Deconstructing the Label Advantage Effect

Teun van Gils (Teun.vanGils@mpi.nl)
Neurobiology of Language Department

Guillermo Montero-Melis (Guillermo.MonteroMelis@mpi.nl)
Neurobiology of Language Department & University of Stockholm

Peter Hagoort (Peter.Hagoort@mpi.nl)
Neurobiology of Language Department

Markus Ostarek (Markus.Ostarek@mpi.nl)
Neurobiology of Language Department, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands

Abstract

Is language unique in how it evokes conceptual content, and if so, why? In an influential study, Lupyan & Thompson-Schill (2012) report that labels (nouns like "dog") have a stronger cuing effect in picture verification tasks than "equally informative and predictive nonverbal cues", like the sound of a barking dog. Here we sought to better understand the factors that lead to a label advantage. First, while we replicate the label advantage itself, our data do not support the assumption that labels and environmental sounds are equally informative. Instead, we show that different cue types are associated with target images to different degrees, and that labels show the strongest association. Moreover, the degree of association is a better predictor of reaction times than cue type. Thus, we conclude that labels are not *more effective* at activating the same semantic content than non-verbal cues, but rather activate *different* semantic content.

Keywords: linguistics; neuroscience; psychological science; language and thought; perception; label-feedback hypothesis

Introduction

Language breaks down our world into a heterogeneous universe of symbols, of discrete labelled categories, allowing us to communicate complex concepts and ideas with relative ease. Each of these symbols (e.g. "dog") conveys a concept, but this conceptual content can also be activated through our senses (the sound of a barking dog), and can be used in non-linguistic tasks (e.g. flight or fight). This suggests that the underlying semantic system is shared, and that language accesses a conceptual store that can equally be accessed through other means.

However, some accounts suggest that language is unique in the way it evokes conceptual content and interacts with perceptual processing. This potentially implies some dedicated cognitive architecture, similar to what Paivio (1991) proposes in his dual-coding theory, which argues that language is processed differently from other sensory input. According to the label-feedback hypothesis (Lupyan, 2012), language can affect lower-level perceptual processing through a top-down process, allowing linguistic information to reinforce or interfere with bottom-up categorization processes. Importantly, the label-feedback hypothesis envisions a special role for labels, as they are the proposed drivers of this feedback effect. Numerous studies support this type of low-level effect of language on perception (Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013; Ostarek & Huettig, 2017; Lupyan & Spivey, 2010; Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2015; Forder & Lupyan, 2019). These effects cannot be explained entirely through high-level cognitive effects of language, since they affected phenomena that occur too early in processing. They can also not be explained through structural changes, since the effects could temporarily be activated or suppressed by the presentation of a label – hence providing support for the label-feedback hypothesis. In determining the processing architecture of the brain, it is relevant whether labels are unique in this regard, as this would imply dedicated processing of labels.

In their seminal study, Lupyan & Thompson-Schill (2012) presented participants with a cue (e.g., the noun "dog" or the sound of a barking dog) followed by an image that could either match the cue (a picture of a dog) or not (that of a train). In a series of experiments, they showed that noun cues ("dog") led to faster reaction times than sound cues (barking sound), even though the sounds were unambiguously related to the images. A central claim in their study is that labels lead to an advantage over environmental sounds even though both were "equally informative and predictive" of the target category (Lupyan & Thompson-Schill, 2012), referring to this as the *label advantage effect*. To investigate whether this effect was due to the nature of the cue (word versus non-word) or due to its semantic content (label versus non-label), Lupyan & Thompson-Schill (2012) conducted an additional experiment that included different types of cues: nouns and verbs (e.g. "dog" and "barking"), imitated sounds (e.g. "arf-arf"), and environmental sounds (the sound of a barking dog). Generally, the noun cue retained an advantage and proved to be more effective than any other cue, suggesting that the effect was caused by the fact that it was a label for the object category, rather than the word-like nature of the cue.

