Access to meaning from visual input:

Object and word frequency effects in categorization behavior.

Klara Gregorová[1, *], Jacopo Turini[1, *], Benjamin Gagl[1, 2 ^] & Melissa Le-Hoa Võ[1, ^]

1. Department of Psychology and Sports Sciences, Goethe University,

Frankfurt am Main, Germany

2. Department of Special Education and Rehabilitation, University of Cologne,

Cologne, Germany

* Joint first authors, ^ Joint senior authors

Corresponding author info: Jacopo Turini, Scene Grammar Lab, Institut für Psychologie, PEG Room 5.G105, Theodor-W.-Adorno Platz 6, 60323 Frankfurt/Main, Germany, turini@psych.uni-frankfurt.de, +49 (0)69 798-35310

Data and materials: https://osf.io/d3j9h/files/; Word count: 13,191;

Previous dissemination

Data, results, and interpretation from the current study have been previously shared in the form of preprint on PsyArXiv and ResearchGate; besides, they have been presented in the form of poster during the virtual Vision Science Society meeting of 2021.

Acknowledgments

We want to thank Michelle Greene for generously making her dataset available to us (Greene, 2013) as well as three anonymous reviewers for their valuable and constructive comments on this work. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 222641018 SFB/TRR 135, subproject C7 to MLV and Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project "The Adaptive Mind").

Abstract

Object and word recognition are both cognitive processes that transform visual input into meaning. When reading words, the frequency of their occurrence ("word frequency", WF) strongly modulates access to their meaning, as seen in recognition performance. Does the frequency of objects in our world also affect access to their meaning? With object labels available in real-world image datasets, one can now estimate the frequency of occurrence of objects in scenes ("object frequency", OF). We explored frequency effects in word and object recognition behavior by employing a natural vs. man-made categorization task (Experiment 1) and a matching-mismatching priming task (Experiment 2-3). In Experiment 1, we found a WF effect for both words and objects but no OF effect. In Experiment 2, we replicated the WF effect for both stimulus types during Cross-modal Priming but not during Uni-modal Priming. Moreover, in Cross-modal Priming, we also found an OF effect for both objects and words, but with faster responses when objects occur less frequently in image datasets. We replicated this counterintuitive OF effect in Experiment 3 and suggest that better recognition of rare objects might interact with the structure of object categories: While access to the meaning of objects and words is faster when their meaning often occurs in our language, the homogeneity of object categories seems to also impact object recognition, particularly when semantically processing contextual information. These findings have major implications for studies wanting to include frequency measures into their investigations of access to meaning from visual inputs.

Keywords: word recognition, object recognition, frequency, distinctiveness, priming.

Introduction

Visual recognition is the cognitive process that maps sensory input from the retina onto meaningful representations stored in semantic memory (Clarke et al., 2013; Grill-Spector & Weiner, 2014); this process supports many tasks like action planning, navigation, reading, social interaction, etc. The types of visual input for these tasks, e.g., objects, scenes, written words, or faces, already pose a high level of complexity, so that research in cognitive science has often investigated different types of visual input separately, focusing on the specificities of each domain (Capitani et al., 2003; Downing et al., 2006). Notably, investigations of the ventral visual stream, i.e., the core neural substrate of high-level vision, compared the brain activation in response to these different types of stimuli (for a review, Grill-Spector & Weiner, 2014). Their main finding was that different stimulus types activated distinct but neighboring regions (e.g., fusiform face area; Kanwisher, et al., 1997; the visual word form area, Dehaene & Cohen, 2011). At the same time, other researchers focused on comparing different visual inputs, e.g., objects and words, to understand the process of accessing the same semantic representation, i.e., the identical meaning (Shelton & Caramazza 1999; Shinkareva et al., 2011; Devereux et al., 2013; Fairhall & Caramazza, 2013). We followed this approach and investigated how different types of visual input can access identical meanings. We were particularly interested in the frequency of occurrence in the world, operationalized by both word and object frequency measures.

In the fields of visual word recognition (Balota et al., 2004) and reading (Kliegl et al., 2006; Rayner, 2009), the so-called “word frequency effect” is a well-established finding. The word frequency (WF) effect shows that words that occur more often in our language (e.g., the article "the") are processed faster than words that are rare (e.g., “platypus”). Common naming and lexical-semantic categorization tasks, e.g., lexical decision tasks (Balota et al., 2004; for a

review, Brysbaert et al., 2011; Brysbaert et al., 2018), consistently show WF effects, i.e., longer response times and more errors for low frequency words. Even though there have been various attempts to identify more reliable estimates of WF and its nature, it is generally agreed upon that the WF effect emerges as an effect of learning and exposure to a language (for a review, Brysbaert et al., 2018). Thus, despite different assumptions and implementations, most models of visual word recognition and reading took WF into account as a crucial parameter representing the difficulty in accessing lexical representation in the so-called "mental lexicon" (Forster & Chambers, 1973; Morton, 1979; McClelland & Rumelhart, 1981; Coltheart et al., 2001; Engbert et al., 2005).

