Activation of basic-level categories in the human brain based on different cue types

Jacob Dudek, Gerrit Bartels, Elias Harjes, Joel Haupt

Cognitive Science, Universität Osnabrück

Experimental Psychology Lab

Prof. Dr. Michael Franke

July 31, 2020

Abstract

In human cognition, language serves as a means to fix beliefs and real-world instances to a mental concept. For example, seeing a robin and mapping it to the concept “bird”.

We believe that there is more to recognizing instances of categories than plain categorical mapping, i.e., receiving an input, processing, and assigning a given stimulus to its corresponding output category label. There seems to be some influence of language that elicits a different cognitive performance (Thierry, Athanasopoulos, Wiggett, Dering & Kuipers, 2009; Lupyan, 2012a; Boutonnet, Dering, Viñas-Guasch & Thierry, 2013; Lupyan & Ward, 2013; Kok & de Lange, 2014; Kok, Failing & de Lange, 2014; Francken, Kok, Hagoort & de Lange, 2015). This is also mirrored in the *reaction time* of recognizing the *basic-level category*[1] of an object (Edmiston & Lupyan, 2015). Likewise, in our study, we found a difference in reaction times of linking *verbal labels* (spoken words) and *environmental sounds* (e.g., a dog bark) to a picture, which suggests that different kinds of cues activate a different mental representation. Furthermore, verbal labels improve the performance in visual recognition tasks (Lupyan, 2012) by clarifying and retrieving the corresponding category more effectively. Other features of an object or *concept*[2] (non-verbal cues) can similarly activate the target category, though, these features will be associated more closely with a specific exemplar than a basic-level category, like in the case of environmental sounds (Lupyan & Thompson-Schill, 2012).

The findings from our *confirmatory study* that replicated Experiment 1A from the paper “What makes words special? Words as unmotivated cues” (Edmiston & Lupyan, 2015), supports the theory that verbal label cues activate conceptual knowledge of basic-level categories more effectively than environmental sound cues.

Keywords:
Basic-level category, environmental sounds, verbal cues, non-verbal cues, mental representation, language, confirmatory study

0. Introduction

The theory that language activates conceptual representation in human cognition differently than non-verbal cues (e.g., environmental sounds) is an extensively studied topic in cognitive psychology. Humans tend to form their conceptual knowledge based on certain features of the object, for example, they associate an animal that has wings with the ability to fly (Rogers & McClelland, 2004; Keil, 1992). Therefore, if an animal has wings but is not able to fly (e.g., an ostrich or a kiwi), this could influence the processing speed of information, since it would require an extra step in cognitive deduction for the question whether the animal can fly. In addition, environmental sounds are an example of features that can be associated with a category.

There has been little research on how environmental sounds (e.g., sound of a cello) carry semantic meaning in comparison to language, even though both cue types carry semantic knowledge and help us in recognizing objects (Louah Sirri et al., 2020).

In Boutonnet B.’s and Lupyan G.’s study from 2015, it was shown that when being primed with a verbal label in a simple visual verification task, the conceptual representation of the category is activated more quickly compared to non-verbal cues. The expected reason for this effect is that verbal labels activate all related objects of the category instead of just one specific exemplar.

The previously mentioned verbal labels and environmental sounds as primers for basic-level categories reveal an important distinction. Following Edmiston’s and Lupyan’s (2015) terminology, we refer to environmental sounds as *motivated cues* and spoken words as *unmotivated cues* as proposed in semiotics (Kockelmann 2005). Motivated cues provide information about the specific instance of the respective category. For example, a dog bark activates conceptual knowledge about a specific instance of the category dog representing the likely *sound source*. However, spoken words are unmotivated cues, because the spoken word “dog” is an abstraction from the

[1] The basic-level category stands for the most commonly used and most culturally salient word to name an object. For example, usually people refer to a kind of furniture that you can sit on as a “chair” (the basic-level category) instead of saying the more general category “furniture” (superordinate category) or the more specific one, e.g. “desk chair” (subordinate category) etc.

