstract, decontextualized way—a difference that we argue can be explained by labels acting as "unmotivated cues".

0. Introduction

Consider how we recognize rain. We can feel its coldness on our skin. We can tell it is raining by the glistening pavement. We can also recognize rain by the sound of drops falling onto a roof or bouncing off a sidewalk and in the words of Ruth Millikan (1998), "falling on English speakers, here is another way [rain] can sound: 'Hey, guys, it's raining!'" (p. 64).

The concept of rain can be activated by inputs from multiple perceptual modalities. Concepts can also be activated via language. We present four experiments aimed at contrasting these two ways of activating familiar concepts: through nonverbal environmental sounds, and through auditory verbal labels. Do these cues activate the same knowledge—two pathways to the same concept of rain? And if not, why not?

The question of whether verbal and nonverbal cues activate concepts differently was studied by Lupyan and Thompson-Schill (2012) who found that compared to words such as "dog," equally familiar and unambiguous environmental sounds such as a <dog bark>[1] resulted in consistently slower recognition of subsequently presented pictures of dogs. This "label-advantage" persisted for delays as long as 1.5 s after cue offset for familiar categories and was also observed for new categories of "alien musical instruments" for which participants learned either names or corresponding sounds—evidence that the advantage did not arise from the sound cues being less familiar or more difficult to process. We thus have a situation where, equating for overall associative strength, labels nevertheless activate concepts more effectively than nonverbal cues. This special status of labels is taken by some to be a given because while nonverbal cues like dog-barks can be thought of as simply associative, "words refer; they do not merely associate" (Waxman & Gelman, 2009, p. 259). An alternative is that reference is born from associations, and what is special about the word-referent relationship is the unique pattern of associations it implements. We argue that a critical distinction between verbal labels and other cues to category membership is that dog barks, the sound of rain—indeed *any* nonverbal cue—is that such cues vary in a lawful way with properties of the events that caused them.[2] In contrast, words do not covary with their referents. This seemingly small difference may go far in explaining the special status of labels.

Adopting terminology from semiotics (e.g., Kockelman, 2005), we call environmental sounds "motivated cues" where motivation

* Corresponding author at: Department of Psychology, University of Wisconsin—Madison, 1202 W. Johnson St., Madison, WI 53706, United States.
E-mail address: pedmiston@wisc.edu (P. Edmiston).

[1] We encase environmental sounds in <chevrons> and use quotes to indicate spoken labels.

[2] Environmental cues are well described as Peircean *indices*, being causally linked with specific objects or events, occurring at specific times (Atkin, 2005).

http://dx.doi.org/10.1016/j.cognition.2015.06.008
0010-0277/

describes a principled/predictable relationship between different aspects of an object, or the relationship between its form and function. The three dogs depicted in Fig. 1 all bark, but the barks vary in predictable ways with the dog, e.g., larger dogs have lower-pitched barks. People have surprisingly rich knowledge of such motivated relationships. For example, people can determine the length, shape, and material of hidden objects from the sounds they make when dropped or hit (Carello, Anderson, & Kunkler-Peck, 1998; Kunkler-Peck & Turvey, 2000). In contrast, the acoustic form of a word *does not* lawfully vary with aspects of the exemplar to which it refers. Although individual tokens of the word "dog" are informative of features of the speaker (i.e., we can use the surface properties to infer the speaker's gender, age, etc.), people do not normally say "dog" in a deeper voice when referring to a German Shepherd or elongate it to mean a Dachshund (Taylor, 2012) [3]. As depicted in Fig. 1, any instance of "dog" can be used to refer to any and all dogs. On the terminology we have adopted, words are "unmotivated cues."

We do not simply mean that words are arbitrary, i.e. that "dog" refers to dogs by convention (cf. Hockett, 1966). People need to learn that dogs bark just as English-speakers need to learn that dogs are called "dogs"—both of these mappings are arbitrary as far as learners are concerned. However, while individual barks vary lawfully with their causes, tokens of the word "dog" transcend such within-category variability. In so doing, the word "dog" sheds idiosyncrasies of particular category exemplars, activating the concept of dog—mental states related to dogs—in a more abstract and categorical way. Preliminary support for this account comes from the item analyses conducted by Lupyan and Thompson-Schill (2012) in which labels were shown to not simply activate concepts more quickly, but acted to more selectively activate features most diagnostic of category membership.

Here, we provide initial tests of the hypothesis that sounds fail to activate concepts as effectively as category labels because sounds are motivated cues to a particular source. We address this question by manipulating the relationship between sound cues and their likely causes to determine the circumstances that give rise to the label-advantage. By demonstrating that environmental sound cues are limited in activating conceptual knowledge by their relationship with a particular source, we hope to better understand the degree to which language provides a unique way of manipulating mental content.

In Experiments 1A–1B we test the hypothesis that compared to labels, environmental sounds activate category exemplars corresponding to a likely sound source even when participants are tasked with treating environmental sounds as category-level cues. In Experiment 2 we test whether in addition to exemplar-level congruence (e.g., the sound of an electric guitar indexing an electric rather than an acoustic guitar), sounds activate visual information corresponding to the sources of the sound *at a particular time*. In Experiment 3, we compare the effects of hearing task-irrelevant category label and environmental sound cues on within-category individuation using eye-tracking, a more implicit measure.

