MEMORY & COGNITION

Words and nonverbal sounds: differences in the activation of low-level visual information

Journal:	*Memory & Cognition*
Manuscript ID	Draft
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Falcón, Alberto; Universidad Autónoma del Estado de Morelos, Ramírez, Michelle; Universidad Autonoma de Baja California, Psychology Montano, Ulianov; Universidad Autónoma del Estado de Morelos
Keywords:	conceptual activation, language and thought, nonverbal sounds, linguistic labels

January 24, 2022

Journal of Cognitive Psychology
Editor

I am pleased to submit an original research article entitled ***Words and nonverbal sounds: differences in the activation of low-level visual information*** by Alberto Falcón, Ulianov Montano, and Michelle Ramírez for consideration for publication in the *Memory and Cognition.*

Recently, previous work has shown that words play a crucial role on the nature of the conceptual system and more specifically on how words activate conceptual knowledge. In our present work we try to contribute to show how words and nonverbal sounds differ on the conceptual activation they are promoting. We don't think we are challenging any general theory, however we do think we are contributing on establishing a more robust theoretical and methodological path for this kind of studies.

This manuscript has not been published and is not under consideration for publication elsewhere. We have no conflicts of interest to disclose.

Author contribution

Falcón and Ramírez contributed to conceptualization and design of the study. Ramírez contributed to the acquisition of data. Falcón and Ramírez contributed to organization of data. Falcón contributed to analyses. Montano wrote the first draft. All authors contributed to interpretation of results, wrote sections of the manuscript, manuscript revision, read and approved the submitted version.

Data availability

The datasets generated during and/or analysed during the current study are available in the WORDS AND SOUNDS repository, https://github.com/falcognition/Words-and-sounds.git Best regards,

Alberto Falcón, PhD.
Professor,
falcon@uaem.mx
Laboratorio de Comunicación Humana y Cognición
Facultad de Comunicación Humana
Universidad Autónoma del Estado de Morelos
Morelos, México

Words and nonverbal sounds: differences in the activation of low-level visual information

FALCÓN, Alberto[1*] (Orcid: 0000-0003-4884-6518)

MONTANO, Ulianov[1] (Orcid: 0000-0003-0603-7915)

RAMÍREZ, Michelle[2] (Orcid: 0000-0003-0582-5653)

[1]Laboratorio de Comunicación Humana y Cognición, Facultad de Comunicación Humana, Universidad Autónoma del Estado de Morelos, Cuernavaca, Mor, México

[2]Universidad Autónoma de Baja California, Tijuana, BC, México

***** Corresponding Author

falcon@uaem.mx

Abstract

Linguistic labels have been shown to facilitate visual recognition and categorization more than nonverbal sounds do. One of the explanations of this advantage is that, in contrast with nonverbal sounds, labels activate categorical representations, while sounds activate more specific co-varying representations. While this may be a reasonable claim, we argue that a distinction of what kind of low-level visual information is being activated by these auditory cues is necessary. University students participated in an image verification task, in which they listened to a cue signal followed by an image. Participants had to decide whether the image they saw corresponded to the nonverbal sound or label they previously heard. Results showed that nonverbal sounds allow recognition of visual images as efficiently as nouns and verbs importantly depending on low level information related to the dynamics of the image. Thus the present study presents evidence that action visual information plays a role in the representations activated by words and nonverbal sounds. In addition, the evidence shows that nonverbal sounds seem to activate information similar to the one activated by verbs. These findings may be relevant for having a more accurate idea of exactly what kind of low-level information is entailed in the representations activated by different kind of sounds (i.e., verbal vs. nonverbal).

Keywords: conceptual activation, language and thought, words, nonverbal sounds, linguistic labels

Words and nonverbal sounds: differences in the activation of low-level visual information

1. Introduction

Evidence shows that there is a bidirectional influence between thought and language. Among the models of how thought and language interact, the one by Lupyan and collaborators offers a simple yet powerful account of the deep influence of language on cognition (see Lupyan & Lewis, 2009). The model states that, firstly, linguistic labels and environmental sounds activate different types of mental representations, and secondly, that while the representations activated by sounds are motivated (i.e., they are source-correlated, specific and detail-rich representations), the representations activated by labels are categorical (i.e., they are more abstract and oblivious to the specificities of exemplars). Now, studies have also shown that linguistic labels do activate more than just categorical information. Basic label categories words (e.g., dog) are known to be better cues than and super-ordinate level categories (e.g., animal). There is also evidence that words activate specific visual information such as shape (see below). Since words seem to activate specific information such as shape or basic/superordinate category level, and since language contain words such as verbs whose explicit purpose is to refer to actions, it is likely that information about whether the referred object is in a steady state or is undergoing a dynamic change can also be activated by words (as well as by nonverbal sounds). To investigate this possibility, we tested the hypothesis that action information may be a component of the representations activated by *verbs* and nonverbal sounds. Thus, the present study intends to determine whether action information is a constituent of conceptual information activated by sounds and verbs, in contrast to conceptual information activated by nouns. In the following, we elaborate and document the ideas introduced in this brief preview.

Language is for communication. This view is common wisdom and is also pervasive in the scientific study of language (Gleitman & Papafragou, 2005; Klemfuss et al., 2012) . An alternative view is that language is for thinking (Lupyan

& Thompson-Schill, 2012). In recent years, empirical evidence has helped consolidate this last view. A growing number of studies show that conceptual and linguistic development influence each other. It has been observed that conceptual development constrains linguistic development (Gentner, 1982; Snedeker & Gleitman, 2004). More significantly, learning words impacts conceptual development (Casasola, 2005; Gentner & Goldin-Meadow, 2003; Gumperz & Levinson, 1991; Levinson, 1997; Lupyan, Rakison, & McClelland, 2007; Spelke, 2003; Waxman & Markow, 1995; Yoshida & Smith, 2005). For example, Casasola (2005) found that infants form abstract spatial categories only when presented with a word they know. Lupyan, Rakison, and McClelland (2007) found that learning labels for novel categories facilitated category learning.