[2] A concept is the mental representation of a category. It refers to all the knowledge that one has about a category.

specific instance of the category. There is no specific knowledge activated in the brain, but rather a much broader scheme for the basic-level category dog, which comprises multiple instances that are treated as if they were the same (Fig. 1). Moreover, variations in sound pitch in spoken labels will not lead to activating specific instances of the category.

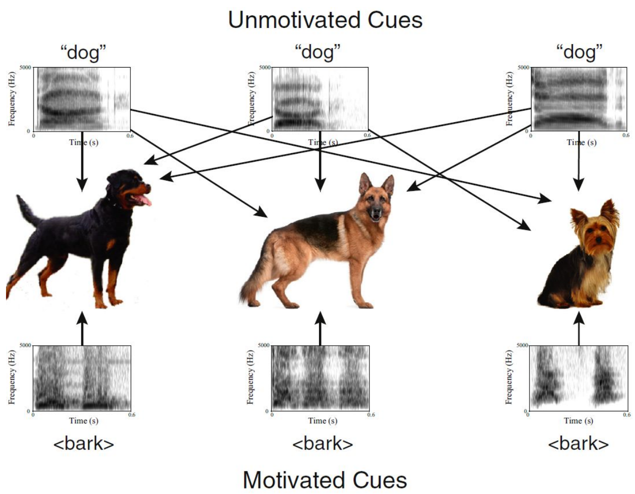

Fig. 1. Image from P. Edmiston, G. Lupyan / Cognition 143 (2015) 93–100 (Fig. 1.). The verbal cue "dog" refers to all instances of the dog while each bark only refers to one specific instance.

1. Experiment

We designed and conducted our verification task experiment according to Experiment 1A in Edmiston's and Lupyan's study (2015).

In our experiment, we tested how fast participants verified whether a given auditory cue matches the basic-level category of an instance shown in a picture. Thereby, we investigated which kind of auditory cue, environmental sound or verbal label, primed the participant more effectively. Our aim was to better understand how the conceptual categorization of objects is realized in human cognition. We expected that different cues cause different activation patterns in the brain. Consequently, from a neurobiological view, this would activate a different assembly of neurons (Lupyan & Thompson-Schill, 2012).

Hypothesis 1: We hypothesized that environmental sounds as auditory cues cause longer reaction times than spoken category labels when trying to assess and match their basic-level category to the image that is being displayed. We expected that environmental sounds activate a more specific category instance linking the sound to its likely source, although participants are instructed to only treat it as its basic-level category. This would mean environmental sounds are less effective cues than spoken category labels.

Hypothesis 2: We also hypothesized that *congruent*[3] environmental sounds exhibit a faster reaction time than *incongruent* ones because there might be less mental effort involved in assigning congruent sounds to their corresponding basic-level category.

Hypothesis 3: As *exploratory hypotheses*, we wanted to find out whether the assumed effect from Hypothesis 1 is still present if we exclude all incongruent trials or whether the hypothesized difference in reaction times is solely caused by them.

Hypothesis 4: Furthermore, we wanted to explore whether the assumed effect from Hypothesis 1 is also present when only considering *mismatching trials*.

1.1 Methods

1.1.1 Participants

A total of 35 people participated in our experiment. Participants consisted of family, friends, and fellow students.

1.1.2 Materials & Design

We used 24 color photographs consisting of 6 basic-level categories: *bird, car, dog, instrument, phone* and *typing*. Each basic-level category comprised 4 images, where two of them represent a different instance of the respective category (e.g., two images of a *Chihuahua* and two images of a *Rottweiler* for the basic-level category dog, see Fig. 2). For each category, we had one congruent and one incongruent environmental sound and one spoken word for each basic-level category (bird, car, dog, instrument, phone, typing*)* as auditory cue. The peak amplitude of all sounds, spoken words and environmental sounds, was normalized to -10db and the duration of each environmental sound was set to 1s, using the free audio

---

[3] In our experiment, a trial is congruent if the environmental sound cue does not only match with the basic-level category but also with the specific instance of the image that matches the likely sound source.
Contrarily, incongruent means that the given environmental sound cue matches with the basic-level category but not with the specific instance of the category shown in the picture.
E.g. <Chihuahua bark> congruent for the Chihuahua images, incongruent for the Rottweiler images.
<Rottweiler bark> congruent for the Rottweiler images, incongruent for the Chihuahua images. Each representing the likely sound source for their respective instance of the basic-level category "dog".

editing tool Audacity. We compressed the size of every image to 50kb by reducing the number of pixels while keeping its ratio using photoshop. To control for cue variability, we used a female and male synthetic voice for each category name to account for the bias that sound pitches could possibly have on participant's responses.