1. Experiments 1A and 1B

In Experiments 1A–B we use a picture-verification task modeled after Lupyan and Thompson-Schill (2012). In this task, people need to recognize the basic-level category of a picture as quickly and accurately as possible. We hypothesized that environmental sounds used as cues will lead to slower recognition than label cues because sounds necessarily cue a more specific category exemplar and thus are not as effective cues for an entire category of objects as category labels. We tested this hypothesis by manipulating the within-category specificity of the mapping between the cues and picture targets.

1.1. Methods (Exp. 1A)

1.1.1. Participants

43 University of Wisconsin–Madison undergraduates participated in Experiment 1A in exchange for course credit. Sample sizes were on the same order as those from a previous study utilizing a similar paradigm (Lupyan & Thompson-Schill, 2012).

1.1.2. Materials

The auditory cues comprised basic-level category labels and environmental sounds for six categories: *bird, dog, drum, guitar, motorcycle,* and *phone*. For each category, we obtained two distinct environmental sound cues, e.g., <classical guitar strum>, <electric guitar strum>, and two separate images for each subordinate category, e.g., two electric guitars for <electric guitar strum>, two acoustic guitars for <electric guitar strum>. To control for cue variability, we also used two versions of each spoken category label: one pronounced by a female speaker, one by a male speaker. All auditory cues were equated in duration (600 ms.) and normalized in volume. The images were color photographs (four images per category). The materials, obtained from online repositories, are available for download at http://sapir.psych.wisc.edu/stimuli/MotivatedCuesExp1A-1B.zip.

1.1.3. Procedure

On each trial participants heard a cue and saw a picture. We instructed participants to decide as quickly and accurately as possible if the picture they saw came from the same basic-level category as the word or sound they heard Participants were tested in individual rooms sitting approximately 24″ from a monitor such that images subtended $\sim 10 \times 10^{\circ}$. Trials began with a 250 ms. fixation cross followed immediately by the auditory cue, delivered via headphones. The target image appeared centrally 1 s after the offset of the auditory cue and remained visible until a response was made. Each participant completed 6 practice and 384 test trials. If the picture matched the auditory cue (50% of trials) participants were instructed to respond 'Yes' on a gaming controller (e.g., <cellphone ring> or "phone" followed by a picture of any phone). Otherwise, they were to press 'No' (e.g., <cellphone ring> or "phone" followed by a dog). All factors (cue type, congruence) varied randomly within subjects. Auditory feedback (buzz or bleep) was given after each trial.

1.2. Results of Experiment 1A

All participants performed very accurately on all items ($M = 97\%$). Response times (RTs) shorter than 250 ms. or longer than 1500 ms. were removed (292 trials removed, 1.77% of total).

We fit RTs for correct responses on matching trials ('Yes' responses) with linear mixed regression using maximum likelihood estimation (Bates, Maechler, Bolker, & Walker, 2013), including random intercepts and random slopes for within-subject factors and random intercepts for repeated items (unique trial types) following the recommendations of Barr, Levy, Scheepers, and Tily (2013). Reported below are the parameter estimates (b) and confidence intervals for each contrast of interest. Significance tests were calculated using chi-square tests that compared nested models—models with and without the factor of interest—on improvement in log-likelihood.

[3] There is some evidence that speakers modulate pronunciations of words in a graded fashion, e.g., speaking faster or slower in proportion to the speed of the object they are describing (Shintel & Nusbaum, 2007; Shintel, Nusbaum, & Okrent, 2006). Such instances of motivated mappings in language may be viewed as the rest of the world "bleeding through" (Lupyan & Casasanto, 2012).

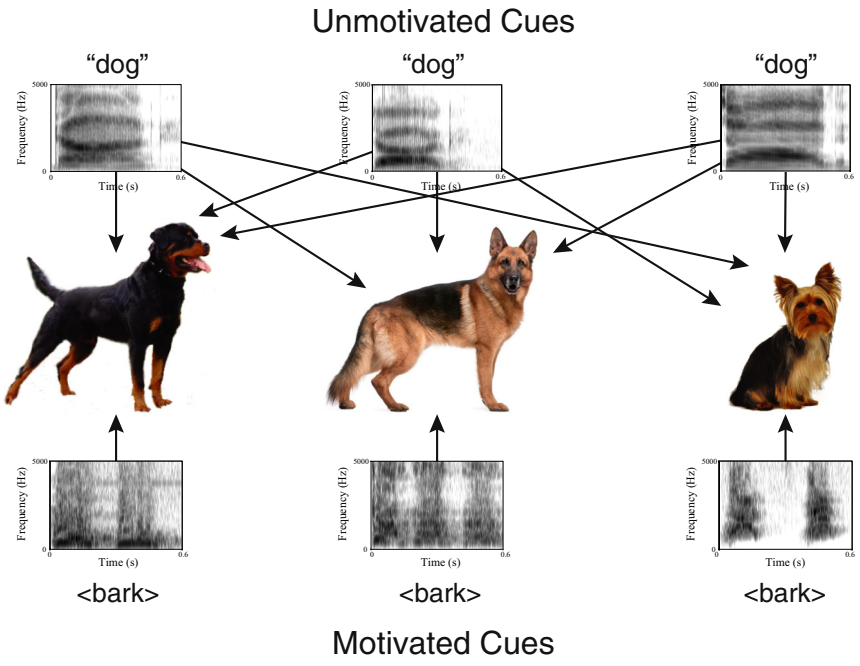

Fig. 1. Schematic of motivated and unmotivated cues to the concept dog. Motivated cues such as dog barks provide information about an individual category exemplar. In contrast, the form of the word "dog" abstracts away from the particulars: any "dog" can refer to any dog. Words are "unmotivated" in that their surface forms do not covary lawfully with properties of their referents.