On the activation of concepts, Lupyan and collaborators have found significant results. For instance, they have found that labels, compared to nonverbal sounds, seem to facilitate recognition of images (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). Lupyan and collaborators have suggested two hypotheses in connection to how words are related to concepts. First, that words and environmental sounds activate different types of representations. Second, that the representations activated by sounds are motivated, that is the properties of their representations co-vary with the properties of the sound source, while the representations activated by labels are categorical, they are more abstract and insensitive to the properties of the exemplars that can be labeled with the word. According to Edmiston and Lupyan (2015) sounds

used as cues lead to slower recognition than words, words show an advantage, because sounds activate a more specific representation.

Now, the mechanisms by which language and cognition interact are still not completely understood. The hypothesis that the mental representations activated by words are distinctively categorical, different from representations activated by nonverbal sounds, has the advantages of being simple, of capturing the intuition that words are special, and has a wide empirical support.

However, there is evidence that information other that categorical information plays a significant role in the representations activated by words. For instance, the cascade model of visual-linguistic interactions (e.g., Ferreira et al., 2008; Huettig & Altmann, 2005; Huettig & McQueen, 2007; Huettig et al., 2011), suggests that words evoke both visual and semantic representations: eye tracking studies that have shown that, upon hearing a word, subjects presented with an array of images show a bias in orienting toward semantically related objects (e.g., a picture of socks after hearing the target word *belt*). More evidence that extra-categorical information plays a role in word-activated representations can be found in Vales and Smith (2015), who showed that 3-year-old children were faster in identifying targets when hearing appropriate words. They also observed a significant interaction between the factors of shape and hearing words: labels enhanced target detection among shape-dissimilar distractors. More recently, Noorman (2018), using behavioral and EEG measurements in an image verification task, found that hearing a word activates representations of its referent's shape, which impacts the verification task during the initial 100ms, and that non-categorical visual information affects processing in later stages.

The foregoing results show that words activate specific non-categorical information. This raises the question of whether auditory cues may also activate action information, since the very

production of sounds is usually associated to actions or dynamic exchanges of energy. This fact seems to have been neglected in previous studies. For instance one of the hypotheses in Edmiston and Lupyan (2015) is that sounds lawfully co-vary with their source. Edmiston and Lupyan argue that a guitar sound, for example, informs the listener about the type of guitar (electric, acoustic, etc.) that is the source of the sound. By contrast the word ‘‘guitar’’ abstracts away the type of guitar.

Now, although Edmiston and Lupyan's observation about the difference between the sound of a guitar and the word "Guitar" is mostly correct, we suggest that their observation can be qualified: we wish to draw attention to the fact that sounds are usually produced by *actions* and not merely by *objects*. In producing sounds, some sort of active exchange of energy must occur. For instance, a guitar by itself does not emit a strumming sound: it is only in the process of being strummed by someone that the guitar emits the strumming sound. Similarly, a dog (which is another stimulus discussed and tested by Edmiston and Lupyan) in a resting state does not emit a sound: it is only the dog in the act of barking that emits a bark sound. Electric guitars do not possess the feature of sounding like an electric guitar strum. It is not the guitar *as an object*, but rather the *action of strumming* that is accompanied by a strumming sound. In general, objects in a steady resting state do not emit sound. Sounds are usually the result of dynamic processes. In other words, sounds are produced by agents or objects undergoing a process of physical change. Sounds only co-occur with acts, actions or changes, at least in the typical experience of people. This should draw attention to the possibility that, since sounds co-occur with actions, sounds may systematically activate action information.

Thus, controlling action information in an experiment's materials is important to avoid possible confounds in image verification tasks. However, previous studies did not implement complete controls of action information (Edmiston & Lupyan, 2015).

The foregoing discussion raises the question not only of whether action information plays a role in the concepts activated by auditory cues, but also of how special words are (that is, how effective respect to other types of cues) when controlling for action information. In other words, whether the effectiveness of words (compared to sounds) is preserved once action information has been adequately controlled. Moreover, since verbs are words that carry action information, and in that respect are in principle more similar to environmental sounds in the information they activate, a closely connected question is how effective verbs are when compared to sounds. Thus, we have reasons to believe that the conclusions obtained by previous studies (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012) can be further refined and improved by explicitly controlling for the action content in sounds and imagery used in the verification task, and by testing not only nouns and nonverbal sounds, but also verbs as auditory cues.

Hence, to test the effect of action information, we use an image verification task in which images and auditory cues were congruent or incongruent with respect to specific actions. For instance, a bark sound is action-congruent with the image of a dog in the act of barking, but it is action-incongruent with the image of a sleeping dog. The efficacy of sounds, according to Lupyan and collaborators' hypothesis, should be sensitive to action-congruence, but the efficacy of words should not. However, if words, as we have documented, activate specific action information, it is possible that the efficacy of words also vary depending on action-congruence. Verbs (at least some of them) are words specifically devoted to refer to actions. We test verbs by manipulating the congruence between verbs and images. Thus, we test two types of linguistic

cues: nous, which are cues that, in principle, are free from action information, and, verbs, which are cues that are rich in action information. Images of basic categories congruent with these cues were selected to be used in the image verification task.

Therefore, in order to confirm that words activate special types of representation (contrasting with the specific concepts activated by sounds) we should expect that nouns result in faster verification responses (compared to nonverbal sounds): 1) when presented with incongruent action images, 2) when presented with figurative images and 3) at least similar verification speed when presented with congruent action and realistic images. Regarding verbs, these should cause a faster verification response when images of congruent actions are presented, independently of the depiction style.

2. Method

The experiment was an adaptation of image verification experiments previously conducted by Lupyan and collaborators (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012).