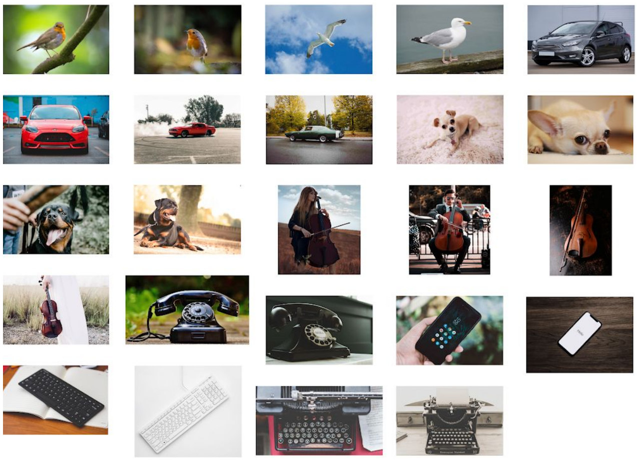

Fig. 2 Color photographs that were used in the experiment.

The utilized materials can be found here: https://github.com/jmdudek/XP-Lab2020-What-makes-words-special-Group-35/tree/master/Experiments/02_main/stimuli

1.1.3 Procedure

Participants used their own computers or laptops to view the browser-based experiment. In every trial, participants saw a fixation cross in the middle of their screen for 250 ms and immediately after were primed with a sound cue – a spoken word or an environmental sound. After the offset of the sound, a target picture appeared with a one-second delay and disappeared after the participant made a decision by judging whether the auditory cue *matched the picture*[4] that was displayed or not.

Participants indicated their decision by either pressing the "q" or "p" button on their keyboard. "q" indicated a "yes" response, for example, the sound of a cell phone ring or the spoken word "phone" followed by a picture of any phone (same basic-level category). Contrarily, "p" indicated a "no" response, for example, the sound of a cell phone ring or the spoken word "phone" followed by a dog image (different basic-level category).

Participants were instructed to minimize reaction time, while still answering as accurately as possible.

There were congruent and incongruent environmental sounds for each basic-level category, in which both were supposed to lead to a "yes" response. The experiment started with 6 practice trials in which the participants received feedback via a popup window telling them whether their given answer was correct or not. In case of an incorrect answer, the popup window also gave a brief explanation why. After the practice trials, participants were tested in 144 main trials, in which the cues matched the pictures in 50% of the cases. Trials were presented as a random sequence of factor combinations of cue type and congruence.

1.2 Results of the Experiment

We performed our statistical analysis using *R*, as well as *RStudio* and based our inference on mixed-effects bayesian linear regression modeling using the R package *brms*. For each model, we specified maximal *random effects* that were licensed by the data. In case the *95% Credible Interval* of the respective random effect included 0, we excluded the *RE* and compared the resulting model to the previous one via Leave-one-out cross-validation, using the *LOO* package. For comparison, we always included the model without the random effects structure. Out of all models we then chose the one with the highest *ELPD* score to perform the hypothesis testing. For visualization and data wrangling purposes we made use of the R packages *tidyverse*, *tidybayes*, *bayesplot* and *HDInterval*.

After filtering the data for main, correct and matching trials, reaction times shorter than 250 *ms* or longer than 1500 *ms* were removed (154 trials removed, 6.46% of total). Subsequently, we computed the natural logarithm of our *RT* values such that the resulting distribution is more well behaved and symmetric. When looking at the mean error rate of all participants (M = 96.00%), we found that everyone performed very accurately, and no one was guessing randomly.