Results for Experiment 1A is shown in Fig. 2. All 43 participants were faster to verify category matches when cued with a verbal label than when cued with an environmental sound. On the environmental sound trials, participants verified images that better matched the likely source of the sound (*congruent* sound trials) more quickly than on *incongruent* sound trials, $b = -38.71$ ms., 95% CI [−55.33, −22.10], $\chi^2(1) = 17.61$, $p < 0.001$. Critically, when cued by a label, participants responded faster than even the sound congruent trials, $b = -25.53$ ms., 95% CI [−40.89, −10.17], $\chi^2(1) = 9.96$, $p = 0.002$.

1.3. Rationale for Experiment 1B

One possible confound present in Experiment 1A is that the presence of incongruent sound trials (e.g., hearing a <cell phone ring> and responding 'Yes' if a rotary phone was presented) may have led participants to be more careful on these trials, inflating the RTs. In order to test this possibility, we conducted another study (Experiment 1B), identical to Experiment 1A except for omitting incongruent sound trials.

1.4. Methods (Exp. 1B)

1.4.1. Participants

25 University of Wisconsin–Madison undergraduates participated in Experiment 1B in exchange for course credit.

1.4.2. Materials

The same materials were used in Experiment 1B.

1.4.3. Procedure

Trials were structured exactly as in Experiment 1B except for excluding the incongruent sound trials. That is, if participants were cued with a <cellphone ring> they responded 'Yes' to a picture of a cellphone but never had to respond 'Yes' to a picture of a rotary phone.

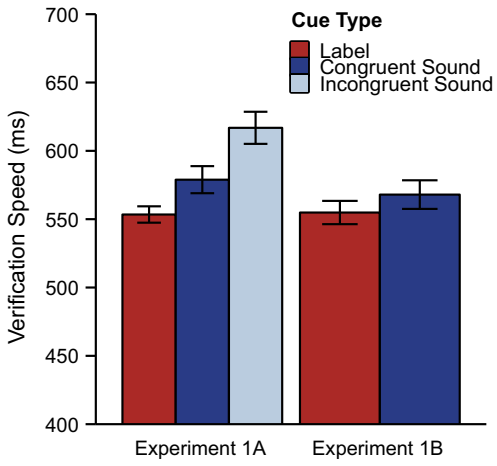

Fig. 2. Picture verification performance in Experiments 1A and 1B in response to sound and label cues (correct "Yes" responses only). Error bars signify within-subject 95% CIs (Morey, 2008).

1.5. Results of Experiment 1B

Participants again performed very accurately on all items ($M = 96\%$). Response times (RTs) shorter than 250 ms. or longer than 1500 ms. were removed (101 trials, 1.05% of total).

Data were analyzed in the same manner as Experiment 1A. The advantage of labels over congruent sounds observed in Experiment 1A was replicated in Experiment 1B, $b = -20.03$ ms., 95% CI [−32.40, −7.66], $\chi^2(3) = 10.53$, $p = 0.015$ (Fig. 2). 22 of the 25 participants showed this effect. That removing the incongruent sound trials did not eliminate the label advantage suggests that the presence of the incongruent sound trials in Experiment 1A was not causing participants to slow down on all sound trials.

1.6. Discussion of Experiments 1A and 1B

There are many ways to activate knowledge. The results of Experiments 1A–1B show that labels are more effective than sounds at activating knowledge in a *categorical* way. Environmental sounds can—and as evident from the results of Experiments 1A–1B—do activate highly specific knowledge. People are faster to recognize an electric guitar after hearing an electric guitar strum and faster to recognize an acoustic guitar after hearing an acoustic guitar strum. But when the task is simply recognizing any guitar as a guitar, hearing the verbal label "guitar" is more effective still. We have argued that what makes labels more effective than sounds at activating such categorical states is that labels are unmotivated—the word "guitar" denotes an entire class of guitars. There is no analogous relationship for environmental sounds (or for that matter, any perceptual cue) because, we argue, such cues are *motivated*—their form is correlated with the entity with which they are associated or which they denote.

In Experiment 1 we manipulated the level of congruence between environmental sounds and their causes, showing that the label advantage is obtained even when the picture and the sound are well-matched. In Experiment 2 we go one step further by investigating the temporal relationship between the two cue types and their referents. The goal of Experiment 2 was to investigate whether the difference in knowledge cued by verbal labels and environmental sounds derives in part from the tighter temporal correspondence of sounds and their sources compared to labels and their referents.

