# Verbal and Nonverbal Cues Activate Concepts Differently, at Different Times

**Pierce Edmiston (pedmiston@wisc.edu)**
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

**Gary Lupyan (lupyan@wisc.edu)**
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

**Abstract**

For many everyday objects, nonverbal cues are just as effective as verbal cues for identifying a category member. Yet this similarity belies the evidence that labels activate conceptual representations differently than nonverbal sounds (Lupyan & Thompson-Schill, 2012). Here we investigate the specifics of that relationship. While labels activate a category-typical representation, natural sounds activate a representation of the source of the sound, and improve recognition of such an image when simultaneously presented.

**Keywords:** categorization, concepts, sounds, recognition, cross-modal effects, language

## Introduction

Most concepts are multimodal and can be activated in a variety of ways (Hoffman & Ralph, 2013). For example, the concept DOG can be activated by seeing a wagging tail, hearing a bark, or petting its furry coat. However, the concept DOG can also be activated by hearing the word 'dog'—without seeing, hearing, or touching an actual dog. This raises the question of how concepts activated by nonverbal sensory cues compare to those activated by verbal category labels.

In the experiments reported here we compare how verbal and nonverbal cues activate representations of purportedly the same concepts. In particular, we focus on visual aspects of familiar animals and artifacts as cued by natural sounds: auditory events with a distinct source (e.g., cat meowing, chainsaw revving), and how these same concepts are activated by verbal labels: words like "cat" and "chainsaw."

The mechanisms underlying recognition of nonverbal sounds and of speech appear to be quite similar. Recognition of both words and natural sounds varies as a function of familiarity, frequency, and context (Ballas, 1993; Stuart & Jones, 1995). Perception of both natural sounds and speech is influenced by signal ambiguity and noise in similar ways (Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2010; Gygi, Kidd, & Watson, 2004). Labels and natural sounds also seem to activate the same semantic networks. Both types of cues elicit similar N400 event-related potentials (Cummings et al., 2006; Van Petten & Rheinfelder, 1995)—even when the identification of the natural sound is incidental to task demands (Orgs, Lange, Dombrowski, & Heil, 2008). Functional imaging during similar sequential processing tasks reveals largely overlapping cortical areas recruited in processing labels and natural sounds (Dick et al., 2007). Lastly, patterns of naming deficits in patients with aphasia suggest the labeling of everyday objects and the visual recognition of natural sound sources rely on similar cognitive resources (Goll et al., 2010; Saygin, Dick, Wilson, Dronkers, & Bates, 2003).

The perception of meaningful nonverbal sounds and of words is thus dependent on many of the same properties and activate largely the same semantic networks. But although it may seem that verbal and nonverbal cues are in important respects equivalent, there are critical differences. One such difference is that natural sounds, unlike labels, have a causal relationship with a specific physical source (Ballas, 1993). Recognizing these relationships requires learning, but the relationship between a referent and its natural sound is not arbitrary. We call these relationships "motivated": that is, they are determined by physics (e.g., thunder) or driven by biology (e.g., large dogs—and agitated dogs—have deeper barks). Auditory perceivers are able to exploit such "motivated" relationships and surmise features of a hidden physical source, such as the size of a barking dog (Taylor, Reby, & McComb, 2008), the shape of resonating plates (Kunkler-Peck & Turvey, 2000), or the hardness of percussion mallets (Freed, 1990). The perception of these auditory sources is surprisingly accurate, reflecting the lawful relationships between signals and sources in the environment (Fowler, 1990). Importantly, sounds covary lawfully *within* as well as *between* categories. For example, a barking sound informs us not only that its source is a dog, but can inform us of the approximate size of the dog.

In contrast, the relationship between labels and their referents is "unmotivated." By this term we do not simply mean that words are arbitrary, i.e., that "dog" refers to dogs by convention (cf. Hockett, 1966). But rather that there exists a word "dog" that denotes the entire category of dogs rather than a particular type or instance (dachshund, German shepherd, dog-on the left, dog-far away, etc.). In short, barks index specific occurrences of dogs. Even though we can interpret them at a higher, more categorical level, the surface properties of a specific bark still indexes a *particular* dog. Verbal labels, on the other hand, abstract over these specifics. When we say "dog" we can leave all that information unspecified. On this view, labels may activate concepts in a more categorical way. This prediction has been supported by a variety of findings (Lupyan, 2012). For example, Lupyan & Thompson-Schill (2012) found that label cues resulted in faster visual processing over equally predictive nonverbal cues. This advantage persisted across a number of cue-to-image delay periods and extended to artificially created objects with novel labels and "natural" sounds, suggesting that labels do not activate conceptual representations faster but differently than nonverbal cues. In our view, labels activate representations that emphasize the differences between categories, and thus play a facilitative role in category learning (Lupyan, Rakison, & McClelland, 2007). These categorical representations enable faster recognition of category-typical objects (Lupyan & Swingley, 2012), but blur within-category differences reflected in biased exemplar memory (Lupyan, 2008).