To not constrain the parameter space we chose to use the *flat priors* that are implicitly defined by brms. All our models ran 4 chains with 2000 iterations, 1000 warm up samples and a thinning factor of 1. We encountered no problems with *divergent iterations* and all chains converged with an *Rhat* value far below 1.1 for every coefficient.

[4] An auditory cue matched the picture if it belonged to the same basic-level category, i.e. any sound of a bird or the spoken word '"bird" followed by a picture of any bird.

1.2.1 First Hypothesis

"Environmental sounds cause longer reaction times than spoken category labels". 24 out of 35 participants produced faster reaction times when being primed with spoken labels. Similar results were obtained from the posterior estimates of our model in Fig. 3. After having looked at the 95% Credible Interval for the slope coefficient (Fig. 4.), we found that zero is excluded (*CrI* [17.65, 71.58], *M* = 43.79 *ms*[5]). With only positive values we can be certain that environmental sounds cause longer reaction times than spoken category labels. Hence, the data at hand lends credence to our first hypothesis.

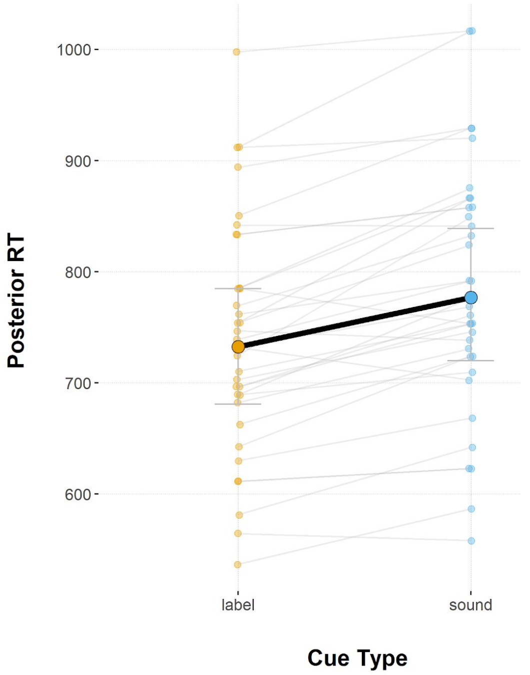

Fig. 3 Posterior means and 95% CrI for both cue types with estimates for each subject. RT in ms.

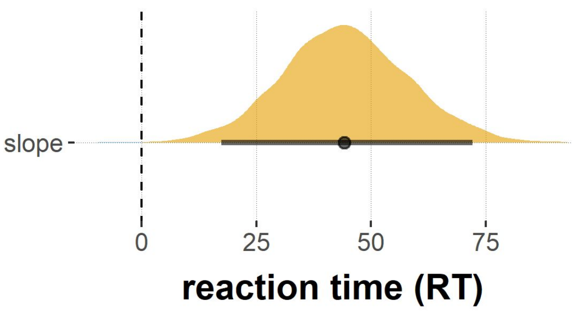

Fig. 4 Population-level estimate for the change in RT (ms) when switching from spoken labels to environmental sounds.

1.2.2 Second Hypothesis

"Congruent environmental sounds exhibit a faster reaction time than incongruent ones". Again 68.57% (*n* = 24) of our participants exhibited the hypothesized effect. The respective brms model estimated the slope coefficients' 95% *CrI* to be entirely positive (*CrI* [40.85, 96.78], *M* = 68.60 *ms*, Fig. 5) and therefore provides significant evidence that participants exhibit longer reaction times when being cued with an incongruent sound.

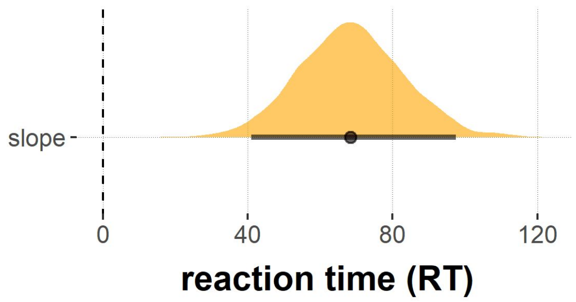

Fig. 5 Population-level estimate for the change in RT (ms) when switching from congruent sounds to incongruent ones

Summarizing the results from our two main hypotheses (Fig. 6), we found significant evidence that label cues result in faster reaction times than sound cues. Furthermore, when only considering environmental sounds we found that participants primed with congruent cues responded more quickly compared to when being primed with incongruent ones.