2. Experiment 2

If we hear a bird chirp, it is likely that an actual bird is chirping nearby. In comparison, we may hear "bird" in the absence of actual birds (the so-called "displacement" property of language, Hockett, 1966). The psychological reality of the temporal link between environmental sounds and their sources is supported by studies showing that hearing sounds improves visual detection of category members when the sound and visual target are presented simultaneously (Chen & Spence, 2010).

In Experiment 2 we predicted that sounds could serve as more effective cues (perhaps even as effective as labels) when two conditions were met: (1) the visual image following the nonverbal sound cue depicts the object in a "sound-making" state, e.g., not just a dog following a <bark>, but a dog with an open mouth (cf. Zwaan, Stanfield, & Yaxley, 2002), and (2) the sound and visual image occur simultaneously.

Note that if the difference between sounds and labels were simply a difference in overall amount of experience or familiarity with the auditory cues, we would not expect the cue-to-image offset (temporal congruence) to affect labels and sounds differently. If, however, sounds act as indices, activating specific category exemplars (a *specific* dog, a *specific* guitar) at a specific time (close to occurrence of the sound), people's recognition performance following label and sound cues may indeed depend on the temporal relationship between the cue and its referent.

2.1. Methods

2.1.1. Participants

42 University of Wisconsin–Madison undergraduates (29 female; 18–22 years old) participated in Experiment 2 in exchange for course credit. The data from one participant was incomplete due to experimenter error and was removed from analysis. The stopping criterion was the same as in Exps. 1A–1B; participant recruitment was uninformed by the data.

2.1.2. Materials

Participants were presented auditory cues and color images from familiar animal and artifact categories (*baby, bee, bowling ball, car, cat, chainsaw, dog, keyboard,* and *scissors*). Auditory cues comprised a spoken label (pronounced by a female speaker) and an environmental sound for each category. As in Experiments 1A–1B all cues were matched in duration (600 ms.) and normalized in volume. We gathered images for each category that varied in how well they resembled the likely source of an environmental sound. Continuous sound-image congruence ratings were obtained from participants recruited online through Amazon's Mechanical Turk. Participants ($N = 20$) listened to each environmental sound and rated the congruence between the sound and each image (8–10 per category) on a 5-point Likert scale. To control for a possible confound between sound-image congruence and category typicality in our materials, we recruited an additional group of participants ($N = 22$) to rate the same images on typicality (e.g., "How typical is this dog compared to dogs in general?"). In selecting the final images we sampled across the typicality range (correlation between sound-image congruence and category typicality: Pearson's $r = 0.27$). All sounds, images, and typicality/congruence ratings are available for download at http://sapir.psych.wisc.edu/stimuli/MotivatedCuesExp2-3.zip.

2.1.3. Procedure

The picture verification task used in Experiment 2 was similar to Experiments 1A–1B with two key changes. On half the trials, the auditory cue and picture were presented simultaneously. On the remaining trials, there was a 400 ms. delay between auditory cue offset and picture onset. Second, there was now only a single auditory cue per category (a single instance of the word "dog", and a single <bark>). The images varied in congruence with the environmental sound cue, but now in a more subtle and graded way rather than the congruent/incongruent distinction used in Experiment 1A (see inset of Fig. 3). Note that because there was now only a single environmental and a single verbal cue for each category, there was even less room for ambiguity in what constituted a matching response (cf. Exp. 1A). Participants completed 6 practice trials and 426 test trials. The picture matched the cue 75% of the time. We increased the percentage of matching trials in Experiment 2 to allow full counterbalancing of within-subject variables while limiting the length of the experiment to 30 min. Auditory feedback (buzz or bleep sound) was given after each trial.

2.2. Results

All participants performed very accurately on all items ($M = 97\%$), except responses to the environmental sound cue for scissors cutting paper ($M = 91\%$). Over half of the participants (24 out of 41) reported difficulties with the scissors sound cue during debriefing and we therefore excluded these trials from analysis for all participants (<5% of total). As in Experiments 1A and 1B, response times (RTs) shorter than 250 ms. or longer than 1500 ms. were removed (220 trials, 1.33% of total). RTs for correct responses on matching trials were fit with linear mixed regression as in Experiments 1A and 1B. Random intercepts and random slopes for the effect of cue type were allowed to vary by subject, with by-item random intercepts for repeated items (unique combinations of auditory cue and image target across subjects).

The results for the two cue conditions, split by cue-to-image delays are shown in Fig. 3. Participants were much faster to verify category members on the delayed than simultaneous trials, as expected if the additional delay provided extra time to activate visual knowledge, which aided in the recognition of the picture, $b = -181.13$ ms., 95% CI $[-193.57, -168.70]$, $p \ll 0.001$. At both delay periods, participants verified images more quickly when