The strongest claims regarding the label advantage effect point towards a special cognitive treatment of labels, arguing that labels are not better cues for the target images, but that they affect their processing more strongly because labels are "special" and processed differently from other conceptual cues. This reasoning relies on the assumption that verbal and non-verbal cues were equally informative, as they showed in an image concordance rating task conducted as a part of their study (Lupyan & Thompson-Schill, 2012). This is crucial, because it would not be surprising to show that cues which are rated to be better fits for a set of images, are better cues in a picture verification task using those images. Nonetheless, this less surprising possibility might be the more likely one: labels differ from environmental or other non-verbal cues in presentation, usage, and acquisition. Thus, they are likely to activate different instantiations of the same conceptual categories. Here, we test the notion that the strength of cuing effects is not determined by whether cues are labels or not, but by how (well) cues trigger a concept that matches the target picture.

Interestingly, in their study, Lupyan & Thompson-Schill (2012) also found verbs to lead to slower reaction times than environmental sounds in matching conditions, an unexpected result that was left mostly unexplained. Verbs are also labels, although perhaps not for the object in question, but for associated actions. Additionally, sound cues (sound of a dog barking) more accurately reflect the verb ("barking") than the noun ("dog"), since sounds are often related to actions, and only refer to the object or animal in question indirectly. Hence, the presented verbs and sounds might be more conceptually similar to each other, but less similar than the noun and a generic target image. This argument would entail that cues are not actually equally informative.

This account that the semantic content of cues can help explain a processing advantage is in line with research by Edmiston & Lupyan (2015), which suggested that the label advantage could be modulated by having cues and targets match more or less closely. The authors proposed that environmental sounds convey specific information about particular exemplars whereas labels activate diagnostic information about the given category as a whole. For example, the pitch of a barking dog provides information about its size in addition to information about the source being a dog. In a picture verification paradigm, this might lead us to have more specific expectations about a target image, and consequently to shorter reaction times when these specific expectations are confirmed.

In a series of experiments, we test whether the label advantage reported in Lupyan & Thompson-Schill (2012) is compatible with a more integrated account of the language system and the conceptual system, where language accesses and modifies a shared conceptual store without the need for a fundamental conceptual processing difference. The original experiment by Lupyan & Thompson-Schill (2012) only tested a relatively small number of stimuli (10) and participants (14). Here, we revisit the assumption that verbal and non-verbal cues are equally informative, with increased power, and find it is not supported by our results. We then replicate the original label advantage, but find that it is better explained by the concordance scores between cues and targets, than by the nature of the cue itself (label vs non-label). In our last experiment, we manipulate the visual targets to make them more concordant with the environmental sounds. While we do not reverse the label advantage in favour of environmental cues, we find an interaction effect similar to that reported in Edmiston & Lupyan (2015). Importantly, we find concordance scores to again be the best predictor of reaction times.

Are labels and non-labels equally informative?

In the original research by Lupyan & Thompson-Schill (2012), a norming study was conducted in order to quantify which