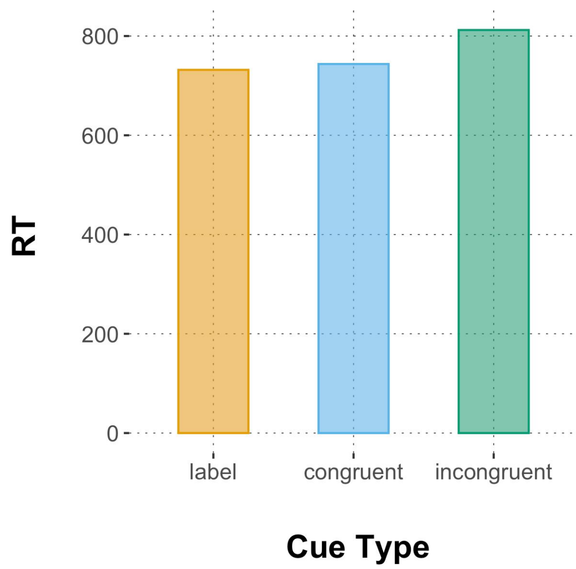

Fig. 6 Estimated mean RTs (ms) for label vs congruent vs incongruent auditory cues

[5] Since we log-transformed the reaction times, we needed to retransform the CrI's to obtain an interval in milliseconds. Use all posterior samples: exp(intercept+slope)-exp(intercept) -> calculate mean, lower and upper 95% CrI.

1.3 Exploratory Analysis

1.3.1 Third Hypothesis

"The assumed effect from Hypothesis 1 is still present if we exclude all incongruent trials". Using the same data set, but excluding all incongruent trials, we ran the same model as for our first hypothesis and obtained the following, non-significant, results: (*CrI* [-19.57, 46.61], *M* = 13.80 *ms*, Fig. 7). This suggests that the effect observed in 1.2.1 might solely be caused by the incongruent trials or at least the combination of congruent and incongruent ones. However, this needs to be investigated in a separate experiment.

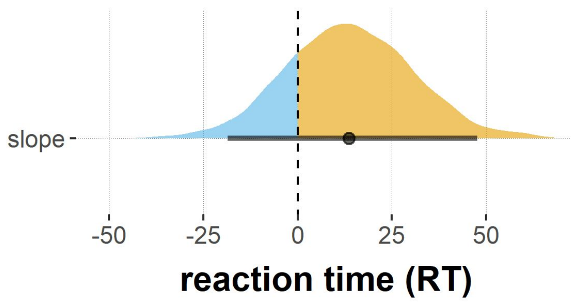

Fig. 7 Population-level estimate for the change in RT (ms) when switching from spoken labels to congruent sounds

1.3.1 Fourth Hypothesis

"Assumed effect from Hypothesis 1 is also present when only considering mismatching trials". When only considering mismatching trials we observed a similar effect as in 1.2.1. Once again, environmental sounds seem to be less effective primers compared to spoken category labels although the effect is less pronounced (*CrI* [17.05, 48.85], *M* = 32.96 *ms*). We suggest further investigation of this phenomenon in a dedicated study to confirm our findings.

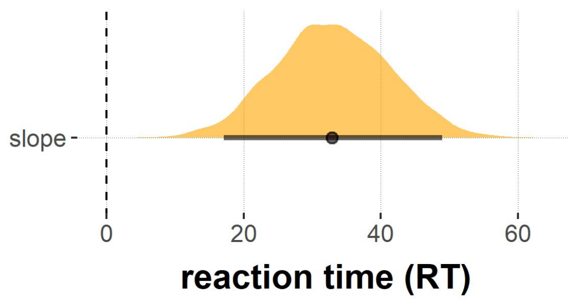

Fig. 8 Population-level estimate for the change in RT (ms) when switching from spoken labels to environmental sounds considering mismatching trials only

2. Summary & Discussion

In our experiment, we tested the theory that language activates a different mental representation than environmental sounds. To do so, we tried to confirm the phenomena that were found in Edmiston's and Lupyan's paper (2015), by replicating Experiment 1A. Additionally, we performed an exploratory analysis to get further insight into the data and the underlying processes of the hypothesized effects.