However, what is not clear from these previous results is how "unmotivated" and "motivated" cues differ in activating different instances of purportedly the same concept. If "unmotivated" verbal cues activate more categorical representations, then what do "motivated" nonverbal cues activate? Given the inherent causal link between a natural sound and its particular physical source, we predicted that natural sound cues would lead to faster processing of images depicting the production of the auditory cue. The results ended up being more interesting.

## Experiment 1

Hearing a sound characteristic of an animal or artifact may automatically activate particular instances of that category. Consider the kind of chainsaw one might expect upon hearing a chainsaw sound (Fig. 1). Here, we asked whether verbal and nonverbal cues lead to different expectations about subsequent visual information. In Experiment 1 we investigated if label and natural sound cues influence visual processing differently based on the action depicted in target images. In line with previous research, we predicted that when presented a label cue, participants would respond faster to category-typical images. Conversely, we predicted that when presented a natural sound cue, participants would respond faster to sound-matched images.

### Methods

**Participants** 14 University of Wisconsin—Madison undergraduates participated for course credit.

**Materials** Auditory cues were spoken labels and natural sounds for 12 target categories of familiar animals and artifacts used in Lupyan & Thompson-Schill (2012).[1] Visual images were four color photographs for each category: 2 category-typical images and 2 sound-matched images. The images were normed to ensure unambiguous categorization. In addition, participants in a separate norming study rated each picture on two dimensions (category typicality and sound match) using 5-point Likert scales. For category typicality, participants viewed e.g., a dog, and were asked: "How typical is this dog of dogs in general?" For sound match ratings, participants listened to e.g., a bird chirping, saw a picture of a bird, and were asked: "How well does that sound go with this picture?" Each participant performed either category-typicality or sound-matching.


Figure 1: Sample stimuli from Experiment 1. Does hearing the sound of a revving chainsaw activate a representation of a chainsaw in action?

**Procedure** Participants completed a category verification task in which an auditory cue—either a spoken category label (e.g., 'cat') or a natural sound (e.g., <meow>)—preceded a visual image. Participants determined if each cue-image pair matched on a category level by pressing 'Yes' or 'No' using a labeled gaming controller. For example, if they heard a chainsaw revving or the spoken word "chainsaw" and then saw a picture of a chainsaw, they would press the 'Yes' button. The picture disappeared after each response, and performance feedback was given. Cue type (Label, Natural Sound) and picture type (category-typical, sound-matched) varied randomly within-subjects. There were a total of 576 trials per subject (50% matching). Each trial began with a 250 msec fixation cross followed by the auditory cue. The target image appeared 1 sec after auditory cue offset. This long delay ensured that participants had ample time to process sounds and labels (see Lupyan & Thompson-Schill, 2012). The experiment took 30 minutes to complete.

### Results and Discussion

Overall accuracy was high (96%). Only correct RTs on matching trials were included. RTs less than 250 msec or greater than 1500 msec were excluded (<4% of correct trials). We fit the data with linear mixed regression (Bates, Maechler, & Bolker, 2012) to predict response times (RTs) from cue type (Label, Natural Sound) and image ratings (category-typicality or sound-match; $z$-scores) with random subject and item effects (target category). As expected (Lupyan & Thompson-Schill, 2012), responses to label cues ($M$=609 msec) were reliably faster than responses to natural sound cues ($M$=639 msec), $F(1,13)=22.14$, $p<0.001$.[2] Also as expected, category-typicality interacted with cue type, $F(1,13)=10.45$, $p=0.001$. An analysis of simple effects revealed a significant effect of category-typicality for label cues, $t(13)=-4.63$, $p<0.001$ (Fig 2, left), but not for natural sound cues, $t(13)= -0.48$, $p=0.63$. In contrast, sound-match did not interact with cue type, $F(1,13)=0.67$, $p=0.41$ (Fig 2, right).

