

UNIVERSITY OF WISCONSIN—MADISON

DEPARTMENT OF PSYCHOLOGY

FIRST YEAR PROJECT

Comparing cues to the same concept

A difference in motivation

Author:

Pierce EDMISTON

Advisor:

Gary LUPYAN

Committee:

Tim ROGERS

C. Shawn GREEN

September 20, 2013

Abstract

Introduction

Conceptual knowledge is acquired through experiences in a variety of sensory-motor modalities (Hoffman & Ralph, 2013). Consequently, the same concept can be activated by a variety of cues. Are all such cues to the same concept basically equivalent, differing only in strength of association?

We investigated this question by comparing two auditory cues to the same concept: verbal category labels (e.g., the word “dog”) and unambiguous natural sounds (a dog <bark>). Natural sounds index a particular source; we say they are “motivated” cues to a concrete instance in the environment. Unlike natural sounds, category labels are “unmotivated”: cues that functionally abstract across individual instances. Similar divisions have been drawn in the history of psychology (e.g. Potter & Faulconer, 1975), but the purpose of this paper is not to deny that experience causally determines the effectiveness of various cues in activating a given concept. Instead, the experiments described here explore how a difference in “motivation” opens a fascinating variety of ways to experience the world.

If our goal is to demonstrate how different cues activate the same concept differently, our first step is finding distinguishable cues to the same concept. Verbal labels and natural sounds are *a priori* equivalent cues to a concept if only in that they are both acoustic signals that are learned in similar ways (e.g., in that recognition of both isolated speech and natural sounds varies as a function of familiarity, environmental frequency Ballas, 1993). At the same time, learning to recognize natural sounds cannot be reduced to the act of applying a known label a new percept (Stuart & Jones, 1995, 1996; Schön, Ystad, Kronland-Martinet, & Besson, 2010; Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2010).

For concepts with unambiguous natural sounds (e.g., a cat <meow>; a doorbell <ring>), both cues have been shown to activate similar semantic networks. For example, both labels and sounds elicit comparable N400 event related potentials (Cummings et al., 2006; Van Petten & Rheinfelder, 1995)—a coarse index of semantic processing—even when the identification of the natural sound is incidental to task demands (Orgs, Lange, Dombrowski, & Heil, 2006, 2007).

Given these parallels between category labels and natural sounds, it is surprising that category labels appear to activate concepts in a unique way. For example, Lupyan and Thompson-Schill (2012) found that participants asked to verify an image as a member of a cued category responded significantly faster to label cues than to sound cues. This label advantage held across multiple cue-to-image delays and even extended to novel categories of alien instruments with learned labels and “sounds”—evidence that the difference between labels and sounds is not entirely attributable to a difference signal ambiguity.

How then should we understand the relationship between category labels and natural sounds? There are of course many important differences between labels and sounds; labels are discrete, reproducible, and often learned in the absence of any physical referent, to name a few. Most importantly, however, is that natural sounds have a causal relationship with a particular physical source (Ballas & Howard, 1987). To use a simple example, big dogs—and agitated dogs—have deeper barks. Of course, big humans—and agitated humans—also have deeper “barks” (Rendall, Vokey, & Nemeth, 2007), but the size of a speaker in no way constrains the meaning of speech. The meaning of a motivated cue, however, is still highly

constrained (Hall, Kingma, & Peters, 2013).

One might even posit the meaning of a motivated cue *is* this constraint. To borrow an analogy from Fowler (1990), instead of “seeing” a lawfully distorted array of electromagnetic signals, we see an object in the world (that structurally distorts light waves). Similarly, instead of “hearing” an acoustic waveform, we hear the source of the sound.

Human perceivers are able to exploit such regularities in the acoustic environment, not only in determining the size of dogs (Taylor, Reby, & McComb, 2008), but also for determining somewhat obscure physical dimensions like the shape of resonating plates (Kunkler-Peck & Turvey, 2000), the length of wooden dowels (Carello, Anderson, & Kunkler-Peck, 1998), or even the hardness of percussion mallets (Freed, 1990).

If auditory perception—indeed all perception—is a complex process of recreating the source (Clark, 2012), then what is the source of a category label? First we must acknowledge that speech perception is not strictly about acoustic information (Massaro, 2004; McClelland, Mirman, & Holt, 2006). Next, abandoning the notion of word meanings as entries in a mental lexicon (Elman, 2004) makes calling category labels “unmotivated” a bit clearer. Labels activate categorical representations of concepts. This abstract notion of “categorical” has been explored in a series of studies (see Lupyan, 2012). Labels activate categorical representations that sharpen between-category boundaries (Lupyan, Rakison, & McClelland, 2007) while blurring within-category differences (Lupyan, 2008).

We believe the distinction between motivated and unmotivated cues pins down an important difference between language and nonverbal communication (Burling et al., 1993). In the four experiments that follow, we intend to quantify this difference. The null hypotheses that will be tested are (a) that natural sounds are simply less effective but functionally identical to category labels, and (b) that natural sounds are “equal but opposite” category labels; labels are cues to linguistic and cultural categories, while natural sounds are cues to some other type of category.