We were able to find significant evidence in support of hypothesis 1, that verbal cues cause faster reaction times than environmental sounds. This was expected, since spoken labels directly provide the basic-level category that would otherwise have to be inferred from the sound cue. This disadvantage is even more pronounced, as the inference from sounds is more difficult due to their natural ambiguity (Schmitz A., 2012). Hence, when being primed with a dog bark, the majority of our participants most probably produced longer response times due to the additional cognitive deduction step and the inherent ambiguity of environmental sounds.

We obtained similarly significant results for our second hypothesis, where we conjectured that congruent trials exhibit faster response times than incongruent ones. Our findings suggest that the main influence on the reaction times is the incongruence between the environmental sound and the depicted instance. This disparity might cause confusion when having to compare the basic-level categories, which simultaneously introduces a possible confound for the results obtained in hypothesis 1.

This confound was examined thoroughly in our exploratory analysis. By analyzing hypothesis 3, we found that participants did not perform significantly better in trials with verbal cues than in congruent trials. This suggests that the trend from hypothesis 1 was mainly caused by incongruent trials or at least the combination of congruent and incongruent ones (Fig. 6). It is important to note that we did not conduct a separate experiment to only compare verbal cues and sounds that result in congruent trials, like Experiment 1B in Edmiston's and Lupyan's paper from 2015. This might have led to a bias in reaction times, since incongruent trials might cause confusion among participants, possibly leading to longer reactions in following trials.

Lastly, we examined whether the label advantage still persists when being presented with a mismatch between the basic-level categories of the auditory cue and the depicted instance in the image.

For the same reasons as in hypothesis 1, we expected faster reaction times for labels, which is also significantly reflected in our results.

However, in order to obtain credible evidence for these results, replication in a dedicated study is necessary, since they only stem from our exploratory research.

Because our experiment was designed in English, and the majority of our participants were non-native English speakers, this might have had an impact on the understanding of the spoken words. Further investigation is needed to determine if this is a potential explanation for the eleven participants that exhibited longer reaction times for spoken words than for environmental sounds, since this contradicts with the findings of many other papers.

Further research could help to better understand the role of unmotivated cues in comparison to motivated cues. For example, it could be interesting to take the idea of the experiment and turn it around, i.e., having the participant assess specific instances instead of basic-level categories (e.g., chihuahua instead of dog). If the verbal cues still lead to a shorter reaction time it could imply that verbal labels are also more effective in priming for *subordinate categories*. If the contrary occurred, we could infer that environmental sounds convey more information about the corresponding specific instances and therefore prime more effectively.

Another study worth considering would be to use *borderline cases*[6]. According to the *prototype theory*, the membership of an exemplar is defined purely in terms of the distance from a prototype. This motivates the question, whether the difference between the reaction times of verbal labels and environmental sounds still remains if borderline cases are incorporated as instances for the basic-level categories in the experiment – classifying a penguin as the basic-level category bird by either being primed by a verbal cue or by the noise the penguin makes. This could also bring up the question of whether the verbal cues might even cause a longer reaction time than congruent environmental sounds, due to the *fuzzy membership* of the depicted instance. If this effect turned out to be significant, this could lead to the assumption that the effects found by Edmiston and Lupyan (2015) are purely based on the selection of the instances for the basic-level categories. To further investigate this problem a new study could be conducted, in which it is extensively stated that there is neither an advantage for the verbal label nor for the environmental sound. If there was still a significant difference in reaction times in favor of verbal labels compared to environmental sounds, this could add further evidence towards Edmiston's and Lupyan's (2015) hypothesis.