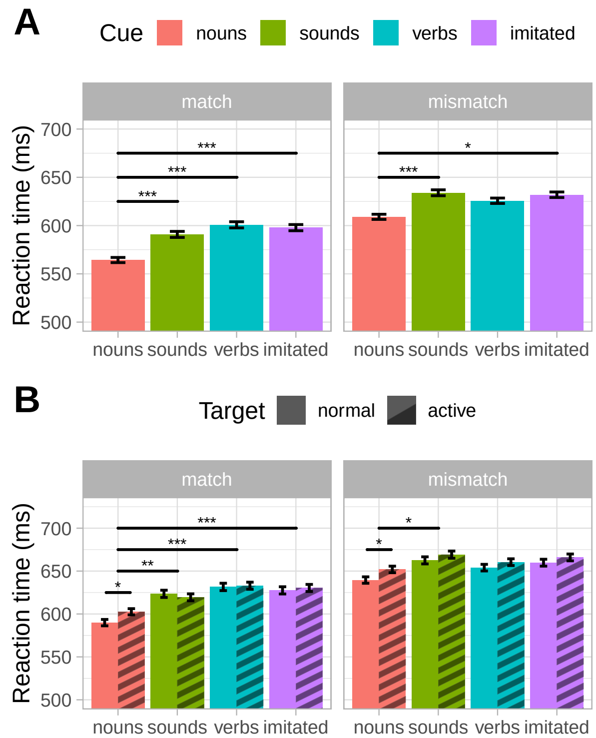

Figure 1: Nouns (red) lead to faster reaction times than any other cue type. (A) Exp. 1: similar to the original Lupyan & Thompson-Schill (2012) study, a noun advantage is observed regardless of whether cue and target match, and responses in the mismatch condition are significantly slower. Contrary to the original study, no verb disadvantage is observed. (B) Exp. 2: the noun advantage effect was reduced in the matching condition when targets were selected to more closely match the semantic content of the non-verbal sound cues (active), as compared to more neutral targets (normal). Bars represent the mean reaction time and the error bars represent the standard error of the mean, both calculated in logarithmic space.

sounds uniquely referred to a specific object or animal, and how well cues referred to these. As the original results supported the claim that the different cue types were equally informative, we conducted an online replication of this study. We followed the design of the original norming study, but with stimuli translated to Dutch, and with additional items and more participants. We obtained image concordance scores for all cues, as differences in these scores could contribute to an explanation of the label advantage effect.

Method

Participants 40 participants (ages 19-44, 32 females) were recruited from the MPI subject database for an online experiment.

Materials We tested 23 items in total and, per item, included one drawing, one active and one normal photograph, one sound, and three voice recordings for the noun, verb and sound imitation conditions. The active photographs were chosen to correspond more closely to the verb and sound cues, e.g. an image of a barking dog to match "barking" and "arf-arf". Stimuli were selected to be suited for all subsequent experiments of this study.

Procedure Participants did the experiment in the Frinex online environment (Withers, 2016), from their own computer. First, the sounds were presented to participants in a randomized order, and they were asked (in Dutch) to respond what animal or object typically produces this sound. We requested singular single-word noun responses, and limited the text input to deny submission of multiple words or common verb or plural word endings.

Subsequently, after hearing/seeing a cue, participants were requested to imagine the animal or object belonging to that cue. They were then shown either a normal or active target photograph, and asked (in Dutch) to rate how well the photograph matched the image they pictured, using a 5-point Likert scale. We asked this question separately for the sound cue, for the spoken noun, verb and sound imitation cues, as well as the drawing cue (so 5 times per photo, or 10 per item). We followed the imagery concordance task procedure used for norming by Lupyan & Thompson-Schill (2012) and described in their appendix. All cue/target combinations were presented in three randomized blocks with short breaks in between.

Analysis Any stimuli where the label accuracy in response to presented sounds was lower than 50% were dropped from the set of stimuli, and stimuli where the cue/target image concordance ratings resulted in an average Likert score significantly below 3.0 (using a Bonferroni-corrected t-test with $\alpha = 0.05$) were also dropped. All other image concordance scores were entered as factors into the corresponding analyses in the further results of the current study.

Results

Based on the collected labels for sounds and concordance scores between cues and images, we selected 16 items that were easily nameable ($> 50\%$ correct labelling), and had sounds, imitated sounds, nouns and verbs that were rated as highly congruent (rating not significantly lower than 3 on a 5-point scale). These items, categorizable into tools and animals, were (with mean and standard deviation, calculated across cue and target types): bell (3.9 ± 1.4), bird (4.1 ± 1.0), cat (4.3 ± 0.9), chicken (4.0 ± 1.1), clock (4.0 ± 1.1), cow (4.2 ± 1.0), dog (3.7 ± 1.2), duck (4.3 ± 0.9), frog (4.1 ± 1.1), gun (4.1 ± 1.0), hammer (4.0 ± 1.2), motorcycle (3.8 ± 1.2), rooster (4.2 ± 1.0), sheep (4.2 ± 0.9), snake (4.0 ± 1.1), and train (3.8 ± 1.1). Half of these items (8) matched the ones used in the original study (Lupyan & Thompson-Schill, 2012), and we find no major differences between the ratings of the original (3.95 ± 1.17) and the additional items (3.97 ± 1.16), indicating that our newly introduced items are not significantly different from the original materials ($p = 0.26$, two-sided t-test).