1 Experiment 1

Do natural sounds activate a representation of the source of the sound? Consider the type of chainsaw one might expect after hearing a chainsaw revving (Fig. 1). In Experiment 1 we compared natural sounds to category labels in the processing of active and silent images. If natural sounds index a particular causal source, then natural sound cues will lead to faster processing of sound-producing images.

1.1 Methods

Participants 14 University of Wisconsin–Madison undergraduates participated in Experiment 1 in exchange for course credit.

Materials Two audio clips (label_{female}, sound) and four color photographs (two silent, two action) for each of twelve categories of familiar animals and artifacts¹ were used in

¹Categories used in Experiment 1: BABY, BEE, BIRD, BOWLING BALL, CAR, CAT, CHAINSAW, DOG, KEYBOARD, RIVER, SCISSORS, TOILET

Figure 1: Sample images for CHAINSAW used in Experiment 1. Do label and sound cues lead to different expectations about subsequent visual information?



Experiment 1.

To verify the validity of the image type manipulation, participants in a separate image rating study evaluated each picture on one of two dimensions (category typicality or sound match) using a 5-point Likert scale. For category typicality, participants viewed e.g., a dog, and were asked: “How typical is this dog of dogs in general?” For sound match, participants listened to e.g., a bird <chirp>, saw a picture of a bird, and were asked: “How well does that sound go with this picture?” As expected, the silent images were rated higher on category typicality ($M = 4.57, SD = 0.29$) than on sound match ($M = 3.49, SD = 0.66$), while active images were rated higher on sound match ($M = 4.37, SD = 0.52$) than on category typicality ($M = 4.05, SD = 0.27$).

Procedure Participants were tested in isolated rooms and wore headphones for the auditory cues. Participants were asked to verify images as members of a cued category. If the cue and image matched (e.g., a <meow> followed by a picture of a cat), participants pressed the button for ‘Yes’ on a gaming controller; if not (e.g., the word ‘cat’ followed by a picture of a car), they pressed the button for ‘No’.

Each trial began with a 250 msec fixation cross. Images appeared 1 second after the offset of the auditory cue and disappeared after each response. This long delay ensured that participants had ample time to process the sounds and labels (see Lupyan & Thompson-Schill, 2012). Auditory performance feedback was provided on all trials.

Participants completed a total of 576 trials (50% matching). Cue type (label, sound) and image type (silent, active) varied randomly within-subjects. The experiment took 30 minutes to complete.

1.2 Results and Discussion

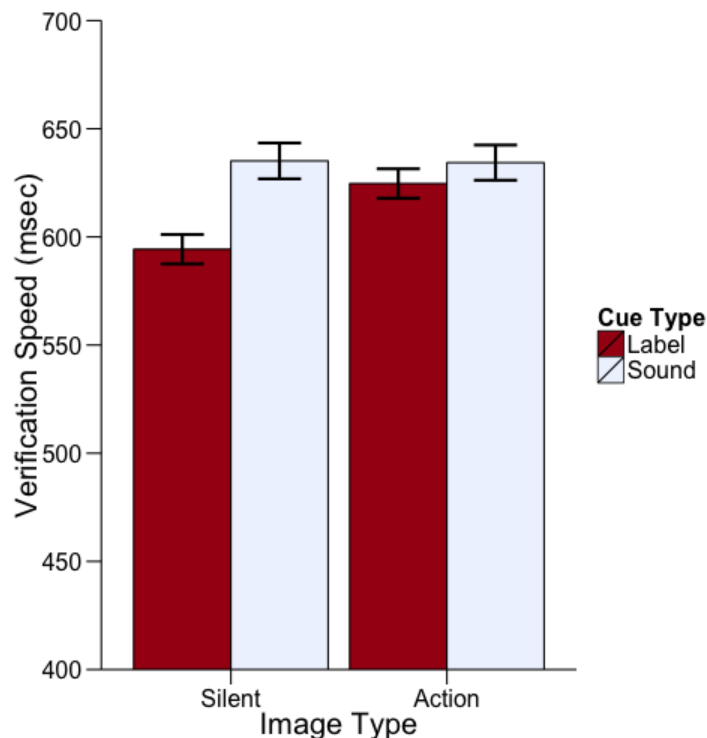
The following criteria are applied to data in Experiments 1-3: Only correct responses on matching trials (‘Yes’ responses) were analyzed. Verification speeds less than 250 msec or greater than 1500 msec were considered errant and excluded (3.11% of total trials). Overall response accuracy was high, $M_{acc} = 96\%$. By-item and by-subject accuracies more than 3 standard deviations below the group mean were considered invalid. Only responses to the natural sound cue for RIVER were removed $M_{acc}(river) = 0.91$, %3.98 of total trials.

We fit the data with linear mixed regression (Bates, Maechler, & Bolker, 2013) predicting verification speeds from the interaction between cue type (label, natural sound) and image

type (action, silent). I allowed intercepts and slopes for all cue type \times image type effects to vary by-subject, and intercepts to vary by-item (picture file). Correlations between random effect parameters were not estimated. All p -values were computed using Monte Carlo Markov Chain sampling with 10,000 simulations.