Findings of other papers suggest that the label advantage still holds and even amplifies when the delay between sound offset and the appearance of the target image is increased (Lupyan & Thompson-Schill, 2012). This rules out the argument that sound cues are solely less familiar cues that need more time to be processed. Moreover, it supports the assumption that verbal cues are not processed faster in the human brain but rather activate different mental representations than environmental sounds (Lupyan, 2012). The theory of different activation is also supported by another finding of an experiment where participants had to learn novel categories for previously unseen objects. As a reference, they learned either verbal cues or sound cues to associate with the categories. Even for these "alien" objects, it was found that verbal cues helped people remember the corresponding category better than sound cues did (Lupyan, Rakison & McClelland, 2007).

To summarise, by confirming the findings from Experiment 1A (Edmiston & Lupyan, 2015), we were able to add substantial support to the hypothesised effects. However, the results in our exploratory analysis differed to some extent with respect to the original findings of Experiment 1B and further produced contradicting results.

[6] Exemplars that are of far distance to the prototype

3. References

3.1 Papers

Boutonnet, B., Dering, B., Viñas-Guasch, N., & Thierry, G. (2013). Seeing objects through the language glass. Journal of Cognitive Neuroscience, 25(10), 1702-1710.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. Journal of Neuroscience, 35(25), 9329-9335.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. Cognition, 143, 93-100.

Francken, J. C., Kok, P., Hagoort, P., & De Lange, F. P. (2015). The behavioral and neural effects of language on motion perception. Journal of cognitive neuroscience, 27(1), 175-184.

Keil, F. C. (1992). Concepts, kinds, and cognitive development. mit Press.

Kok, P., & de Lange, F. P. (2014). Shape perception simultaneously up-and down regulates neural activity in the primary visual cortex. Current Biology, 24(13), 1531-1535.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. Journal of Cognitive Neuroscience, 26(7), 1546-1554.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. Frontiers in psychology, 3, 54.

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.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. Proceedings of the National Academy of Sciences, 110(35), 14196-14201.

Rogers & McClelland, 2004. Précis of Semantic Cognition: A Parallel Distributed Processing Approach. Behavioral and brain sciences 31, 689–749.

Schmitz, A. (2012). A Primer on Communication Studies.(2012).

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. Proceedings of the National Academy of Sciences, 106(11), 4567-4570.

3.2 R Packages

aida: R package with helper functions for courses on Advanced topics in and Intro to Data Analysis. R package version 0.2.1.

bayesplot: Gabry J, Mahr T (2019). “bayesplot: Plotting for Bayesian Models.” R package version 1.7.1, https://mc-stan.org/bayesplot

brms: Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01 & Paul-Christian Bürkner (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal, 10(1), 395-411. doi:10.32614/RJ-2018-017

HDInterval: Mike Meredith and John Kruschke (2018). HDInterval: Highest (Posterior) Density Intervals. R package version 0.2.0. https://CRAN.R-project.org/package=HDInterval

loo: Vehtari A, Gabry J, Magnusson M, Yao Y, Gelman A (2019). “loo: Efficient leave-one-out ross-validation and WAIC for Bayesian models.” R package version 2.2.0, https://mc-stan.org/loo

Rstudio: R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

tidybayes: Tidy Data and Geoms for Bayesian Models_. doi: 10.5281/zenodo.1308151 (URL: https://doi.org/10.5281/zenodo.1308151), R package version 2.0.3, http://mjskay.github.io/tidybayes/

tidyverse: Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

3.3 Software

Audacity: Audacity® software is copyright © 1999-2020 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.

_magpie: https://magpie-ea.github.io/magpie-site/

3.4 Web Resources

All the images that were we used stem from: https://www.pexels.com/

Netlify: An all-in-one platform for automating modern web projects. https://www.netlify.com/

Sounds (all sounds were normalized and cut to 1s in length):

CC: "ringtone.wav" by davidferoli
CC: "5_cello_E2.wav" by flcellogrl
CC: "Aggressive Guard Dogs" by Oneirophile
CC: Bird Whistling, Robin, Single, 13.wav by InspectorJ

Synthetic voices: https://www.text2voice.org/+