[1] Target categories for Experiment 1: *bird, bee, toilet, scissors, dog, chainsaw, bowling ball, cat, car, keyboard, river, baby*.

[2] All $p$-values were generated using Markov chain Monte Carlo sampling (1,000 simulations).

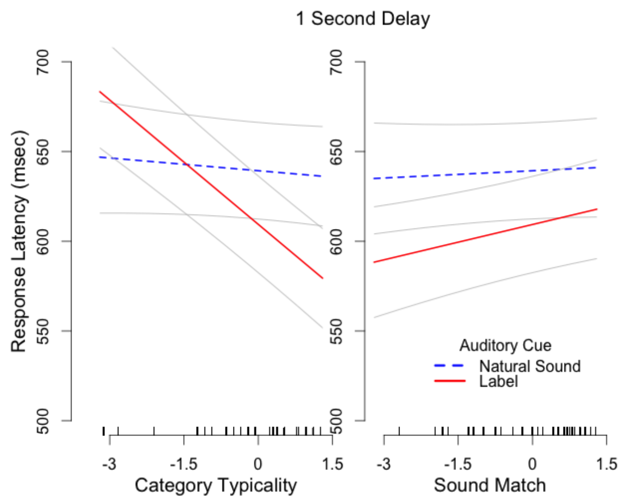


Figure 2: Significant interaction between cue type and category typicality, but not between cue type and sound match. Error bars denote ±1 SE of predicted values (Mazerolle, 2012).

To summarize, the type of auditory cue influenced the recognition of a category member as a function of the category typicality of the image. These results replicate previous findings that labels facilitate visual processing more effectively than nonverbal cues (Lupyan & Thompson-Schill, 2012) and that labels facilitate visual processing as a function of category typicality (Lupyan & Swingley, 2012). However, this interaction was driven exclusively by the effectiveness of labels in facilitating the visual processing of category-typical images. Recognition performance on trials cued by a natural sound was not predicted either by category typicality or by sound-match.

These results demonstrate that people's recognition times vary as a function of category-typicality, but only when cued by a label. Recognition times did not vary as a function of sound-match when cued with a natural sound. The results clearly show that labels and sounds activate familiar concepts differently and that labels appear to activate a representation that is more categorical/typical. Unexpectedly, sounds do not appear to do the same as a function of the fit between the sound and the image. This finding is investigated further in Experiment 2.

## Experiment 2

Our second experiment extends the first in two important ways. First, we compiled a more extensive set of stimuli containing a larger variety of category-typical and sound-matched images to cover the entire 2-dimenional space of category typicality and sound-matching (Fig. 3). Second, we varied the cue-to-image delay. We did this because environmental sounds, unlike labels *index* the animals/objects that produce them. While labels often occur in the absence of the referent (we talk about things not presently in view), sounds are much more temporally correlated with the presence of the referent. If we hear a bark, chances are a dog is in the vicinity.

In Experiment 2, we investigated if label and natural sound cues facilitate visual processing based on the fit between an auditory cue and an image, and on the delay between the cue and the image. In line with the results of Experiment 1, we predicted a label cue would improve processing of category-typical images. We also predicted that a natural sound would improve processing of sound-matched images—that is, where the image depicted an animal or artifact that was the likely source of the natural sound—and that this effect would be greater when the cue and image were temporally coupled, that is, presented simultaneously.

## Methods

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

**Materials** Auditory cues comprised spoken labels and natural sounds for 10 of the 12 target categories used in Experiment 1 (categories *river* and *toilet* were excluded). One of the goals of this experiment was to include pictures from the entire space of category-typicality and sound-match, which required compiling a new set of images. Picture ratings were collected via Amazon's Mechanical Turk (mTurk). mTurk workers ($N$=42) heard either the 10 spoken labels or the 10 natural sounds to be used in the Experiment, and were given 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 corresponding to the quadrants depicted in Fig. 3.