Responses following label cues ($M = 609$ msec) were reliably faster than responses to natural sound cues ($M = 639$ msec), $t(13) = 3.47, p = 0.003$. This effect was moderated by image type such that the label advantage was greater for silent, category typical images than for active images, $t(13) = -2.35, p = 0.032$ (see Fig. 2).

Figure 2: Participants were asked to verify images as members of a cued category. Responses following label cues were significantly faster than responses following sound cues. This label advantage was greater for the silent, category typical images than the action, sound-matched images. Unexpectedly, active images were not verified more quickly than silent images following a natural sound cue.



The results of Experiment 1 supported the hypothesis that labels activate concepts differently (Lupyan & Thompson-Schill, 2012). Unlike sounds, labels activate a more category-typical representation. However, cuing with a natural sound did not speed verification of images depicting the action that created the sound.

2 Experiment 2

A possible explanation for failing to find faster verification speeds for action images following a natural sound cue in Experiment 1 is that natural sounds as motivated cues do not index

the act, they index the identity. For example, a dog <bark> may not activate “dog with jaw open” as much as it activates “dog of a certain size” (Fig. 3). In Experiment 2 we tip the scales in favor of natural sounds. If natural sounds index a particular source, then verification speeds for sound-congruent images should be faster than for sound-incongruent images.

Figure 3: Sample images for DOG used in Experiment 2. Participants heard natural sound cues associated with both within-category variations. As a control for similar within-category variation, label cues were also recorded from two speakers (male and female).



2.1 Methods

Participants 43 University of Wisconsin—Madison undergraduates participated in Experiment 2 in exchange for course credit.

Materials Four digital audio clips ($\text{label}_{\text{male}}$, $\text{label}_{\text{female}}$, sound_A , sound_B ; see Table 1) and four color photographs for each of six categories of familiar animals and artifacts were used in Experiment 2. All audio clips trimmed to 600 msec.

Category	Version A	Version B
bird	hawk	songbird
dog	rottweiler	yorkie
drum	bongo	drumkit
guitar	acoustic	electric
motorcycle	dirtbike	Harley
phone	cellphone	rotary

Table 1: Categories with distinct within-category sounds used in Experiment 2. A sound-congruent image was one which matched the natural sound cue at this more precise level. If cued with $\langle \text{bird}_{\text{hawk}} \rangle$, a picture of a hawk is sound-congruent but a picture of a songbird is sound-incongruent, even though both images elicited the same response.

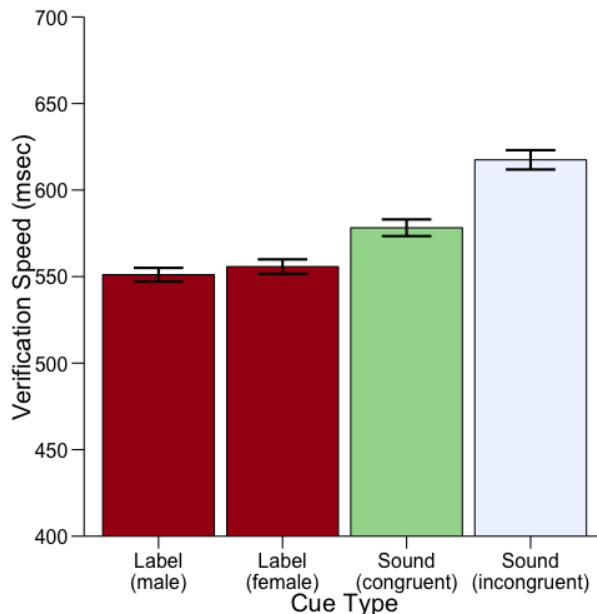
Procedure Experiment 2 was conducted in the same manner as Experiment 1. The instructions were modified to emphasize category-level matches; participants were told that if they heard e.g., a guitar chord they should respond ‘Yes’ to any subsequent picture of a guitar. As in Experiment 1, images appeared 1 second after auditory cue offset.

Participants completed a total of 383 trials (50% matching). Cue type ($\text{label}_{\text{male}}$, $\text{label}_{\text{female}}$, sound_a , sound_b) and picture type (sound-congruent, sound-incongruent) varied randomly within-subjects. The experiment took 30 minutes to complete.

2.2 Results and Discussion

Correct responses on matching trials were analyzed as in Experiment 1 (1.60% of total trials excluded). We fit the data with linear mixed regression predicting response times from the four cue types: $\text{label}_{\text{male}}$, $\text{label}_{\text{female}}$, $\text{sound}_{\text{congruent}}$, $\text{sound}_{\text{incongruent}}$. We created a set of three orthogonal contrasts with the hypothesis that verification speeds would be fastest following both label types and slowest following incongruent sounds with congruent sounds falling between. The contrast of interest was significant, $t(42) = 8.78, p = 0.0001$, and the remaining contrasts did not account for a significant amount of residual between-group variance, $\chi^2(4) = 8.64, p = 0.07$, (Abelson & Deborah, 1997). This result verifies (a) that congruent sounds resulted in faster verifications than incongruent sounds, (b) that labels resulted in faster verifications than congruent sounds, and (c) that there were no differences between responses to male and female speakers (see Fig 4).