Comparing the image concordance scores between cue and target types (data not shown), we observe significant differences between both different cue types ($F(4,156) = 35.6$, $p < 0.001$) and different target types ($F(1,39) = 6.64$, $p < 0.05$), as well as an interaction between the two ($F(4,156) = 41.0$, $p < 0.001$). For normal targets, nouns result in the highest image concordance scores (4.5 ± 0.8), followed by sounds (3.9 ± 1.2), and both other cues led to significantly lower scores (imitated: 3.6 ± 1.2, verbs: 3.4 ± 1.2). For active targets, sounds result in a higher concordance score than any other cue type (4.24 ± 1.0), closely followed by nouns (4.16 ± 1.0), and both other cues led to significantly lower scores (imitated: 3.8 ± 1.2, verbs: 3.8 ± 1.2).

In sum, this shows that active target images match more closely with sounds than with nouns, and hence confirms our suspicion that the original materials used target images that were biased in favour of noun cues. More generally, these results show that there are significant differences between cue types, as well as that there is an interaction effect between cue and target type on the observed concordance scores. Both of these factors contradict the claim that the different cue types are equally informative, and may hence play a role in explaining the label advantage effect.

Experiment 1

We replicate experiment 2 from Lupyan & Thompson-Schill (2012) to confirm their findings and to analyse whether concordance scores contribute to their results. In the original study, the motivation for this experiment was to disentangle a label (noun) advantage from a word or speech advantage by introducing non-label word cues (verbs) and non-word speech cues (imitated sounds). The results from their experiment suggest that there is a reaction time advantage for noun cues over the other cues, while there is a reaction time disadvantage for verb cues. The verb disadvantage was only present in matching trials (where cue and target match), while the noun advantage persisted in both matching and mismatching trials. Sound and imitated sound cues performed comparably in both cases, and led to response times in between the noun and verb cues. We will investigate whether concordance scores can account for the observed differences in reaction times, and whether they can explain the presence of a verb disadvantage effect.

Method

Participants 63 participants (ages 18-49, 50 females) were recruited from the MPI subject database for an online experiment, based on a power analysis. One participant was excluded from further analysis because fewer than 75% of trials were usable.

Materials We created stimuli similar to those used by Lupyan & Thompson-Schill (2012), following the same selection criteria, and after norming selected 16 items that met the requirements (see norming experiment).