Object recognition models (Riesenhuber & Poggio, 2000), on the other hand, are primarily concerned with assigning images to different categories, irrespective of their frequency of occurrence (Morrison et al., 1992; Criss & Malmberg, 2008; Taikh et al. 2015). In the rare cases when studies compared recognition performance of written words and matched object images, typically frequency effects were investigated based on word measures. For example, Taikh and colleagues (2015) found faster object than word recognition performance in a semantic categorization task, but WF only affected word recognition performance (Taikh et al., 2015). When behavioral investigations used naming aloud tasks, object recognition performance also showed frequency effects based on word-based estimates (Bates et al., 2001; Almeida et al. 2007; Taikh et al. 2015). However, the WF effects found in object naming studies are likely related to the process of accessing the verbal output representation (i.e., the spoken word; Almeida et al., 2007). Thus, tasks that involve linguistic representations, e.g., as part of the output modality, might be more sensitive for word frequency effects on object recognition performance.

A potential limitation of previous investigations comparing the two domains (words and objects) is that these studies included only frequency estimations that rely on linguistic input:

i.e., large text corpora (e.g., books and newspapers, like dlexDB, > 20 million words; Heister et al., 2011) or spoken language corpora (e.g., from tv-movie subtitles, like SUBTLEXDE, about 25,4 million words; Brysbaert et al., 2011). Typically, WF estimates represent the number of occurrences per million words. Across languages, WF estimates have a better explanatory power for reaction time data from word recognition tasks when extracted from TV and movie subtitles than from book and newspaper texts (e.g., for German, see Brysbaert et al., 2011; for English, see Brysbaert & New, 2009). This finding likely reflects that participants in psycholinguistics experiments (often young students) are more exposed to popular TV shows and movies than the content of classic text corpora, which often include highly specialized texts. Thus, subtitle-based WF measures are, to date, the best representation of the number of occurrences of words in everyday life (Brysbaert et al., 2011; 2018). However, it is still unclear how these more precise measures estimated from subtitles might also explain recognition performance in the object domain. Furthermore, it is essential to explore if newly developed frequency measures, based on the occurrence of objects in images of real-world scenes, could also be valid estimates of access to meaning or not, and could also shed more light on the phenomenon underlying the WF effect. Thus far, the lack of such object frequency (OF) measures has likely been due to a lack of easy access to fully labeled image databases.

Recent advances in computer vision have made annotated image datasets with segmentations and labels of all objects within a scene readily available. Usually, these labels come from human annotators (Russel et al., 2008). For example, the ADE20K dataset contains over 20,000 real-world images from 900 different scene categories, with hundreds of thousands of object annotations categorized into more than 2,500 object categories (Zhou et al., 2019). Despite having been developed for computer vision research, these datasets allow us to extract quantitative measures about contextual regularities of objects in the environment (e.g., objects that appear more often in a specific scene category). These newly available object-in-scene

statistics have inspired new investigations regarding which aspects of a scene our cognitive system exploits to efficiently process objects and scenes (Greene, 2013; Võ et al., 2019). Notably, we can now efficiently compute an object-based frequency measure based on these image datasets. This OF measure uses the same logic as word-based frequency: counting the number of occurrences of a labeled object in a given image dataset.

It is important to note that current research on WF measures suggests that corpora should include at least 20 million words (Brysbaert et al., 2011) in order to yield a reliable frequency estimate. We cannot expect such a high number of objects for the currently available annotated image datasets, and we should consider that - as is the case even with wellestablished text corpora - every measure computed from a dataset represents only an approximation of real-world properties. In the specific case of real-world image datasets, biases could arise not just from the limited number, but also from limited variety of scene categories, limited points of view of photographs, artificiality of image composition, lack of clutter, etc. Nevertheless, there have been some successful attempts to use measures from existing image datasets to model neural response to object recognition (e.g., from ADE20K; Bonner & Epstein, 2021; Bracci et al., 2021). Thus, in this study, we explore the potential of these newly computed object-based frequency measures on capturing aspects of visual recognition behavior and compare them to well-established word-based frequency measures. To do so, and to limit biases from specific datasets, we employed not only one, but two measures of OF computed from two datasets that differ in size and quality of annotations (Greene, 2013; Zhou et al., 2019), as well as two measures of WF from datasets that differ in the source of the linguistic input (Brysbaert et al., 2011; Heister et al., 2011). The effect of these measures on accessing meaning during visual recognition was assessed in three experiments.

The first experiment used a semantic categorization task in which participants had to decide whether a concept, presented via an object image or via a written word, was natural or

artificial (i.e., man-made). During the procedure, we recorded response times and error rates from participants. The response time data allowed us to investigate whether word-based or object-based frequency measures modulated the speed of semantic access. We expected to replicate the WF effect for words. Besides, we wanted to test whether a WF effect on object recognition would emerge without an explicit linguistic response. Importantly, for the first time we explored possible effects of newly developed OF measures on both object and word recognition behavior.