**Procedure** The procedure was the same as in Experiment 1. Trials varied by cue type (Label, Natural Sound), picture type (4 per category), image delay (Simultaneous or 400 msec from auditory cue offset), and cue-to-image

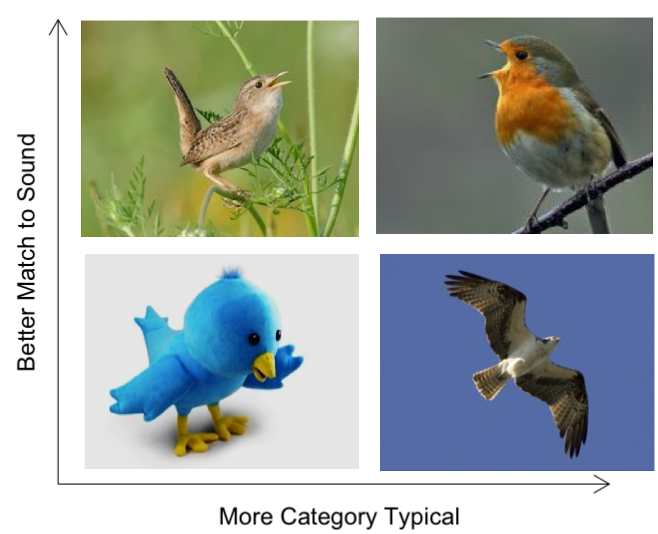


Figure 3: Sample stimuli from Experiment 2. Category-typicality varies independently of sound-match.

categorical match for a total of 427 trials per subject (75% matching[3]). Each trial began with a 250 msec fixation cross. On half the trials, the auditory cue and picture were presented simultaneously; on the remaining trials the picture was presented 400 msec after the offset of the auditory cue. The experiment took 30 minutes to complete.

## Results and Discussion

Overall recognition accuracy of the images was high ($M$=97%), except for trials in which pictures of scissors were cued by a sound of scissors cutting paper ($M$=91%, $SD$=1.8). Participants also reported difficulties with these trials during debriefing (24 out of 56 participants; next most frequent was 5 for *bee*), and these trials were removed from subsequent analyses (<5%). We excluded trials using the same exclusion criteria as in Experiment 1 for analyses (<2% of correct trials removed). Again, the data were fit with linear mixed regression allowing random subject and item effects (target category).

**Delay and Cue Type** We first investigated how the effect of cue type varied by image delay. We fit a model predicting RTs from the interaction between cue (Label, Natural Sound) and image delay (Simultaneous, 400 msec), and found the interaction was significant, $F(1,41)=6.38$, $p=0.01$. As is clearly visible in Fig. 4, it took much longer to recognize an image presented simultaneously than it did to recognize the image 400 msec after auditory cue offset. However, the label advantage was larger for simultaneous presentations ($M_{sound}$-$M_{label}$= 34 msec) than for delayed presentations ($M_{sound}$-$M_{label}$= 19 msec).

**Category Typicality** We then tested if the category typicality of the image influenced response times differently by cue-type and by image delay. We fit a model predicting RTs from cue type (Label, Natural Sound), delay (Simultaneous, 400 msec), and category-typicality ($z$-scores; see Methods). Category typicality was a reliable predictor of RTs, $F(1,41)=30.74$, $p<0.0001$ (Model including category-typicality as a predictor: $\chi^2(4)=34.67$, $p<0.001$). Importantly, this effect held 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 sequential trials (Fig. 5A). Interestingly, these results indicate that responses following natural sound cues were influenced by category-typicality of the image during simultaneous and 400 msec delayed trials.

**Sound Match** We next tested whether the sound-match of the image influenced response times differently by cue-type and by image delay. We fit a model predicting RTs from cue type (Label, Natural Sound), delay (Simultaneous, 400 msec), and sound-match ($z$-scores; see Methods). Response times varied by sound-match as a function of both cue type and image delay, $F(1,46) = 6.23$, $p=0.01$ (Model including sound-match as a predictor: $\chi^2(4)=34.67$, $p<0.001$). 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, $t(46)=-3.47$, $p<0.001$, (Fig. 5B, left). However, there was no such cue type × sound-match interaction at the 400 msec delay, $t(46)=-0.44$, $p=0.66$ (Fig. 5B, right). That is, sound-match mapped with RTs to natural sounds and not to labels when the delay was simultaneous, but not with a 1 sec delay.

To summarize: the category-typicality ratings and the image fit to sound ratings for each picture correlated with response times in the predicted directions. 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 (but not the category-typicality of the image) correlated with the response time to that image, *but only when the cue-to-image pair was presented simultaneously.* That is, hearing a natural sound at the same time improved processing a particular kind of visual image: a picture depicting an object that could have made the sound that was heard. These results show that the ways in which an auditory cue influences recognition of visual images depends on both the fit of the image to the auditory cue and the time course of the presentation.