We manipulated the congruence between the action information in images and auditory cues; we also manipulated congruence in depiction style, and used verbs as an additional type of cue. In these manipulations we implemented consistent criteria for sound-image congruence, taking into account the specificities of images. This should contribute to a more precise understanding of how special words are when systematically controlling for relevant features such as action information type of image, and also when comparing nouns, nonverbal sounds and verbs.

2.1. Participants.

37 university students (13 men and 24 women) of the Faculty of Human Communication of the Autonomous University of the state of Morelos, in the city of Cuernavaca, Mexico, volunteered in the experiment. Participants were between 18 and 22 years old.

2.2. Materials.

2.2.1. Auditory Stimuli

Auditory stimuli were created from the basic categories *wolf, horse, dog, lion* and *person*. These categories were selected making sure that they have associated auditory cues in the form of nouns, verbs, and nonverbal sounds. For instance, the auditory cues created from the category *dog* were the noun "Dog", the verb "Bark" and the characteristic sound of a bark. In order to test the effect of action information, in addition to the usual basic level categories, in the category person we included four variations consisting of a person performing acts that also have the three types of recognizable auditory cues. The four act-variations comprised a person *yelling, clapping, whistling* and *coughing.*

Nouns. The noun cues comprised the words *wolf, horse, dog, lion* and *person*.

Verbs. The verb cues comprised the words *howl, neigh, bark, roar, yell, clap, whistle* and *cough*.

Sounds. The nonverbal sounds were characteristic and recognizable sounds of the following actions: *howl, neigh, bark, roar, yell, clap, whistle* and *cough*.

Labels (nouns and verbs) where produced by a female voice that spoke native Spanish. Nonverbal sounds where obtained from free online digital sound libraries. All auditory cues where normalized in volume and equated to 800 ms in duration, except for the words howl

(1000 ms), horse (1100 ms), neigh (1100 ms), and the nonverbal sound for *howl* (1000 ms) and *neigh* (1100 ms). These exceptions were made to make sure that all the words and sounds were equally intelligible.

2.2.2. Visual stimuli

Thirty-two images were selected. Images were classified according to the combination of two dimensions of congruence with auditory cues: action congruence (the image is congruent with the action producing the sound) and depiction style (the image is realistic or figurative). Thus, four groups of eight images depicting actions were created. The four groups, characterized by the combination of the two dimensions, are as follows: realistic depiction and action congruence (in this group there were 8 different images), realistic depiction and action incongruence (8 images), figurative depiction and action congruence (8 images) and figurative depiction and action congruence (8 images). See Fig. 1.

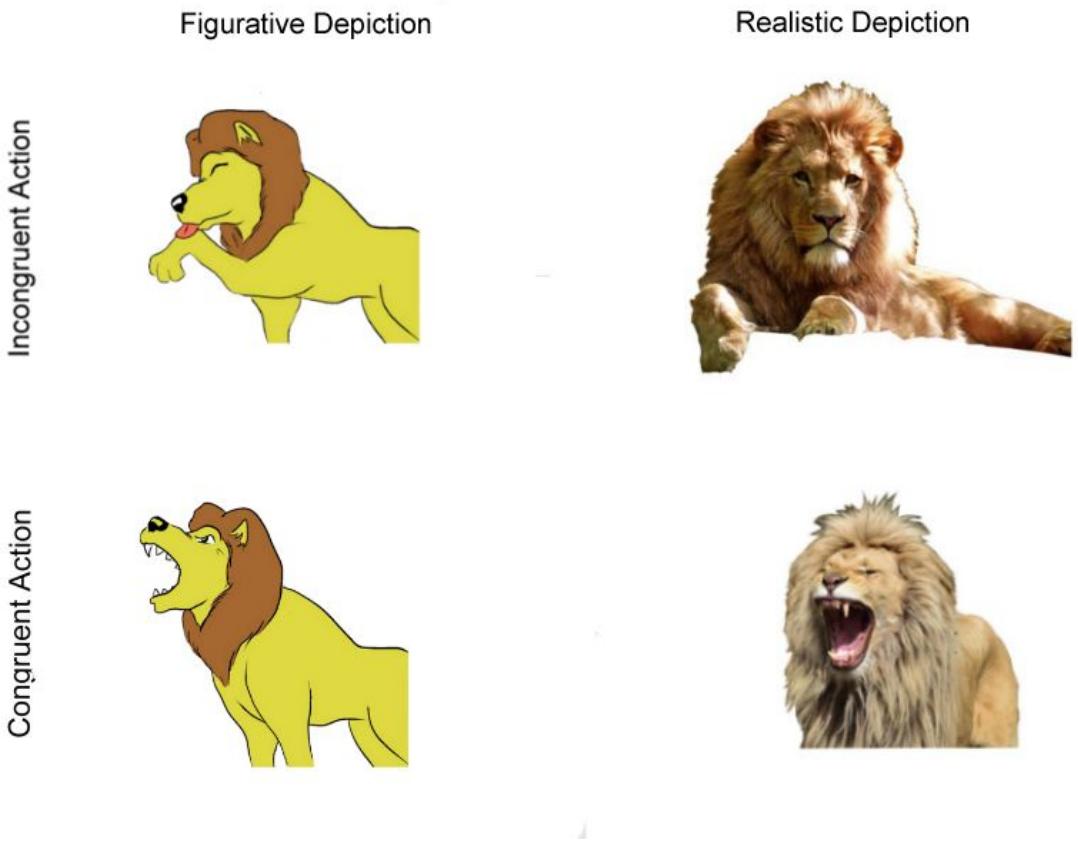

Fig. 1. Example of the four groups of visual stimulus used in the experiment. Each of the four stimuli was cued by either the noun *lion*, the verb *roar* or a roar sound in corresponding trials.

Realistic vs. figurative depiction. This dimension refers to the style in which exemplars are represented in images. It is similar to the controls utilized in Experiment 3 (Edmiston & Lupyan, 2015), in which the experimenters asked on-line participants to evaluate the congruence between sound and image. One of the dimensions in which evaluators seemed to rate the congruence of images was depending on whether the image was a realistic photograph or a cartoon drawing (see Experiment 3, Edmiston & Lupyan, 2015, and their supplemented material).