Figure 4: Cuing with a label—independent of speaker gender—resulted in significantly faster verifications than cuing with natural sounds even when the to-be-verified image matched the natural sound at an arguably more-specified level.



In Experiment 2 we found that natural sounds activated a representation associated with the identity of a likely causal source. However, labels still lead to faster verifications even when the natural sound is arguably more specified.

3 Experiment 3

Do natural sounds activate a representation of the source of a sound *at a particular time*? Motivated cues index a multimodal cause, or, to put it differently, multisensory integration—a cue in one modality facilitating processing in another modality—reflects learned semantic knowledge (Fiebelkorn, Foxe, & Molholm, 2009; Hein et al., 2007; Chen & Spence, 2011; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004). In Experiment 3, we investigated if label and natural sound cues influence recognition speed based not only on the fit between an auditory cue and an image, but also on the delay between the cue and the image. If natural sounds index a particular source at a particular time, then sound-matched images should be recognize more quickly when presented simultaneously with the sound than less sound-matched images.

3.1 Methods

Participants 56 University of Wisconsin–Madison undergraduates participated for course credit.

Materials Spoken label and natural sound audio clips for ten of the twelve² categories used in Experiment 1 were used in Experiment 3. Four new color photographs per category with a “continuous” variety of sound match properties were obtained.

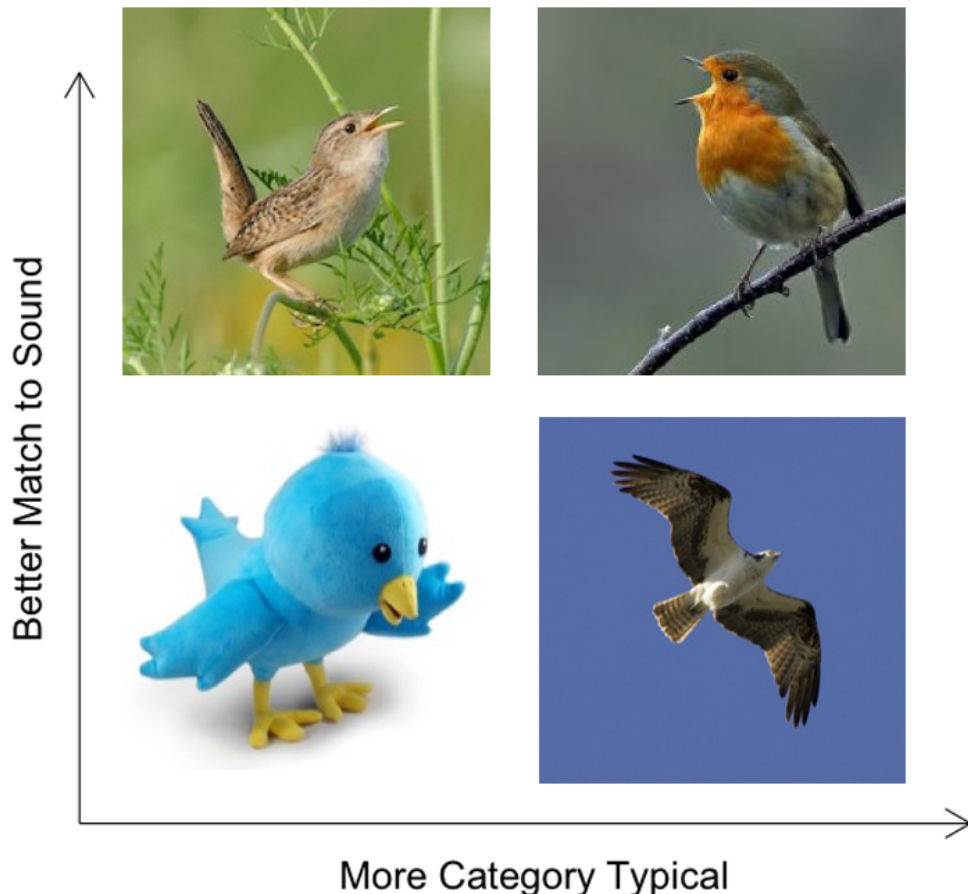
To quantify the image type manipulation, image ratings (category typicality and sound match) were collected via Amazon’s Mechanical Turk (mTurk). mTurk workers ($N = 42$) heard either spoken labels or natural sounds, and were presented 8 to 10 pictures for each category with the following instructions: “Please listen to the following audio clip and report how well each image fits with the audio file.” Ratings were given on a 5-point Likert scale. From these data, we selected 4 images for each category roughly corresponding to the quadrants depicted in Fig. 5. There was a positive correlation between category-typicality and sound-match (Pearson’s $r=0.27$). These ratings were standardized (z -score) and used as predictors in subsequent analyses.