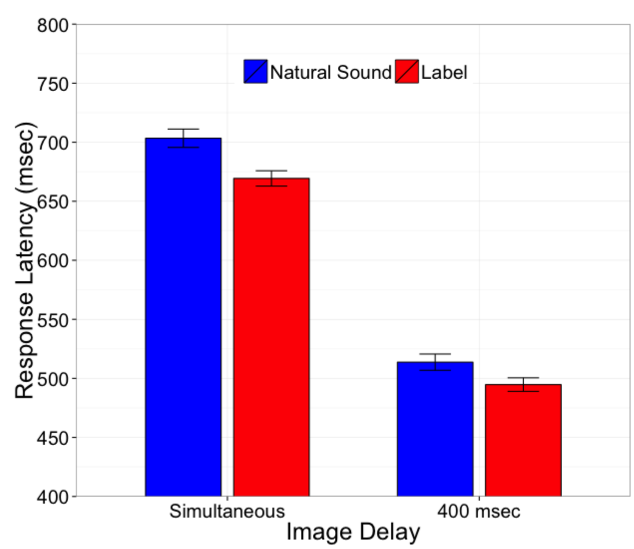


Figure 4: Interaction between image delay and cue type. Error bars denote 95% confidence intervals (Morey, 2008).

[3] This increase in response validity compared to Exp. 1 allowed us to fully counterbalance all trial variables on matching trials while keeping the length of the experiment manageable.

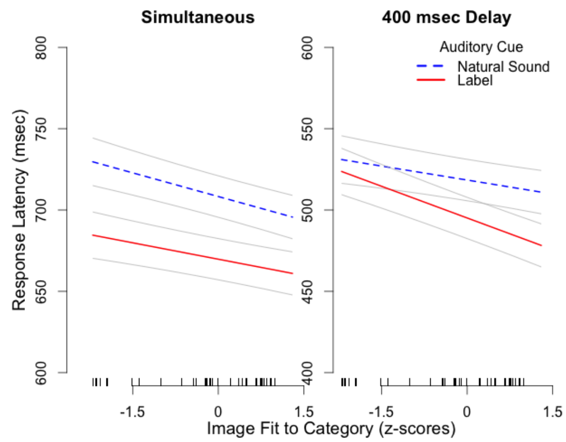


Figure 5A: Response times were faster for more category typical images across both cue types and both image delays.

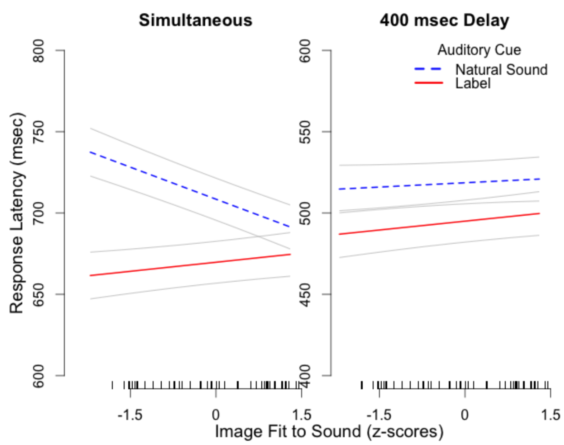


Figure 5B: Response times were faster for sound-matched images when a natural sound cue was presented simultaneously. Error bars denote ±1 SE of predicted values.

## General Discussion

In two experiments we demonstrated that verbal and nonverbal cues systematically differ in how they activate conceptual information, as tested by visual recognition of category exemplars. Experiment 1 revealed more category-typical exemplars were recognized faster following a spoken label cue but not a natural sound. In addition, Experiment 1 revealed that more sound-matched exemplars were not recognized faster following either auditory cue. Importantly, responses following natural sound cues did not vary as a function of category-typicality while those following labels did, suggesting that verbal and nonverbal cues are indeed operating on different typicality gradients. Experiment 2 added to these results with a fuller stimulus set and varying image delays. In Experiment 2, but not in Experiment 1, responses following natural sounds varied with category-typicality. This result is likely due to the shorter delays in Experiment 2 (see Lupyan & Thompson-Schill, 2012). However, the intriguing result from Experiment 2 was that responses to natural sounds did indeed vary by the sound-match of the image, but the relationship was time sensitive. In particular, sound-matched exemplars were recognized faster following a natural sound only during simultaneous presentation.