Auditory stimuli were volume normalized, but unlike the

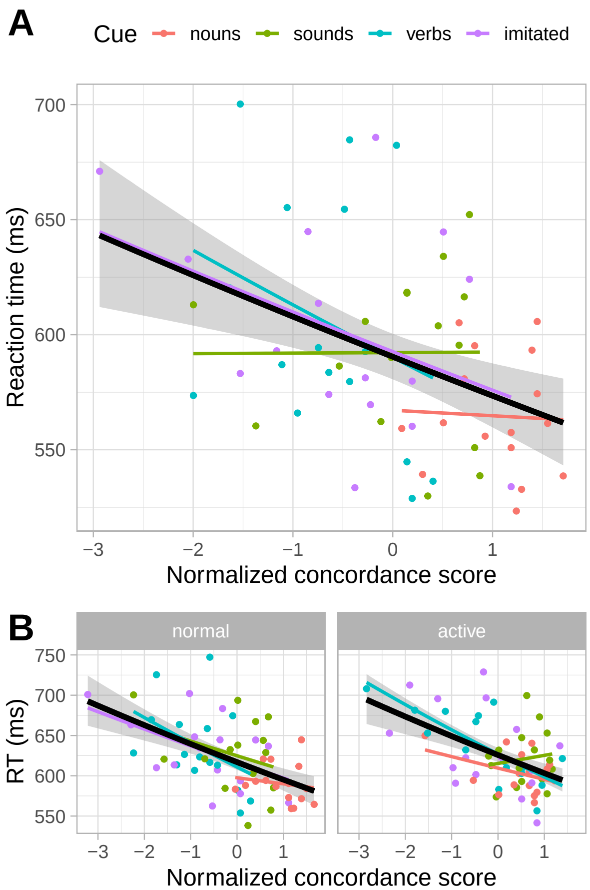

Figure 2: Concordance scores explain reaction times. (A) Exp. 1: reaction time was negatively correlated with cue-target concordance scores, taken from our norming study. The ranges of concordance scores spanned by each cue type are limited, as they are not independent from to each other, with labels notably more reliably being concordant with target images. (B) Exp. 2: this negative correlation remained regardless of the semantic closeness between non-verbal cues and target images, indicating that our concordance scores accurately reflect these differences. A notable exception are the sound cues in trials with active target images, where concordance score seems to be positively correlated with reaction time. Dots represent different cue-target combinations, coloured by cue type and plotted by their normalised mean concordance score and reaction time, the latter calculated in logarithmic space. Lines represent log-linear fits on specific cue types (coloured) or on all data (black), with the shaded area representing the standard error of the mean. Only data from correct trials on matching conditions is visualised.

original study, we did not match the lengths of the auditory stimuli, because this produced unnaturally sounding audio which could affect results. Despite the fact that cue timing has been shown to be important (Edmiston & Lupyan, 2015) when the delay was very short or even absent, results from Lupyan & Thompson-Schill (2012) (experiments 1A-C) where the cue-target delay was varied suggest that the cue is processed by the time the target is displayed. Additionally, we included cue length as a regressor in all our models, to prevent any effect of cue length on reaction times from confounding our results.

Procedure Participants did the experiment in the Frinex online environment (Withers, 2016), from their own computer. Each trial followed the following sequence: first a cue was presented, followed by a 1s pause, and then a target image was presented that either matched or did not match the presented cue. Cues were all auditory, and could be either a noun, verb, imitated sound, or environmental sound. For non-matching trials, an image from the other objects was randomly selected. Participants were instructed to respond to targets as quickly as possible by indicating whether the displayed target matched the presented cue. Feedback was presented on-screen for 500ms after each trial, indicating whether a correct response was given, followed by a 500ms pause before starting the next trial. Each participant completed 512 trials: 16 items x 4 cues x 1 target x 2 levels of congruence x 4 repetitions. The experiment started with 16 practice trials (using 2 items that were not used in the actual experiment). Trials were presented in four randomized blocks with short breaks in between. In total, the experiment took approximately 40 minutes, including instructions, practice trials, and breaks.

Analysis We analyse reaction times of correct trials using a log-normal mixed-effects model. We use stimuli and participants as random effects, and we consider cue, matching, and all their interactions as fixed effects and slopes for our random effects. In further analyses, we additionally consider norming concordance scores, gender, age, cue length and stimulus order as fixed effects. We used orthogonal contrasts, and all numeric variables were normalized prior to fitting the models. The Welch–Satterthwaite method was used to approximate the effective degrees of freedom, in order to calculate significance.

Results

We find a significant noun advantage ($p < 0.001$), both in matching and mismatching conditions (Figure 1A), replicating the original results by Lupyan & Thompson-Schill (2012). While verbs and imitated sound cues also led to slower responses than noun cues (Figure 1A, $p < 0.001$), the verb disadvantage was not replicated in either the matching or mismatching condition ($p = 0.36$ and $p = 0.63$, respectively). Given our much higher sample size and power, we assume that our failure to replicate this verb disadvantage, which was the largest effect in the original study (Lupyan & Thompson-Schill, 2012), indicates that it was a false positive in the original results.

When concordance scores are included in the model, higher scores significantly predicted faster reaction times (Figure 2A, $p < 0.001$). This was even the case when concordance scores were normalized per cue type, to correct for the correlation between concordance scores and cue type (data not shown,

$p < 0.001$). Critically, inclusion of concordance scores as a covariate eliminated the noun advantage over verbs and imitated sound cues ($p > 0.05$). This shows that the degree to which cue and target are deemed to be congruent can explain the changes in reaction time (Figure 2A). Here, sounds and nouns seem to display a weaker negative correlation between reaction time and concordance score, which may be explained by the limited range of concordance scores within each category, combined with a high variability in reaction times for sound cues. These ranges are limited because concordance scores and cue types are not independent from each other, with labels more reliably being concordant with target images. We conclude that concordance scores can at least partially explain differences in reaction time, but since concordance scores and cue types are correlated, no strong conclusions can be drawn regarding the noun advantage effect itself.