Procedure Experiment 3 was conducted in the same manner as Experiments 1 and 2. Participants completed a total of 427 trials (75% matching). This increase in response validity compared to Experiments 1 and 2 allowed us to fully counterbalance all trial variables on matching trials while keeping the length of the experiment manageable. The target image was either presented simultaneously with the onset of the auditory cue, or delayed the offset of the auditory cue by 400 msec.

Cue type (label, natural sound), picture exemplar (four per category), and delay (simultaneous, delayed) varied randomly within-subject. The experiment took 30 minutes to complete.

²Categories RIVER and TOILET were excluded due to insufficient within-category variation.

Figure 5: Sample images from BIRD used in Experiment 3. Category typicality and sound-match ratings can be conceptualized as independent measures, though they are undoubtedly correlated for categories with unambiguous natural sounds (the most typical dogs are likely to have the most typical barks).



3.2 Results and Discussion

Only correct responses on matching trials (‘Yes’ responses) were analyzed. Verification speeds less than 250 msec or greater than 1500 msec were considered errant and excluded (1.21% of total trials). By-item and by-subject accuracies were analyzed the same as in Experiments 1 and 2. Responses to the natural sound cue for the category SCISSORS, $M_{acc}(scissors) = 0.91$, and the data from one participant, ($M_{acc}(subject) = 0.89$), were removed.

We fit the data with linear mixed regression predicting verification speeds from 2 three-way interactions: cue type \times delay \times category typicality, and cue type \times delay \times sound match. We allowed intercepts and slopes for all cue type \times delay effects by-subject, and intercepts to vary by-item (picture file). Other model parameters were the same as in the previous experiments.

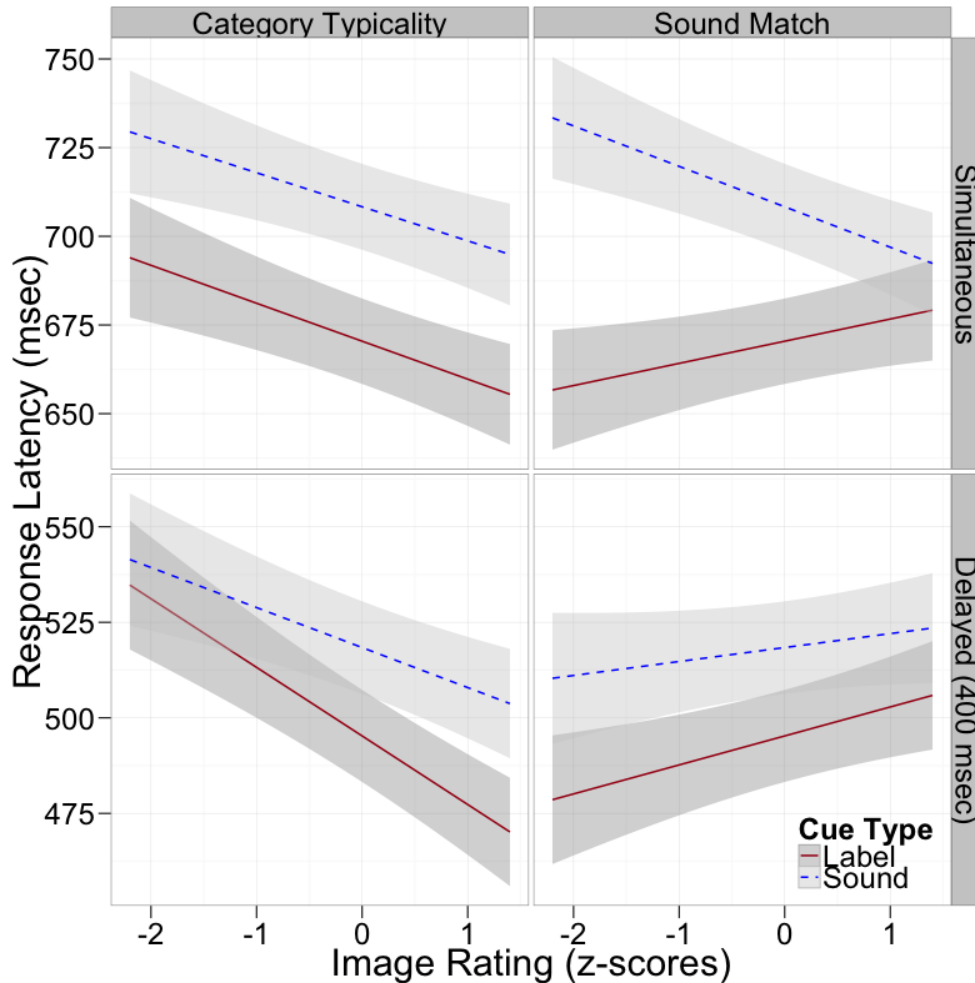
Responses to label cues were again reliably faster than responses to natural sound cues, $t(40) = 8.82, p < 0.0001$. The effect of cue type was moderated by delay, $t(40) = -2.46, p =$

0.016, such that the label advantage was greater on simultaneous trials than it was on delayed trials (Fig. 6).