Based on the latter, we controlled whether the image of the exemplar, (e.g., a lion), is presented in a realistic style (e.g., a photograph of a lion) or in a figurative style (e.g., a cartoon drawing of a lion). Realistic depiction is considered the congruent type of depiction, while figurative depiction is the incongruent type of depiction. By utilizing this second dimension of congruence, we have a common ground with previous studies, and an additional variable to observe the results of manipulation.

Sound-action congruence. We use two realistic images (photographs) per category. One of the two images shows an agent performing an action congruent with a sound or a verb. For instance, the image features a lion in the act of roaring, which is congruent with the roar sound and with the verb *roar*. The other image shows the agent performing an action that is not congruent with the sound or verb. For instance, the image features a lion sleeping, which is not congruent with hearing a roar sound nor with the verb *roar*. We also used two figurative images showing cartoon drawings of the agents. Following the same logic as with the realistic images, two cartoon images were created, showing congruent or incongruent action. For instance, one cartoon drawing of a lion roaring and one of a lion sleeping. Thus, sound-action congruence/incongruence refers to an image in which the agent is presented performing an action that corresponds to a given verb and non verbal sound (e.g., lion roaring while hearing a lion's roar or the verb *roar*) or an agent performing an action different to the one that corresponds to the verb or nonverbal sound (e.g., lion sleeping while hearing a lion's roar or the verb *roar*).

All of the realistic depiction style images where downloaded from the Internet. Figurative style images were purposely designed for this study. Images were edited to have a white background that allowed the agent to be visible and without visual distractors. For the

figurative depiction style, images were purposely drawn and digitally designed for this study with color with *clip studio paint* software, procuring that the design of the drawn images were equal in the way the action was presented to the photographs. Similar dimensions, angles, positions and colors were considered for each image and photograph of the same agent for all categories.

2.3. Validation of stimuli

Imaginability. To establish the reliability of the congruence between image and verb cue, test were carried out. Twenty-five participants were presented with a verb cue following a question that asked them to imagine the image that corresponded to the action that was previously mentioned. Subsequently, an image was shown and participants were asked evaluate the resemblance between the image and what they have imagined, by rating images in a scale from 1 (least resemblance) to 5 (most resemblance).

Familiarity. Words and sounds were presented to twenty-five participants and they were asked to rate their familiarity with such stimuli on a scale from 1 to 10.

Discriminability. A discrimination test was performed for both sounds and for images independently. To verify that subjects specifically knew what action corresponded to the nonverbal sound, 25 participants were presented with a sound and were asked to name the action that resembled the sound. A different sample of 15 participants evaluated discriminability of images. Participants were presented with an image and were asked to name the action that resembled the image.

2.4. Experiment

The experiment was an image verification task. Three kinds of auditory cues levels, namely, nouns, nonverbal sounds and verbs, were tested. Visual stimuli comprised the thirty-two images described in the previous section, varying in the two dimensions of action congruence and depiction style.

Each participant completed 240 trials: 8 categories x 2 levels of action congruence x 2 levels of depiction style x 3 Cue Types (noun, nonverbal sound and verb) x 2 repeats.

2.5. Procedure

On each trial of the verification task each participant listened to a cue signal followed by an image. Each participant was instructed to decide whether the image they saw corresponded to the sound they previously heard. All participants where tested in an individual room with enough space, well light free from auditory disturbances. A laptop with a touchscreen was used to observe and interact during the task by touching the laptop monitor. Each participant sat at about 15” from the laptop monitor. Cue sounds were heard via headphones.

All trials began with a 250 ms fixation cross, following immediately by an auditory cue. The image appeared in the center of the monitor 400ms after hearing the auditory cue and remained in the monitor for up to 4 seconds. Each participant competed 240 trials.

Each participant was instructed to decide if the image they saw matched the cue they heard. Participants were instructed to touch the bottom-left quadrant of the touchscreen when images and sounds matched, and the bottom right quadrant when image and cue mismatched.

3. Results

Analyses were preformed based on the correct responses of the matching trials that met the following criteria: RTs should not be longer than 2500 ms (1% of the total) nor shorter than 250 ms (all correct trials met these criteria).

Data were analyzed with a 3 (cue type; noun, nonverbal sound and verb) x 2 (action congruence and no congruence) x 2 (realistic and figurative depiction style) within-subjects analysis of variance (ANOVA).

Results showed a statistically significant main effect for type of auditory cue; F (2, 54) = 8.788, $p < .001$, and for action congruence F (1, 27) = 11.089, $p < 0.01$. Marginally significant differences were found for the depiction style factor (realistic or figurative); F (1, 27) = 3.833, $p = .061$. Also, a marginal significant interaction between cue and depiction style; F (2, 54) = 2.617, $p = .082$, a significant interaction between cue and action; F (2, 54) = 13.642, $p < .001$, and a statistically significant interaction with cue, depiction style and action congruence; F (2, 54) = 4.315, $p = .018$.

3.1. Analyses for incongruent action trials

3.1.1. Incongruent action and Realistic Depiction

T-test analyses to compare responses to each cue (noun, nonverbal sound and verb) for trials with images featuring incongruent action and realistic depiction style were conducted. These analyses showed significant differences; $t(32) = -2.962$, $p = .006$, $d=.89$ in RTs between noun (M = 1013; SD = 190) and nonverbal sound (M = 1134; SD = 30), between noun cue (M = 1023; SD = 199) and verb (M = 1250; SD = 294); $t(34) = -6.825$, $p < .001$, $d = .90$, and also

between the nonverbal sound cue (M = 1100; SD = 242), and verb (M =1237; SD = .300), $t(31) = -3.105$, p = .004, $d = .50$. These differences are shown in Fig 2.