In order to address this point, we perform an indirect model comparison between three models: one with cue type, one with concordance scores, and one with both as fixed effects. Because the first two models are not hierarchical with respect to each other, we compare both single-effect models with the combined model in order to test the hypothesis that concordance scores alone can explain the results. We do this by estimating the Bayes factor from two linear mixed-effects model, assuming a uniform prior, using the Bayes Information Criterion (BIC). Here, we find that the combined model is strongly preferred over the cue-type model with a Bayes factor of over 10^{18}, and that the concordance-only model is strongly preferred over the combined model with a Bayes factor of over 10^{8}. Thus, we conclude that the noun advantage can be explained by just the concordance scores, indicating that it is determined by how well a cue matches the target image, not by whether the cue is a label.

Experiment 2

The initial Lupyan & Thompson-Schill (2012) study suggested that the label advantage effect was related to the label (noun) itself, not just the fact that it was a word. However, verbs are also labels, although perhaps not for the object in question, but for associated actions. Additionally, our concordance scores show that there are differences in semantic content between the different cue types. To further tease apart the effect of semantic content and cue type, we adapt a paradigm used by Edmiston & Lupyan (2015), where different target images were used that matched the cue more or less closely (e.g. [barking dog] versus just [dog]).

Method

Participants 91 participants (ages 18-44, 72 females) were recruited from the MPI subject database for an online experiment, based on a power analysis.

Materials Materials matched those used in experiment 1, with the exception of an additional target photograph. These photographs were "active", in the sense that they more closely resembled the action associated with the verbs and sounds, e.g. displaying a barking dog.

Procedure The procedure of this experiment was identical to that of experiment 1. Each participant completed 512 trials: 16 items x 4 cues x 2 targets x 2 levels of congruence x 2 repetitions. The experiment started with 32 practice trials (using 2 items that were not used in the actual experiment). In total, the experiment took approximately 40 minutes, including instructions, practice trials, and breaks.

Analysis The analysis was identical to experiment 1, except we added target (active/normal) and all its interactions with the other predictors as fixed effects. Slopes for our random effects included these same factors, but without their interactions, due to computational restrictions.

Results

Reinforcing results from experiment 1, we find a significant noun advantage ($p < 0.001$), both in matching and mismatching conditions, as well as between normal and active target images (Figure 1B). Verb and imitated sound cues again led to slower responses than noun cues (Figure 1B, $p < 0.001$), and a significant effect of matching condition ($p < 0.001$) was found.

Active target images, which more closely matched sound cues (as shown by our norming results), led to slightly slower reaction times overall ($p < 0.05$). Contrary to our expectations, we did not find a complete inversion of the label advantage effect. However, there was an interaction between target type and the size of the noun advantage, due to slower reaction times in response to noun cues but faster reaction times in response to sound cues for active images ($p < 0.05$, Figure 1B). This is in line with our expectations, and reaffirms the hypothesis that the original target images favoured the noun cues, albeit not to the degree as we had expected.

When concordance scores are included in the model, the disadvantage of verbs and imitated sounds with respect to nouns disappears, along with cue-target interaction effects (all $p > 0.05$, Figure 3). This shows that concordance scores at least partly explain the same variance as cue type. The semantic congruence between cues and targets consistently led to faster reaction times (Figure 2B), for both target image conditions, showing that our concordance scores accurately capture the information that determines task performance. Note that both here and in our statistical model (Figure 3), sound cues seem to be outliers, and a positive correlation even seems to be present for active target images following sound cues (Figure 2B, right).