Category typicality was a reliable predictor of RTs, $t(41) = -2.6, p < 0.011$. Importantly, this effect remained constant across both cue types and both image delays. That is, the RT advantage for more category-typical images over less category-typical images was equivalent for label and natural sound cues, on both simultaneous and delayed trials (Fig. 6, left column).

There was a reliable three-way interaction between sound-match, cue type, and image delay, $t(41) = 2.2, p = 0.023$. On simultaneous presentation trials, RTs following natural sound cues decreased as the sound-match of the image increased, while RTs following label cues did not vary by sound-match, (Fig. 6, upper right). However, there was no such cue type \times sound-match interaction at the 400 msec delay (Fig. 6, lower right). That is, sound-match predicted RTs following natural sounds and not labels when the delay was simultaneous, but not with a 400 msec delay.

Figure 6: Speed of verification of category typical and sound producing images as a function of cue type and delay.



To summarize, image ratings for category-typicality and sound-match correlated with

response times based on the cue and the cue-image delay. First, when presented with a spoken label, RTs were predicted by category-typicality of the image, and this effect held across both cue-to-image delay periods. Second, when presented with a natural sound, the sound-match of the image correlated with the response time to that image, *but only when the cue-image pair was presented simultaneously*. That is, hearing a natural sound improved processing of a particular kind of visual image: a picture depicting an object that could have made the sound at the moment the sound was detected. These results support the hypothesis that motivated cues index a particular multimodal source.

These findings afford an intriguing prediction: In Experiments 1-3, task demands required explicitly performing a category match between an auditory cue and an image. However, if natural sounds as motivated cues index a particular multimodal source, then natural sounds should drive perception of a particular source even without explicitly attending to the auditory cues (Orgs, Lange, Dombrowski, & Heil, 2008).

4 Experiment 4

Do motivated cues index a particular source regardless of task demands? Previous studies have demonstrated that task-irrelevant category labels draw visual attention (Salverda & Altmann, 2011) and that natural sounds guide visual search amid distractors (Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008). In Experiment 4 we investigated if labels and natural sounds guide visual attention to different types of images when the auditory cues are made irrelevant to the participants task.

4.1 Methods

Participants 30 University of Wisconsin–Madison undergraduates participated for course credit.

Materials Images and audio clips were the same as those used in Experiment 3.

Procedure Participants searched a 2×2 image grid to find a target image to be highlighted in a green frame. They were informed the experimenters were studying how listening to sounds would help or hurt their speed at performing this task. They were also told the experimenters were curious where they would look during the task, which necessitated the eye tracker.

Each trial began with a fixation circle that turned green when fixated. Participants could only begin each trial by clicking the green dot, ensuring both accurate calibration and foveal origin on each trial. Auditory cues were played over the computer speakers. The target appeared 2 seconds after the onset of the 2×2 image grid. Participants ended a trial by clicking the mouse within the target frame.

All images in a given trial comprised members of the same category. Auditory cues varied randomly within-subject and were present on every trial. Approximately half of the participants (N=13) were in the simultaneous condition in which the auditory cue was played concurrently with the onset of the image grid. These data are analyzed first. The remaining

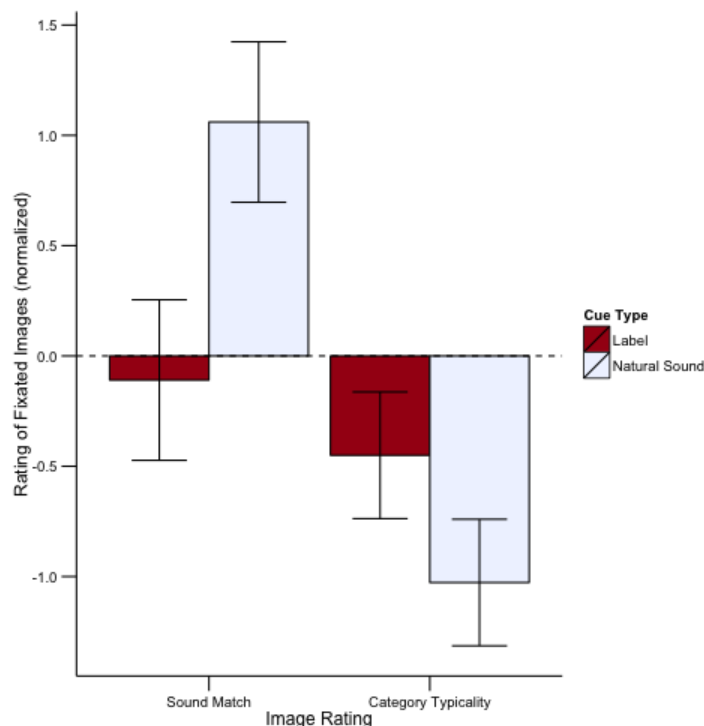
participants (N=17) were run in a subsequent control condition in which the appearance of the images lagged the auditory cue by 400 msec—the same delay as in Experiment 3.

4.2 Results and Discussion

Only fixation events (as defined by the EyeLink’s internal criteria) occurring between the onset of the image grid and the onset of the target frame were analyzed.