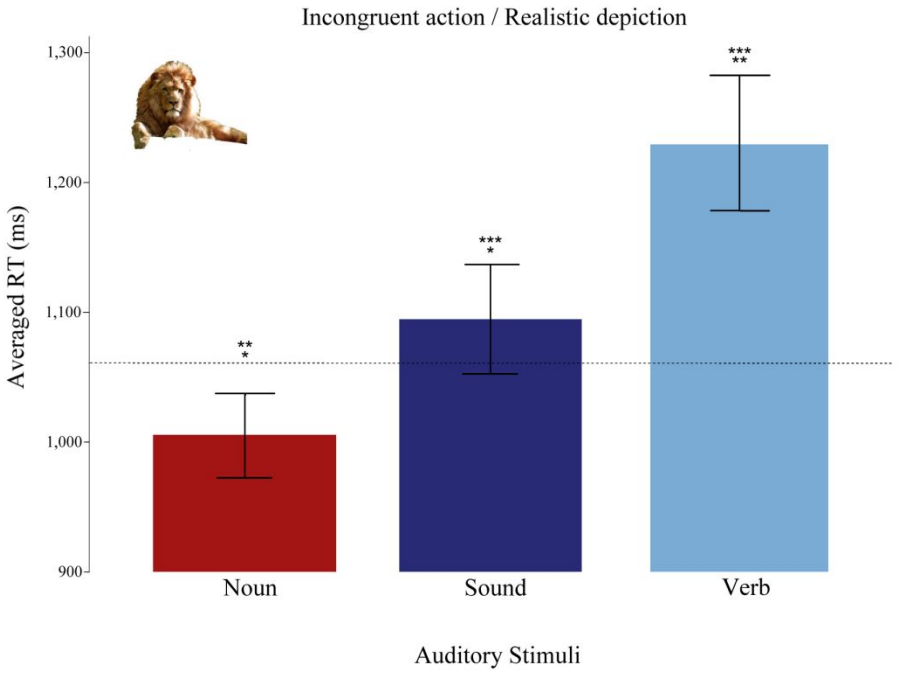

Fig 2. RTs in the picture verification performance for incongruent action and realistic depiction style with each type of cue (nonverbal sound, noun and verb (only correct "yes" responses only). Dotted line represents the averaged RT (1,078 ms) across all conditions. Pairs of combination of asterisks indicate a significant difference with p > .05)

3.1.2. Incongruent action and Figurative Depiction

Similar analyses were conducted for the three types of cue signals, with images featuring now incongruent action and figurative depiction. A significant difference in RTs; $t(33) = -2.275$, p = .030, $d = .37$ between noun (M = 1073; SD = 214) and nonverbal sound (M = 1172; SD = 304) was found. Significant differences; $t(28) = -3.313$, p = .003. $d = .53$ were also found

between noun cue (M = 1041; SD = 188) and verb (M = 1196; SD = 360). Results are illustrated in Fig 3.

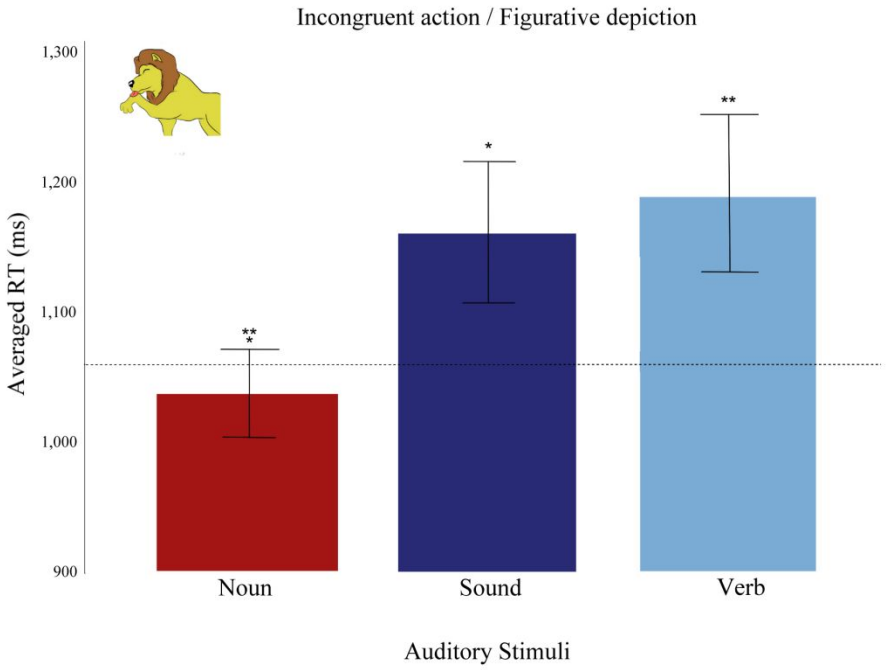

Fig 3. RTs in the picture verification performance for incongruent action and figurative depiction style with each type cue (nonverbal sound, noun and verb (only correct "yes" responses only). Dotted line represents the averaged RT (1078 ms) across all conditions. Pairs of combinations of asterisks indicate a significant difference with $p > .05$.

Overall results from these analyses of responses to images with *incongruent action* information show that nouns have a significant advantage in facilitating image recognition over verbs and sounds (nouns are the most effective facilitators). These results replicate previous findings in Lupyan and Thompson-Schill, 2012. Recall that such results suggested that words seem to activate special types of representations. Now, it must be noticed that in the present

study this effect was observed in the case of *incongruent* actions. Below we show that the effect was not observed with congruent action images.

3.2. Analyses for congruent action trials

3.2.1. Congruent action and Realistic Depiction

Further analyses to determine differential effects in performance elicited by images with *congruent actions* were conducted. For the condition of congruent action and realistic depiction style, Fig. 4 shows mean RTs for nouns, sounds and verb s. Contrary to the results in the incongruent action condition, and also contrary to the results in previous studies, no significant difference between nouns (M = 1110; SD = 311) and nonverbal sounds (M = 1058; SD = 154) was observed; $t(35) = 1.305$, p = .20, $d = .21$. In other words, no advantage for nouns over sounds as recognition facilitators was observed, this is a noteworthy results. Moreover, significant differences between nouns (M = 1110; SD = 311) and verbs (M = 1032; SD = 158) was observed; $t(35) = 2.131$, p = .040, d = .31, verbs being more effective facilitators than nouns. No significant differences were observed between nonverbal sounds (M = 1058; SD = 154) and verbs (M = 1032; SD = 158); $t(36) = 1.605$, p = .117, $d = .16$.