To determine whether concordance scores alone can adequately explain these results, we again perform an indirect model comparison between three models, as in Experiment 1: one with cue type, one with concordance scores, and one with both as fixed effects. Again, we find that the combined model is strongly preferred over the cue-type model with a Bayes factor of over 10^{41}, and that the concordance-only model is strongly preferred over the combined model with a Bayes

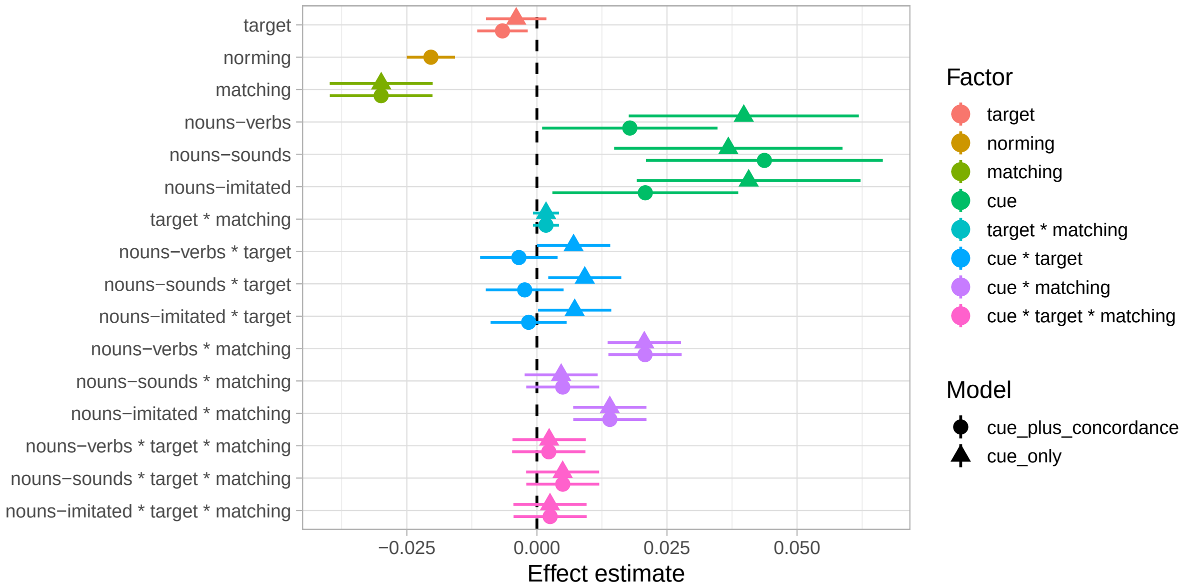

Figure 3: Concordance scores can partially explain the effect of cue type on reaction time (Exp. 2). Incorporating cue-target concordance scores (yellow), taken from our norming study, into our model leads to reduced noun-verb and noun-imitated effects (teal), and eliminates the cue-target interaction effect (blue). The noun advantage effect (nouns-sounds) itself is unaffected. Shapes represent the mixed-effects model estimates of the specified contrasts in log-space, and bars the 95% confidence intervals.

factor of over 10^3, confirming our hypothesis.

Discussion

This study suggests that the label advantage effect can be deconstructed into differences in semantics, and that labels are only special insofar as they typically differ from other cues with respect to their semantic content. Nouns seem to activate information about objects in a neutral (prototypical) state. Thus, when prototypical images are used, ratings show that nouns are more strongly associated with them and they reduce reaction times in picture verification experiments. This matches the general pattern that higher cue-target concordance scores lead to faster reaction times, also for non-labels. When non-prototypical images are used, the label advantage effect is reduced. Thus, our results support the claim that nouns, verbs, and non-verbal cues activate different conceptual information.

We replicated the main result from Lupyan & Thompson-Schill (2012), and consistently find a reaction time advantage for nouns relative to other cue types, but do not find that the effect is there *despite* cues being equally informative. Rather, we propose that it is there because cues are *not* equally informative: contrary to the original Lupyan & Thompson-Schill (2012) study, nouns were rated to match the target images better than non-verbal cues, despite the fact that we followed their design very closely. A subset of our items were taken directly from the original study, and the newly introduced items did not differ significantly from the original ones, making it unlikely that these introduced differences. We also find it unlikely that differences in population or language alone could explain the difference between our results, and thus we assume that the higher power of our norming study may explain this difference.