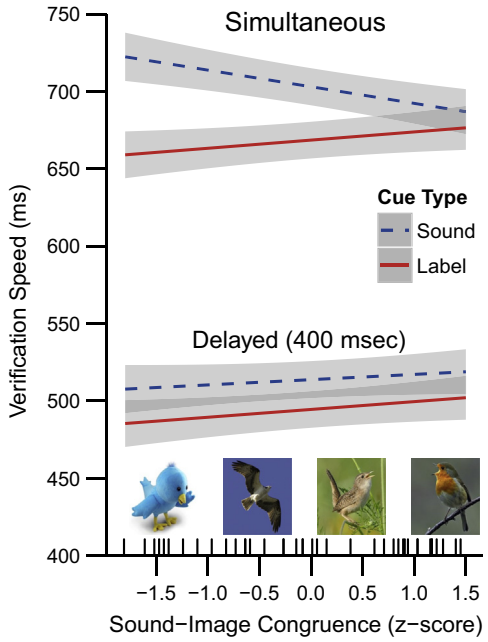

Fig. 3. Verification RTs from Experiment 2. Linear model estimates control for the effect of category typicality of the target images. Shaded confidence bands signify ±1 SE of the predicted RT to verify images as a function of the picture's rated congruence with the natural sound. Rug plot ticks indicate the raw sound-image congruence ratings for the 40 images used in Experiment 2. Sample images display increasing sound-image congruence for category *bird*.

cued by labels than when cued by environmental sounds, $b = -26.72$ ms., 95% CI [−40.49, −12.95], $\chi^2(1) = 13.26$, $p < 0.001$. However, the label-advantage depended in a theoretically meaningful way on the delay and sound-image congruence of the image (Fig. 3). Critically, the label-advantage was abolished when the image was highly congruent with the sound cue, *but only when* the cue and image were presented simultaneously. This qualitative pattern was supported by a three-way interaction between cue type, sound-image congruence, and delay period such that the label-advantage was the smallest when participants verified images that most resembled the source of the environmental sound *and* were presented at the same time as the auditory cue, $b = -14.41$, 95% CI [−26.23, −2.59], $\chi^2(1) = 5.17$, $p = 0.017$.

As expected, participants recognized more typical category members faster than less typical members, $b = -9.76$ ms., 95% CI [−17.23, −2.29], $\chi^2(1) = 6.30$, $p = 0.012$. In order to control for this effect of overall typicality, we included the normalized typicality ratings (*z*-scores) for each image as a covariate in all models reported above.

2.3. Discussion of Experiment 2

In Experiment 2 we tested the hypothesis that the label-advantage should depend both on the level of congruity between the nonverbal sound and the visual image (a bird with a visibly open mouth should be a better match to a tweeting sound than bird in a non-singing posture: Fig. 2), *and* on the timing between the auditory cue and the appearance of the to-be-recognized image. We expected that the label advantage would be reduced when the visual item indexed by the auditory cue was highly congruent *and* the auditory cue and visual target appeared simultaneously. This is precisely what we found. This result adds support to our claim that what distinguishes verbal and nonverbal cues is *how* they are linked with the outside world and not that words are simply more familiar than the environmental sound cues we used in the experiments. Although both a <bird chirp> and "bird" are equally valid cues to birds in our study, the word "bird" activates a representation that abstracts over idiosyncratic states (e.g., whether the bird has its mouth open) and temporal coincidence (whether the bird is present at the time of cue onset). In contrast, environmental sounds appear to activate a more concrete and contextualized category representation.

3. Experiment 3

In the experiments above, we used a picture-recognition task to make inferences about the kind of information activated by the different cue types. In Experiment 3 we undertake a stronger test of the hypothesis that labels activate a more categorical representation of the denoted entities while environmental sounds activate more specific category exemplars in the here and now.

Do environmental sounds and verbal labels lead people to look at different category exemplars even when the task does not require overt categorization of the items? If labels and environmental sounds activate representations of the denoted entities in a different way, one might expect that the difference would be manifested even in the absence of requiring participants to overtly categorize the visual targets. For example, verbal labels have been shown to attract attention toward category members even when completely redundant or task irrelevant (Lupyan, 2008; Lupyan & Spivey, 2010; Salverda & Altmann, 2011). We thus sought to find out if the difference between label and sound cues was reflected in spontaneous eye movements in the absence of overt categorization and when the cues were task-irrelevant.

3.1. Methods

3.1.1. Participants

13 University of Wisconsin–Madison undergraduates (6 female; 18–20 years old) participated in Experiment 3 in exchange for course credit. Sample size was approximated based on previous work using task-irrelevant cues (Salverda & Altmann, 2011). Data analysis did not inform participant recruitment.

3.1.2. Materials

Participants in Experiment 3 were presented the same materials used in Experiment 2: auditory cues (category label, environmental sound) and four images for each of 10 categories of familiar animals and artifacts. Eye movements were tracked with an EyeLink 1000 Desktop unit (SR Research Ltd.). The eye tracker was calibrated at the beginning of the experiment, after the practice trials, and after every 20 test trials if necessary.

3.1.3. Procedure

Trials began with a central fixation cross that turned green when fixated, indicating to the participant that they could begin the trial by clicking on the cross. After the fixation screen, four images arranged on an invisible 2 × 2 grid appeared on the screen. *All images presented on a given trial were from the same category* (e.g., four pictures of dogs). Any category exemplar could appear at any of the four locations in the grid. Simultaneously with the onset of the four images, participants heard a cue—a label or nonverbal sound—which always matched the category. After approximately 2 s., a randomly selected image was highlighted with a thin green frame. Participants' task was to simply click on the image within the frame. The four images on each trial subtending ∼11 × 11° were arranged in a 2 × 2 grid with the centers of adjacent images separated by 19°. Participants sat unrestrained

approximately 22″ from the monitor. Each participant completed 10 practice trials and 160 test trials.