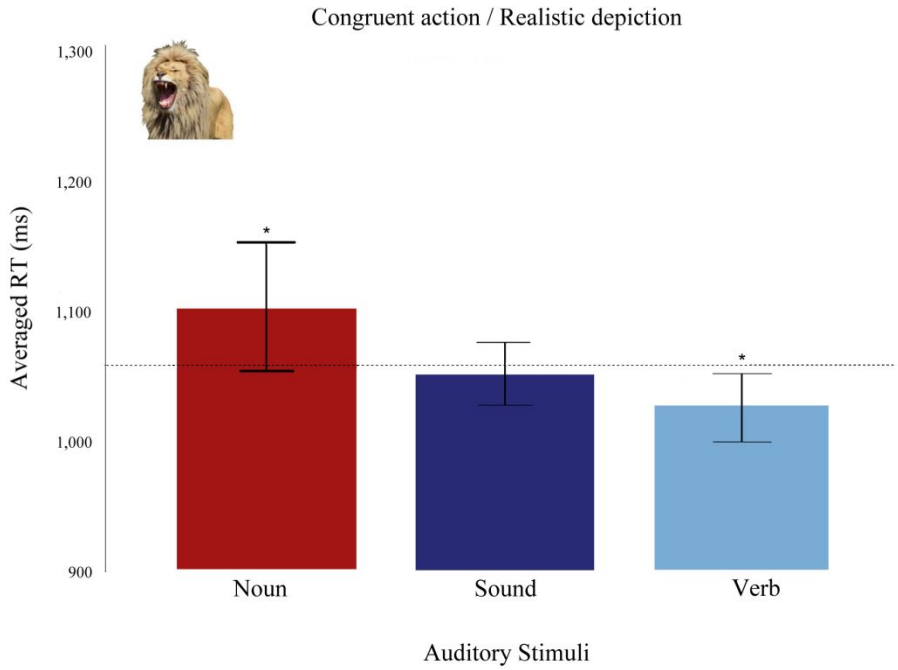

Fig 4. RTs in the picture verification performance for congruent action and realistic depiction style with each type cue (nonverbal sound, noun and verb (only correct "yes" responses only). Dotted line represents the averaged RT (1,078 ms) across all conditions. Pairs of combinations of asterisks indicate a significant difference with p > .05.

3.2.2. Congruent action and Figurative Depiction

For images in the condition of congruent action and figurative depiction style, similar analyses were conducted. Fig. 5 shows mean RTs for this analysis. A significant difference between the noun (M = 1162; SD = 320) and sound cue (M = 1031; SD = 148) was observed; $t(35) = 2.919$, p = .006, $d = .52$, showing that sounds are more effective than nouns. This is again a relevant result, since it contradicts the idea that words have an advantage over sounds in facilitating image recognition. Differences between noun (M = 1162; SD = 320) and verb (M =

1044; SD = 178) were also found; $t(35) = 2.921$, $p = .006$, $d = .45$, which shows that verbs are more effective facilitators than nouns. No significant differences were found between sound and verb cues.

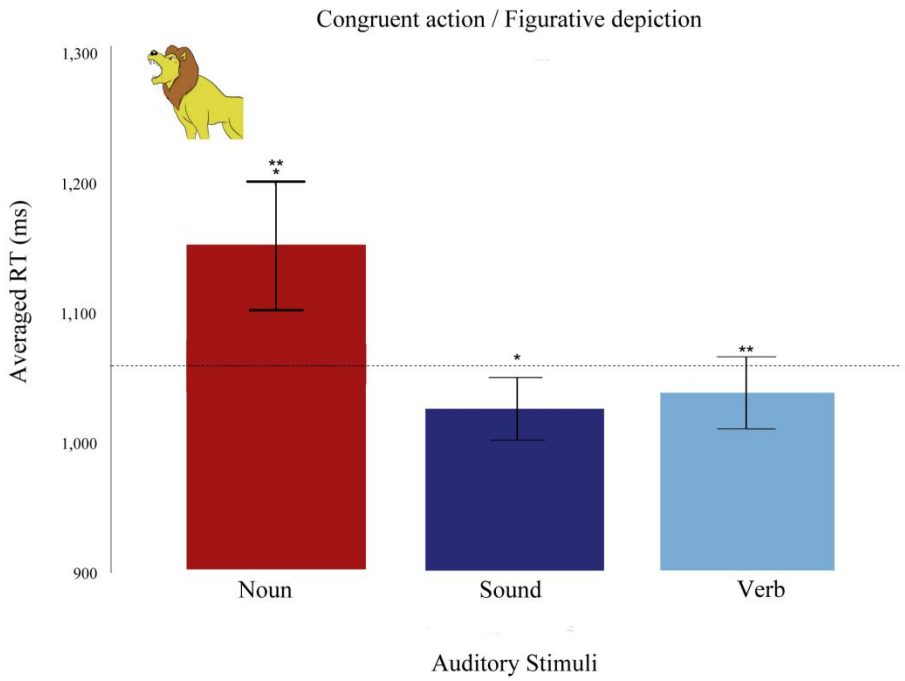

Fig 5. RTs in the picture verification performance for the congruent action and figurative agent with each type cue (nonverbal sound, noun and verb (only correct "yes" responses only). Dotted line represents the averaged RT (1,078 ms) across all conditions. Pairs of combinations of asterisks indicate a significant difference with $p > .05$.