We find that meaningful differences in semantic information, as measured by our concordance scores, are present both for verbal and non-verbal cues, both within and between cue types. This suggests that this flexibility is a feature of a shared semantic system, rather than an idiosyncratic property of label cues. This is in line with statements in the original study by Lupyan & Thompson-Schill (2012) that describe labels as evoking more prototypical categorical representations of concepts. However, it contradicts statements about cues being equally informative, or words being inherently special.

Importantly, our results are not compatible with the view that nouns refer to the general conceptual category (dog) without specifying features of a particular category member. This is because on that account the effect of labels should be the same regardless of the idiosyncratic features of a given target image. In contrast, our second experiment shows that reaction times in the noun condition were slower when active target images were used (e.g., a barking dog) than more closely matched non-verbal cues. This fits better with a view where nouns activate objects in a particular state in a way that differs from verbs *and* from non-verbal cues. This is in line with experiments conducted by Edmiston & Lupyan (2015); they show that the semantic content of cues, and in particular the degree to which the cue and target match, has an influence on the observed effect.

In experiment 4, Lupyan & Thompson-Schill (2012) show the presence of a label advantage effect in a different paradigm where participants first learnt novel cues (either sounds or labels), so that cue predictiveness and familiarity could be controlled. This experiment suggests that labels differ in the semantic content that they evoke even if the learning regime is the same. As proposed by Lupyan & Thompson-Schill (2012), this may be the case because our prior experience with language biases us to learn new words in a more categorical and abstract manner compared to non-verbal cues.

One limitation of the current concordance-based explanation is that the different cue types only span a limited range of the concordance scores. Of course, this limitation follows directly from the task design, since we selected cues with high concordance ratings to ensure that they were sufficiently associated with their target concepts. Despite this, we showed that the concordance scores have a significant effect even if they are normalized per cue type, effectively removing differences in range. For future studies, it may be interesting to sample cues where each cue type spans a similarly broad range of congruence scores, to consolidate our results.

There were some effects of cue type that the concordance scores did not capture; however, this was reflected in a disadvantage for environmental sound cues, rather than an advantage for nouns. This raises the possibility that the differences in reaction times can be characterized more accurately as a disadvantage for environmental sound cues in conjunction with an advantage for trials with higher semantic cue-target concordance. Further research needs to determine whether the label advantage in this experimental paradigm may be explained through a bias towards representations of a specific modality type, e.g. a bias of nouns towards visual representations (dog) and of verbs towards auditory representations (barking). This difference may be relevant because in the present paradigm the target is always an image, and hence potentially favours visual representations. A fully visual version of the experiment could address this issue, where spoken word cues are replaced by written words and the environmental sound is replaced by e.g. a drawing. Additionally, in future research, it may be informative to look at dynamic stimuli, such that the use of verbs as labels can more extensively be investigated.

In the end, words are not neutral cues, but represent an implicit intention to convey categories, and are used in a way that is fundamentally different from the way in which non-verbal cues are used. Meaning arises from how our conceptual representations form and are used in everyday life, and what their associated prototypes and modalities are, and it is likely that the semantic content of labels reflects these differences. These differences then invoke different representations, leading to differentiated top-down modulation of sensory processing and, thus, a difference in task performance. As such, the label advantage effect can be deconstructed into underlying semantic differences, based on a shared semantic system. This provides a simpler explanation of what makes words special, that is tied to semantic content rather than to cue type directly.

Acknowledgments

This work was partly supported by the Swedish Research Council (grant 2018-00245 to GM-M).

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.

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.

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*(7), 1105–1123.

Hoffman, P. (2016). The meaning of 'life' and other abstract words: Insights from neuropsychology. *Journal of Neuropsychology*, *10*(2), 317–343.

Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, *3*.

Lupyan, G. (2017). The paradox of the universal triangle: Concepts, language, and prototypes. *Quarterly Journal of Experimental Psychology*, *70*(3), 389–412.

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193.

Lupyan, G., & Lewis, M. (2019). *From words-as-mappings to words-as-cues: the role of language in semantic knowledge* (Vol. 34) (No. 10). Routledge.

Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236–2253.

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

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 499–508.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255–287.

Withers, P. (2016). *Frinex: Framework for Interactive Experiments.* Berlin, Germany. (Poster presented at IT-Verantwortlichen-Treffen)