3.2. Results

We analyzed eye movements between the onset of the images and cue and before the appearance of the target frame that identified which image was to be clicked. Our key measure was the sound-image congruence of the images people looked at prior to the appearance of the target frame. Trials in which no images were fixated before the onset of the target frame (e.g., when the participant continued to fixate the central cross) were considered uninformative and excluded from analyses (192 trials, 9.6% of total). We first analyzed the time it took for participants to click on the target frame once it appeared. These did not vary by cue type, M_{label} = 1098 ms, M_{sound} = 1124 ms, $b = -19.91$ ms., 95% CI [−68.15, 28.33], $\chi^2(1) = 0.66$, p = 0.42. An examination of the total image fixation time showed that it also did not vary by cue-type, $b = 3.70$ ms., 95% CI [−6.30, 13.69], $\chi^2(1) = 0.51$, $p = 0.47$. These analyses suggest that participants' overall behavior was similar regardless of which cue they heard and that the cues—as expected—did not affect people's overall ability to detect the onset of the frame.

We next turned to the critical analyses of the effect of cue on patterns of looking to each image. To quantify which images participants were fixating, we ordered the four images presented on each trial in terms of increasing sound-image congruence ratings. The proportion of fixations as a function of this rating is displayed in Fig. 4. Participants fixated the image highest in sound-image congruence for longer than the other three images in the trial *only* when they heard an environmental sound, $b = 50.37$ ms., 95% CI [28.20, 72.54], $\chi^2(1) = 19.82$, $p < 0.001$ (inset of Fig. 4). We found no significant differences in fixations among the remaining three images on each trial, $\chi^2(2) = 2.58$, $p = 0.28$. On the label trials, participants split their fixations roughly equally across the four images. Recall that in all cases, the auditory cue was not relevant to the task.

As a final analysis, we examined the effect of cue type on the average sound-image congruence rating of fixated images. Participants looked at images that were on average higher in sound-image congruence rating after hearing a sound, $b = 0.09$, 95% CI [0.02, 0.16], $p = 0.02$, but not after hearing a label, $b = -0.01$, 95% CI [−0.06, 0.03] (see Fig. 5). That is, on sound trials participants tended to fixate images that better resembled the source of that sound. The sound-congruence-by-cue-type interaction for this aggregate analysis was marginal, $b = 0.10$, 95% CI [0.00, 0.21], $\chi^2(1) = 3.40$, $p = 0.065$.

3.3. Discussion of Experiment 3

The results of Experiment 3 showed that hearing task-irrelevant environmental sounds led participants to look at different category exemplars compared to hearing similarly task-irrelevant verbal labels. As predicted, nonverbal auditory cues led participants to quickly fixate on the most likely source of the sound, even though the picture had only a 25% change of being the target. In contrast, verbal labels led to a more even distribution of fixations (see also Salverda & Altmann, 2011). Importantly, the results of Experiment 3 further show that environmental sounds are not simply less familiar cues than verbal labels. Environmental sounds changed the pattern of visual fixations in a search task in a way that is not explainable by these cues being less effective cues to the same underlying concept as activated by verbal labels.

One noteworthy aspect of the results is that unlike Experiment 2 where we saw a graded effect of sound-to-image congruence on reaction times, in the present study the effects of sound-to-image congruence were more discrete. When all the images were visible at the same time, participants were more likely to look at the image that was the *most* congruent with the sound (i.e., the most likely source). This consistent advantage for the most sound congruent images further highlights the extent to which environmental sounds appear to activate a representation of specific category exemplars in the here and now.

4. General discussion

Verbal labels activate conceptual knowledge more effectively than nonverbal cues. In Experiments 1A–1B we showed that the label-advantage persists even when the environmental cues are matched to the visual targets at a subordinate level. At first this seems puzzling. After all, a <Yorkie bark> conveys *more* information than the word "dog," yet people are nevertheless faster to recognize a Yorkie as a dog after hearing "dog" than hearing <Yorkie bark>. On our account, the label-advantage is obtained *precisely because* labels are detached from idiosyncracies of specific category members, and thereby able to selectively activating the features/dimensions most diagnostic of the named category, i.e., most useful in distinguishing between dogs and non-dogs (Lupyan, 2012). While a dog-bark or a guitar strum are both readily recognizable, the information conveyed by these environmental cues is necessarily specific and people appear unable to fully detach environmental sounds from perceptual details corresponding to their causes.

As shown by Experiment 2, the only circumstance in which sounds and labels led to equally fast and accurate categorization was when the cues were placed in an indexical relationship with the referent by (1) using pictures that maximally matched the likely cause of the sound and (2) presenting the sound and picture simultaneously. Experiment 3 extended our findings by showing that environmental cues continued to activate highly specific category instances even when they were not relevant to the task.