First we tested if participants in the simultaneous condition looked at systematically different types of images based on the presented cue *before* the onset of the target. Raw image ratings reported in Experiment 3 were normalized for each trial (that is, for each category) to force random search of the image grid to reflect a net 0.0 for both category typicality and sound match image ratings. We fit linear mixed regression models predicting per-trial average category typicality and per-trial average sound match of fixated images from cue type, with relevant random effects. The average category typicality was significantly less than 0, $t(12) = -3.34, p = 0.003$, but did not differ by cue type, $t(12) = 1.58, p = 0.11$. The average sound match was significantly greater than 0, $t(12) = 2.29, p = 0.027$, and was significantly greater following natural sound cues than label cues, $t(12) = 1.96, p = 0.0502$ (Fig. 7). However, this effect was washed out when delay (data from the control condition) was added to the model $t(29) = -1.21, p = 0.23$.

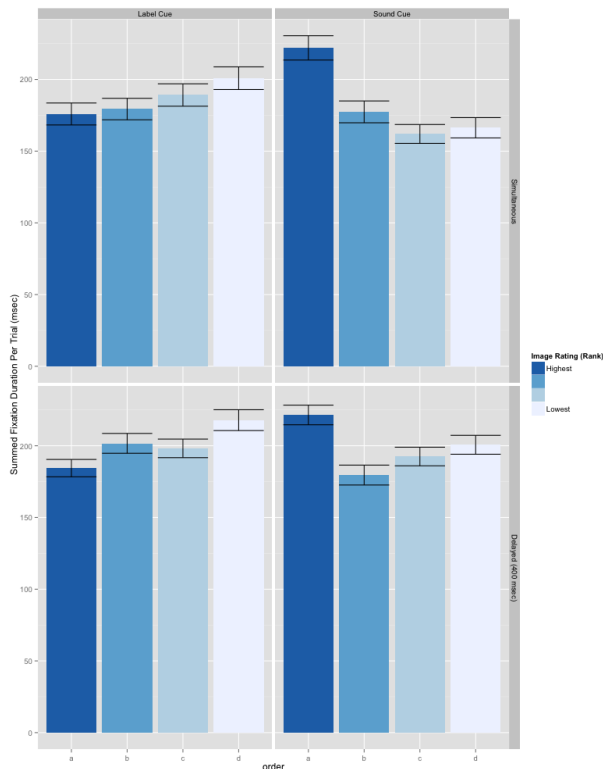
Figure 7: There was a preference for sound matched images following natural sound cues.



Next we tested if participants looked systematically longer at a particular image during a trial. We rank-ordered the images for each trial in terms of decreasing image rating (sound match, category typicality), and summed the per-trial fixation duration for each image.

We fit a linear mixed regression model predicting average per-trial fixation duration from cue type \times rank-ordered image. We created a set of three orthogonal contrasts with the hypothesis that summed fixation duration would be greatest for the highest ranked image in terms of category typicality and sound match. There was a significant interaction between cue type and the contrast of interest, $t(29) = 3.56, p = 0.0004$, and the remaining contrasts did not account for a significant amount of residual between-group variance, $\chi^2(8) = 12.9, p = 0.12$ (Abelson & Deborah, 1997). This effect was strengthened by a similar effect in the delay condition, $t(29) = 8.01, p = 0.0001$ (Fig. 8).

Figure 8: There was a preference for sound matched images following natural sound cues.



In Experiment 4 we found sound cues seem to direct attention toward more sound-typical images, even when processing the sound cue has no relevance for the task. Moreover, sound cues directed attention toward the most likely source of the sound, while labels lead to more even search of all category members. However, the conclusions drawn from Experiment 4 must be qualified by inconsistent data from the control condition.

General Discussion

Four experiments garnered support for a distinction between motivated and unmotivated cues by demonstrating ways in which category labels and natural sounds systematically differ in activating the same concept. In Experiments 1-3 we used category verification tasks to probe the nature of conceptual activations following both labels and sounds. Labels as unmotivated cues activate a more category-typical representation when presented a single

exemplar (Experiments 1, 2, 3), but also make all members of a category appear more similar (Experiment 4). While natural sounds failed to activate a representation of the action making the sound (Experiment 1), one can tease apart the representation of the source of the sound by using within-category variation (Experiment 2). This source is multimodal, leading to the multisensory integration effects (time sensitivity of the natural sound \times sound match interaction) seen in Experiment 3. Lastly, natural sounds are inherently motivated in that processing of an individual causal source is nearly inevitable upon hearing a natural sound (Experiment 4).

The most significant distinction between motivated and unmotivated cues is that motivated cues are obligated to index a multisensory source. As such, natural sounds are not the only motivated cue. Most sensory input is in fact motivated, such images (Beauchamp, Lee, Argall, & Martin, 2004; Brunel, Lesourd, Labeye, & Versace, 2010) and even smell (Gottfried & Dolan, 2003). Of course, motivated cues are also inherent in speech (von Kriegstein & Giraud, 2006; Shintel & Nusbaum, 2007; Rendall et al., 2007), but language can be stripped of its motivated aspects in a way that natural sounds cannot.