Together, the two experiments reported here highlight the role of multisensory integration as a feature of what we have called "motivated" cues. We associate barking with dogs, but the bark informs us about the *particular* dog that made it—a deeper bark is likely to come from a larger dog, and hearing a bark usually temporally coincides with seeing the actual animal. Such contingencies result in audiovisual integration of simultaneous cues that improves detection and recognition more than the sum of the unimodal cues alone (Chen & Spence, 2011; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004). In contrast, word-to-referent mappings are "unmotivated" (cf. Hockett, 1966). Saying "dog" in a deeper voice does not *systematically* imply a larger or angrier dog.[4] So, even though both "dog" and a dog-bark may be unambiguously associated with dogs, the dog-bark communicates information about category specifics (e.g., the type of dog). The word "dog", while varying systematically with aspects of the *speaker* (e.g., the lower the pitch, the more likely the speaker is to be male), does *not* systematically vary with the referent. We can talk about particular dogs, of course, but the word "dog" can and often does remain categorical, abstract.

In addition, these findings establish a heretofore underappreciated relationship between an auditory cue and a sound-matched image in similar cognitive processing tasks. Many attempts to compare semantic and conceptual processing of labels and natural sounds may have conflated category typicality with sound-match by only including a single, typical exemplar per category (e.g., Saygin, Dick, & Bates, 2005).

**Conclusion** We found that verbal and nonverbal cues activate differentiable conceptual representations evident in patterns of RTs 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.

[4] There is intriguing evidence that sometimes, speakers do modulate pronunciations of words in a graded fashion and that listeners are sensitive to these modulations (Nuckolls, 1999; Parise & Pavani, 2011), e.g., speaking faster or slower to describe a faster or slower moving object (Shintel, Nusbaum, & Okrent, 2006). Language can be easily stripped of these features however (e.g., in written form) while still being perfectly understandable.

## References


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.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4.

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.

Chen, Y. C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *JEP: Human Perception and Performance*, *37*(5), 1554–1568.

Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, *1115*(1), 92–107.

Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Bentrovato, S., D'Amico, S., et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience*, *19*(5), 799–816.

Fowler, C. A. (1990). Sound-producing 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.

Goll, J. C., Crutch, S. J., Loo, J. H. Y., Rohrer, J. D., Frost, C., Bamiou, D. E., & Warren, J. D. (2010). Non-verbal sound processing in the primary progressive aphasias. *Brain*, *133*(1), 272–285.

Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America*, *115*(3), 1252.

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

Hoffman, P., & Ralph, M. A. L. (2013). Shapes, scents and sounds Quantifying the full multi-sensory basis of conceptual knowledge. *Neuropsychologia*, *51*(1), 14–25.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *JEP: Human Perception and Perform*, *26*(1), 279.

Laurienti, P., Kraft, R., Maldjian, J., Burdette, J., & Wallace, M. (2004). 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. doi:10.1037/0096-3445.137.2.348

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis, 1–13.

Lupyan, G., & Swingley, D. (2012). Self-directed speech affects visual search performance. *The Quarterly Journal of Experimental Psychology*, *65*(6), 1068–1085.

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.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not Just for Talking. *Psychological Science*, *18*(12), 1077–1083.

Mazerolle, M. J. (2012) AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 1.26. http://CRAN.R-project.org/package=AICcmodavg.

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

Orgs, G., Lange, K., Dombrowski, J.-H., & Heil, M. (2008). N400-effects to task-irrelevant environmental sounds: Further evidence for obligatory conceptual processing. *Neuroscience Letters*, *436*(2), 133–137.

Nuckolls, J. (1999). The Case for Sound Symbolism. *Annual Review of Anthropology*, *28*, 252, 225.

Parise, C., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research*, *214*(3), 373–380.

Saygin, A. P., Dick, F., & Bates, E. (2005). An on-line task for contrasting auditory processing in the verbal and nonverbal domains and norms for younger and older adults. *Behavior Research Methods*, *37*(1), 99–110.

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds Evidence from aphasia. *Brain*, *126*(4), 928–945.

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.

Stuart, G. P., & Jones, D. M. (1995). Priming the identification of environmental sounds. *The Quarterly Journal of Experimental Psychology*, *48*(3), 741–761.

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.