A natural objection to findings of a label advantage in image recognition is that it is caused by a difference in cue familiarity or processing difficulty. On this view, people recognize a picture of a dog faster when cued by "dog" than when cued by <bark> because "dog" is simply a more familiar cue than a barking sound. This is a perfectly plausible explanation, but it cannot explain past and present findings for five reasons. First, Lupyan and Thompson-Schill (2012) showed that the difference between labels and sounds is not predicted by differences in familiarity or ambiguity, is observed when sounds and labels are equated on these factors, and does not decline even after hearing/seeing a specific cue-image pairings 50 times over the course of the experiment. Second, a difference in familiarity (or difficulty in processing) would predict that if the cue-to-image interval is long enough, the ostensibly more difficult-to-process cue to "catch up", i.e. the label advantage would disappear. Yet it does not. In fact, it becomes larger, as expected if labels and sounds activate *different* representations (Lupyan, 2012). Third, the label advantage can be observed for novel categories. After learning novel labels and sounds for novel objects, the labels are more effective category cues than sounds even when people have been exposed to the (novel) labels and sounds in equal amounts and both are equally well-learned. Fourth, as shown in the present work (Experiment 1A–1B), the label-advantage occurs even when the sounds are arguably *more* informative of the upcoming image than the labels. Fifth, in a recent ERP study (Boutonnet & Lupyan, in press) we have examined the electrophysiological responses to pictures appearing after matching or mismatching verbal cues and nonverbal environmental sounds. If the difference between verbal and nonverbal cues were due to the latter being more difficult to integrate with

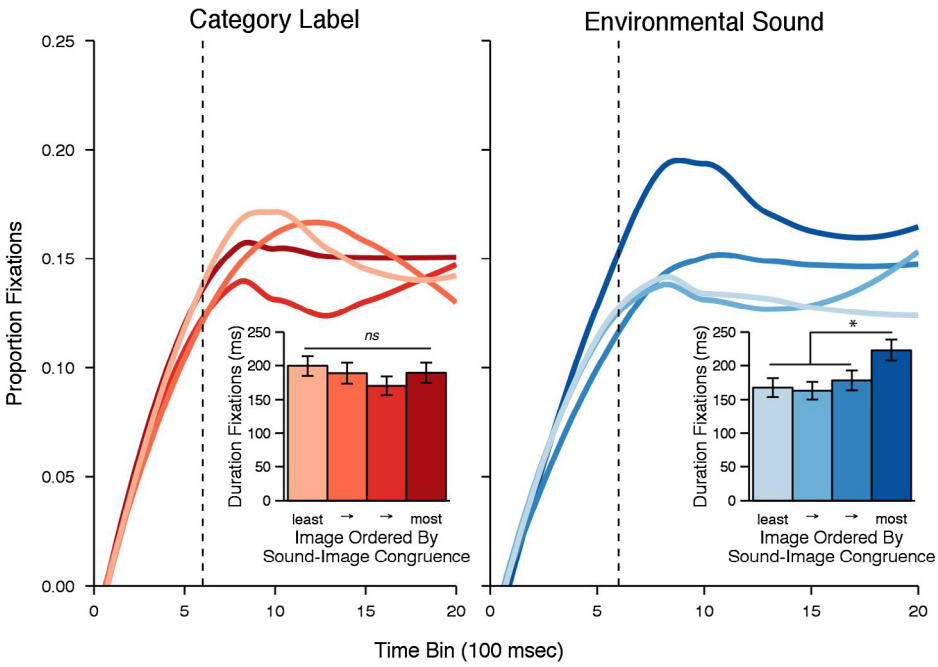

Fig. 4. Proportions and summed duration of fixations to the four images presented on each trial. Line plots display proportion of fixations to each image fit with loess regression. The dashed vertical line marks the offset of the auditory cue. The inset plots display the summed duration of fixation to each of the four images prior to target onset. Error bars signify within-subject 95% CIs.

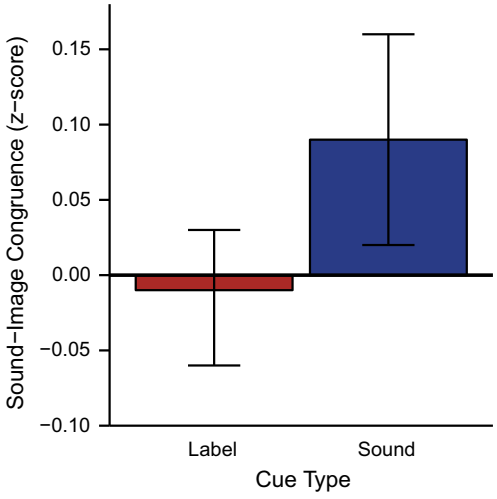

Fig. 5. Average sound-image congruence of images fixated before onset of the target frame. Average sound-image congruence of 0 indicates all images within a category were fixated equally. Error bars signify 95% CIs of the average sound-image congruence for each cue type, controlling for between-participant variance.

the picture, we should expect a smaller N400 response (Kutas & Federmeier, 2011) on mismatching nonverbal-cue trials compared to mismatching verbal-cue trials. We obtained no difference at all in the N400. Instead, labels elicited a larger P1 response which correlated with categorization behavior and distinguished between matching and non-matching trials within 100 ms of picture onset. These results not only help to rule out differences in cue familiarity, but suggest a mechanism by which labels cue categorical states: labels act as perceptual priors, activating perceptual representations that allow for more effective categorization of subsequently presented images.