Contrary to previous findings (Lupyan & Thompson-Schill, 2012), results do not show a systematic advantage of nouns over nonverbal sounds and verbs when the materials exhibit action congruence. Moreover, an advantage favoring nonverbal sounds was observed (see Fig. 6). We further discuss these findings in the following section.

4. Discussion and conclusions

Words are known to facilitate categorization and visual recognition, especially compared to nonverbal sounds. The mechanisms liable for the facilitation advantage of words over sounds are not completely understood, but several studies have suggested that words and sounds activate different types of representations. In particular, words activate categorical (exemplar-agnostic) representations. However, many studies also suggest that words activate the visual representation of words' referents. In this study we investigate whether auditory cues activate representations of actions. Our results show that nouns show an advantage over verbs and sound, but only in the two conditions of action incongruence. However, under action congruence conditions, verbs and sounds show a consistent advantage. This confirms our prediction that sounds and verbs would show an advantage once action information in images were adequately controlled. However, we expected the effect elicited by nouns to be unaffected by action information, the results have disconfirmed this prediction and, by contrast, they present a more complex situation.

Several studies have shown that nouns possess an advantage over nonverbal sounds and other types of labels. From such results, they suggested that nouns activate special types of representations (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). We have replicated the conditions and factors in the antecedent studies, and further controlled the factors of action congruence and depiction style. In the action incongruent trials our results are consistent with Lupyan's findings: nouns showed a significant advantage compared to the rest of the cues, even when we manipulated depiction style (which is a variable that captures whether an image is realistic representation of its referent). However, when the images in the verification task featured information congruent with the actions associated to the auditory cues, the noun's

advantage vanished, and an advantage for sounds and verbs was observed. These results suggest that the conclusion that labels activate a special kind of representations (or that previous results demonstrate "the power of words", as implied in Lupyan & Thompson-Schill, 2012) is supported only under specific conditions.

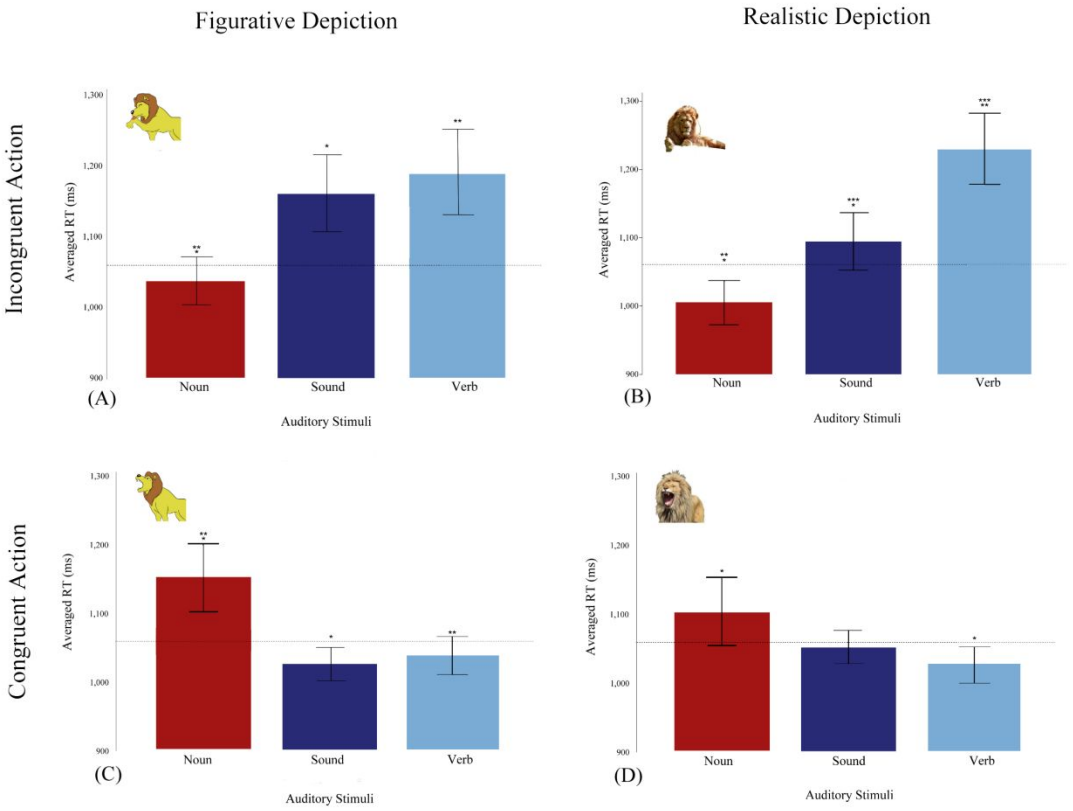

Fig. 6. Overall results on the RTs of the four experimental conditions.

In this study we tested the effect of verbs, which is an issue scarcely addressed in the literature. To probe the differences and commonalities between verbs and nonverbal sounds, we controlled congruence between images and sounds by using visual information typically mapped into nonverbal sounds, nouns and verbs, carefully manipulating the action content of images.

According to Lupyan's suggestions (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012) we expected that (since images controlled to show actions typically associated to verbs), verbs would facilitate image verification, compared to sounds. However, RTs for verbs and sounds showed no significant differences. Thus, regarding verbs we cannot conclude that there is an advantage of labels that denote action over sounds, contrasting thus with previous findings (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). Our results show that nouns show an advantage over verbs only when the action in the image is incongruent with the cues. When there is action in the image is congruent, verbs show a clear advantage over nouns. Thus, our results disconfirm, firstly, the suggestion that verbs are better facilitators than sounds, and, secondly, that nouns are better facilitator than nouns. Nouns and verbs show similar effects in every condition. And verbs are better facilitators than nouns under action congruence conditions.