Conclusion We found verbal and nonverbal cues activate different conceptual representations evident in patterns of response latencies to recognize and verify different category exemplars. In a replication of previous findings, verbal cues facilitated recognition of category-typical images. We extended these findings to discern the specifics of conceptual representations activated via natural sound cues: Natural sounds facilitated visual processing of images that fit with the presented sound, but only if the sound and image were presented simultaneously. Critically, these effects were mediated by time, with natural sound cues improving responses to sound-matched images only during simultaneous presentation.

References

- Abelson, R. P., & Deborah, A. (1997). Contrast tests of interaction hypotheses. *Psychological Methods*, 2(4), 315–328.
- Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., & Besson, M. (2010). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience*, 22(11), 2555–2569.
- Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 250.
- Ballas, J. A., & Howard, J. H. (1987). Interpreting the language of environmental sounds. *Environment and behavior*, 19(1), 91–114.
- Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes [Computer software manual].
- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41(5), 809–824.

- Brunel, L., Lesourd, M., Labeye, E., & Versace, R. (2010, May). The sensory nature of knowledge: Sensory priming effects in semantic categorization. *The Quarterly Journal of Experimental Psychology*, 63(5), 955–964.
- Burling, R., Armstrong, D. F., Blount, B. G., Callaghan, C. A., Foster, M. L., King, B. J., . . . Wallace, R. (1993). Primate Calls, Human Language, and Nonverbal Communication [and Comments and Reply]. *Current Anthropology*, 34(1), 25–53.
- Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998, May). Perception of Object Length by Sound. *Psychological Science*, 9(3), 211–214.
- Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5), 1554–1568.
- Clark, A. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.*
- Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006, October). Auditory semantic networks for words and natural sounds. *Brain Research*, 1115(1), 92–107.
- Elman, J. L. (2004, July). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, 8(7), 301–306.
- Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2009, December). Dual Mechanisms for the Cross-Sensory Spread of Attention: How Much Do Learned Associations Matter? *Cerebral Cortex*, 20(1), 109–120.
- Fowler, C. A. (1990). Soundproducing sources as objects of perception: Rate normalization and nonspeech perception. *The Journal of the Acoustical Society of America*, 88, 1236.
- Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, 87, 311.
- Gottfried, J. A., & Dolan, R. J. (2003). The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron*, 39(2), 375–386.
- Hall, M. L., Kingma, S. A., & Peters, A. (2013). Male Songbird Indicates Body Size with Low-Pitched Advertising Songs. *PLoS ONE*.
- Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007, July). Object Familiarity and Semantic Congruency Modulate Responses in Cortical Audiovisual Integration Areas. *Journal of Neuroscience*, 27(30), 7881–7887.
- Hoffman, P., & Ralph, M. A. L. (2013, January). Shapes, scents and sounds Quantifying the full multi-sensory basis of conceptual knowledge. *Neuropsychologia*, 51(1), 14–25.
- Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008, June). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, 15(3), 548–554.
- Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 279.
- Laurienti, P., Kraft, R., Maldjian, J., Burdette, J., & Wallace, M. (2004, June). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, 158(4).
- Lupyan, G. (2008). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, 137(2), 348–369.

- Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, 3.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not Just for Talking. *Psychological Science*, 18(12), 1077–1083.
- 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.
- Massaro, D. W. (2004). From Multisensory Integration to Talking Heads and Language Learning. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (p. 153). MIT Press.
- McClelland, J. L., Mirman, D., & Holt, L. L. (2006, August). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(8), 363–369.
- Orgs, G., Lange, K., Dombrowski, J.-H., & Heil, M. (2006, December). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition*, 62(3), 267–272.
- Orgs, G., Lange, K., Dombrowski, J.-H., & Heil, M. (2007, August). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology*, 65(2), 162–166.
- Orgs, G., Lange, K., Dombrowski, J.-H., & Heil, M. (2008, May). N400-effects to task-irrelevant environmental sounds: Further evidence for obligatory conceptual processing. *Neuroscience Letters*, 436(2), 133–137.
- Potter, M. C., & Faulconer, B. A. (1975, February). Time to understand pictures and words. *Nature*, 253.
- Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1208–1219.
- 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.
- Schön, D., Ystad, S., Kronland-Martinet, R., & Besson, M. (2010). The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience*, 22(5), 1026–1035.
- Shintel, H., & Nusbaum, H. C. (2007, December). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition*, 105(3), 681–690.
- Stuart, G. P., & Jones, D. M. (1995). Priming the identification of environmental sounds. *The Quarterly Journal of Experimental Psychology*, 48(3), 741–761.
- Stuart, G. P., & Jones, D. M. (1996). From auditory image to auditory percept: Facilitation through common processes? *Memory & Cognition*, 24(3), 296–304.
- Taylor, A. M., Reby, D., & McComb, K. (2008). Human listeners attend to size information in domestic dog growls. *The Journal of the Acoustical Society of America*, 123(5), 2903.
- Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33(4), 485–508.

von Kriegstein, K., & Giraud, A.-L. (2006). Implicit Multisensory Associations Influence Voice Recognition. *PLoS Biology*, 4(10), e326.