We believe that a useful way of thinking about differences between verbal and nonverbal cues is in the degree to which they are motivated. The acoustic properties of environmental sounds *and, arguably, surface properties of all other nonverbal cues* vary lawfully with the properties of the events with which they are associated, that is, these cues are *motivated*. In contrast, verbal labels become dissociated from particular category instances—*are unmotivated*—and thus gain the power to construct more decontextualized and categorical conceptual representations that abstract over idiosyncratic properties in favor of those that are most diagnostic of category membership (Lupyan, 2012).

The role of language in cognition has been a popular topic of speculation (Cassirer, 1962; Clark, 1998; Leavitt, 2011; Whorf, 1956 for historical overview) and a growing body of empirical studies now support the idea that human cognition is, to an important degree, language-augmented cognition (Boroditsky, 2010; Lupyan, 2012). Our findings show that despite all participants being highly familiar with dogs, telephones, etc., the nature of the activated concept varied systematically depending on *how* it was activated. Specifically, environmental sounds stubbornly activated particular exemplars while labels appeared to activate concepts in a more categorical, abstract way.

To summarize: While language can be used to convey very specific information (“a warm drizzle on a rural street,” “the eagle in the sky”; cf. Millikan, 1998; Zwaan et al., 2002), language can also be used in the abstract: “rain”, “eagle”. The real world only contains specific instances of rain, eagles, and yappy Yorkies. The linguistic world transcends this tyranny of the specific, helping to construct a more abstract, categorical mental landscape.

The more categorical representations constructed by language have a range of cognitive uses. By abstracting over idiosyncrasies, language may be instrumental in category based induction (Sloutsky & Fisher, 2004), cross-category comparisons (Ozçalişkan, Goldin-Meadow, Gentner, & Mylander, 2009) and relational reasoning (Loewenstein & Gentner, 2005). More than a pointer to a category, labels may help in constructing the categories during development (Nazzi & Gopnik, 2001; Waxman, 2004; Yoshida & Smith, 2005), and continue to play a similar role in adulthood (Lupyan, 2009; Lupyan, Rakison, & McClelland, 2007).

Acknowledgments

The study was partially funded by NSF #1331293 to G.L. We thank Sol Simpson for help in developing the eye-tracking code. We also thank Callie Porter-Borden, Martin Potter, Oliver Roe, and Kim Knisely for data collection.

References

Atkin, A. (2005). Peirce on the index and indexical reference. *Transactions of the Charles S. Peirce Society, 41*, 161–188.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-2. <https://lme4.r-forge.r-project.org/>.

Boroditsky, L. (2010). How the languages we speak shape the ways we think: The FAQs. In M. J. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 615–632). Cambridge: Cambridge University Press.

Boutonnet, B., & Lupyan, G. (in press). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science, 9*(3), 211–214.

Cassirer, E. (1962). *An essay on man: An introduction to a philosophy of human culture*. Yale University Press.

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114*(3), 389–404.

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 162–183). New York, NY: Cambridge University Press.

Hockett, C. F. (1966). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of language* (2nd ed., pp. 1–29). Cambridge, MA: The MIT Press.

Kockelman, P. (2005). The semiotic stance. *Semiotica, 157*, 233–304.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 279–294.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Leavitt, J. (2011). *Linguistic relativities: Language diversity and modern thought*. Cambridge University Press.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*(4), 315–353. http://dx.doi.org/10.1016/j.cogpsych.2004.09.004.

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108*(2), 566–577.

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review, 16*(4), 711–718. http://dx.doi.org/10.3758/PBR.16.4.711.

Lupyan, G. (2012). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 57, pp. 255–297). Academic Press. <http://www.sciencedirect.com/science/article/pii/B9780123942937000078>.

Lupyan, G., & Casasanto, D. (2012). The meaning of nonsense words. In T. C. Scott-Phillips, M. Tamariz, & E. Cartmill (Eds.), *Proceedings of the 9th international conference* (pp. 490–492). Kyoto, Japan: World Scientific.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science, 18*(12), 1077–1082.

Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics, 72*(8), 2236–2253. http://dx.doi.org/10.3758/APP.72.8.2236.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology-General, 141*(1), 170–186. http://dx.doi.org/10.1037/a0024904.

Millikan, R. G. (1998). A common structure for concepts of individuals, stuffs, and real kinds: More mama, more milk, and more mouse. *Behavioral and Brain Sciences, 21*, 55–100.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*(2), 61–64.

Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization? *Cognition, 80*(3), B11–B20.

Ozçalişkan, S., Goldin-Meadow, S., Gentner, D., & Mylander, C. (2009). Does language about similarity play a role in fostering similarity comparison in children? *Cognition, 112*(2), 217–228. http://dx.doi.org/10.1016/j.cognition.2009.05.010.

Salverda, A. P., & Altmann, G. T. M. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology. Human Perception and Performance, 37*(4), 1122–1133. http://dx.doi.org/10.1037/a0023101.

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition, 105*(3), 681–690. http://dx.doi.org/10.1016/j.cognition.2006.11.005.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language, 55*(2), 167–177. http://dx.doi.org/10.1016/j.jml.2006.03.002.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology-General, 133*(2), 166–188.

Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford University 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. In *From many strands: Weaving a lexicon* (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*(6), 258–263.

Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16*(2), 90–95.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171. http://dx.doi.org/10.1111/1467-9280.00430.