4.1 Representation activated

A way of explaining the results is to revisit Lupyan's suggestion that words activate special types of representation. We may extend this idea by suggesting that different types of words activate different types of representations. Nouns, for instance, may activate representations with information useful for quickly diagnosing whether a given exemplar belongs to the noun's category. Verbs, by contrast, may activate information associated to actions of exemplars in a dynamic state. Moreover, the information activated by verbs may also be similar to the information activated by sounds. In this sense, different types of words may activate information associated to achieving certain perceptual goal. Nouns may activate information relevant to achieving the goal of categorizing exemplar, whereas verbs may activate

information relevant to achieving the goal of reacting to a dynamic environment. This is also consistent with the fact that we recognize that different types of words, nouns, verbs, adjectives, adverbs, have different linguistic functions. Words may activate representation containing visual information relevant performing such functions. Now, this suggestion does not entail that the information activated by words is necessarily different from the information activated by sounds, since it is possible that in achieving a given function it is necessary to use exemplar-specific information. For example, adjectives have the function of associating *specific* information to a category. For instance, in using the terms "brown dog" or "black dog", the adjectives brown and black have the function of attributing the *specific* colors brown and black. Thus, in order to perform their functions, words may activate representations that contain the necessary information.

Verbs by contrast seem to activate information similar to the information associated to sounds. Another advantage of this suggestion is that different types of words activate representations with different contents and functions is that it this suggestion can also accommodate the fact that information can be activated following different temporal profiles. Perhaps action information is activated in the onset of the response, while categorical or other types of visual information is activated at different temporal stages. This would also help to explain why under action congruence conditions nouns do not seem to show an advantage, even when compared with sounds, and that when action information is incongruent nouns show a consistent advantage in the realistic and figurative depictions. As we mentioned, perhaps this "action dominance" effect" occurs to cope with transient stimuli. This would result in a facilitation effect under action congruence conditions, but it would become an interference

under action incongruence conditions. Now, to ascertain any of the suggestions presented above, further research is necessary.

4.2. Conclusion. How special words are? And, what kind of low-level information is activated by words and sounds?

The general picture that emerges from results in the present study is that, as we hypothesized, action information appears to play a role in the representations activated by sounds and words. The presence of action information does have a significant effect across conditions in the picture verification task. This suggests that action information is a consistent element in the representations activated by auditory cues, but mainly by nonverbal sounds and verbs. Verbs and nonverbal sounds can be grouped together as auditory cues that have rich action information content.

These findings should also call attention to the possibility that the word-advantage observed by Lupyan and collaborators might have resulted from confounds in action information; and, therefore, that the different-processes-of-activation explanation might be worth revisiting. However we also dissed the possibility of extending and refining Lupyan's approach. Lupyan and collaborators interpreted their results as suggesting that hearing words or sounds activates *different kinds* of representations, and not that hearing words and sounds may involve *different process* of activation. As we discussed, this is indeed a possibility, and we further extended this idea by suggesting that verbs may activate representations of dynamic processes, and nouns activate representations of steady state objects. Moreover, we suggested that there may exist different types of representations associated to different types of words, and that their

activation may occur to achieve the different linguistic functions associated to the different types of words. This, however does not preclude the possibility that some of the representations associated to words may be similar to representation activated by non-verbal sounds. In this sense, words are not especially distinct from sounds, but rather words may simply co-opt representations adequate to the perform their functions.

Results of the present study replicate its closest antecedents in the literature, but also show that the interaction between linguistic and low-level visual information is more complex. For example, results suggest that action information has a significant role in semantic representations, and, moreover, that this role is manifested only when materials are adequately controlled for action information. This in turn, draws attention to the need to interpret previous results in terms of more complex interactions of different types of information in the activation of concepts. Our results do not necessarily imply a challenge to the view that words are different from nonverbal sounds, but rather they suggest that there are complex interactions between low and high level information. A strong body of evidence shows that words are indeed special and play a critical role in our cognition, or more specifically in the way we represent our world. However word's peculiarities must be carefully studied to more accurately show in what specific ways words are crucial in the formation and activation of conceptual representations.

5. **References**

Casasola, M. (2005). Can Language Do the Driving? The Effect of Linguistic Input on Infants' Categorization of Support Spatial Relations. *Developmental Psychology*, *41*(1), 183–192. https://doi.org/10.1037/0012-1649.41.1.183

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading Technical Report; No. 257*.

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought*. MIT Press.

Gleitman, L., & Papafragou, A. (2005). Language and thought. *Cambridge Handbook of Thinking and Reasoning*, 633–661.

Gumperz, J. J., & Levinson, S. C. (1991). Rethinking linguistic relativity. *Current Anthropology*, *32*(5), 613–623.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Huettig, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–150.

Klemfuss, N., Prinzmetal, B., & Ivry, R. B. (2012). How does language change perception: A cautionary note. *Frontiers in Psychology*, *3*, 78.

Levinson, S. C. (1997). From outer to inner space: Linguistic categories and non-linguistic thinking. *Language and Conceptualization*, 13–45.

Lupyan, G., & Lewis, M. (2019). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *34*(10), 1319-1337.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not Just for Talking: Redundant Labels Facilitate Learning of Novel Categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

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. https://doi.org/10.1037/a0024904

Noorman, S., Neville, D. A., & Simanova, I. (2018). Words affect visual perception by activating object shape representations. *Scientific Reports*, *8*(1), 14156. https://doi.org/10.1038/s41598-018-32483-2

https://doi.org/10.1016/j.cognition.2007.04.003

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. *Weaving a Lexicon*, *257294*.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. *Language in Mind: Advances in the Study of Language and Thought*, 277–311.

Vales, C., & Smith, L. B. (2015). Words, shape, visual search and visual working memory in 3-year-old children. *Developmental Science*, *18*(1), 65–79.

Waxman, S. R., & Markow, D. B. (1995). Words as Invitations to Form Categories: Evidence from 12- to 13-Month-Old Infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Yoshida, H., & Smith, L. B. (2005). Linguistic Cues Enhance the Learning of Perceptual Cues. *Psychological Science*, *16*(2), 90–95. https://doi.org/10.1111/j.0956-7976.2005.00787.x