Observing an OF effect only in object recognition and a WF effect only for words would indicate that recognizing and learning visual stimuli (words vs. objects) occurs separately within each modality (e.g., by means of a verbal vs pictorial representation). Alternatively, if one frequency measure would affect both modalities alike (e.g., WF affecting word and object recognition), this finding would indicate that a frequency measure is not just a proxy for the repeated experience with a modality-specific stimulus (e.g., a word) but for the repeated experience with the semantic representation connected to that stimulus (i.e., its meaning). Therefore, the strength of the semantic representation given by the repeated experience would also be present when that semantic representation is accessed from a different modality (e.g., a picture). This scenario is in line with the idea that semantic representations are shaped by different kinds of experiences: perceptual, motor, affective, but also linguistic. In this view, for example, language is not just a means of representing and communicating conceptual knowledge but has a transformative power on this knowledge as well (Lupyan & Lewis, 2019). These transformations derived from modality-specific experience then generalize to other modalities.

In the second experiment, the same participants completed a priming task in which they had to decide whether the meaning of the prime and target stimuli matched. By implementing either Uni-modal or Cross-modal Priming, we were able to modulate the degree of semantic

processing in the task and examine how frequency effects change as a function of the varying semantic demands. Uni-modal Priming (Scarborough et al., 1977) occurs solely on the perceptual level, as matching prime and target pairs not only have the same meaning but are also identical in their visual appearance (i.e., word primes word or object primes object). Thus, we expected a lower involvement of semantic processing. In contrast, Cross-modal Priming (Tversky, 1969) necessarily requires semantic processing because participants must relate two visually distinct stimuli to one meaning (i.e., object priming word or vice versa) to solve the task. If the effects were most substantial in Cross-modal Priming, this would provide further evidence that the frequency effects reflect an aspect of semantic rather than merely perceptual processing. The same participants of Experiment 1 and 2 also performed a rating study from which we have extracted stimulus-specific measures that we have used as covariates in the analysis.

To avoid potential carry-over effects from Experiment 1 to 2 when testing the same participants, we conducted a third experiment which included two new sets of participants — one performing only the Cross-modal and another performing only the Uni-modal Priming trials. This additionally reduced the number of concept repetitions per person. We again hypothesized that if frequency effects reflect processing of semantic representation rather than only perceptual representation, stronger frequency effects should emerge in the group exposed to Cross-modal Priming rather than Uni-modal Priming. Finally, further sets of ratings were collected from a new group of participants different from the ones of Experiments 1,2, and 3, again with the idea of extracting covariate measures to use during the analysis.

Materials and Methods

Participants

We required all participants taking part in our study to have normal or corrected-to-normal vision, be German native speakers, and have no history of linguistic, psychiatric, or neurological disorders. Additionally, we only included participants who did not report having technical problems during the online procedures and who completed both sessions. Participants were recruited by sharing the link to the studies on through platforms and mailing lists of students at the Goethe University of Frankfurt.

To prevent an overestimation of underpowered correlations, which may be expected when Ns is below 30 participants (e.g., see Yarkoni, 2009), we tested 60 participants (of whom 42 fit the above-mentioned criteria) in Experiments 1 and 2, as well as the rating study judging typicality and familiarity of the used stimuli (age: M = 23.55, SD = 8.88, Range: 15-59 y.; Gender: 34 F, 7 M, one person did not report; 5 bi/multilingual with German as one of the native languages).

For the replication in Experiment 3, we recruited 53 additional participants for the Cross-modal Priming task (age: M = 22.87, SD = 6.83, Range: 18-50 y.; Gender: 43 F, 10 M; 10 bi/multilingual with German as one of the native languages), and yet another 53 participants took part in the Uni-modal Priming task (age: M = 22.66, SD = 4.81, Range: 18-39 y.; Gender: 35 F, 16 M, 2 NB; 13 bi/multilingual with German as one of the native languages). The sample size for the replication (N = 53 + 53 = 106) was obtained by taking the sample size in Experiment 2 (N=42), which had a within-participant design, and adapting it to a betweenparticipants design in the replication, following this formula: $Nbetween = (Nwithin * 2)/ (1 - \rho)$, where 2 represents the number of groups / conditions (in our case: Cross-modal and Uni-modal Priming) and ρ represents the correlation between the two groups / conditions

(in our case, from Exp 2, ρ = 0.208). The formula was then solved for $Nbetween$ = (42 ∗ 2)/ (1 − 0.208) = 106.061 (Maxwell et al., 2017).

Additionally, two distinct groups of participants were recruited to collect further ratings regarding the stimuli used: (1) One group of 20 participants (age: M = 21.65, SD = 2.66, Range: 19-29 y.; Gender: 10 F, 10 M; 4 bilinguals/multilinguals with German as one of the native languages) performed a rating study judging typicality and familiarity of the stimuli. This further set of ratings for typicality and familiarity were collected anew as part of the replication. (2) A second group of 16 participants performed a rating study judging the "Conceptual distinctiveness" (as in Konkle et al., 2010) of the concept used in the studies (age: M = 23, SD = 4.75, Range: 18-33 y.; Gender: 9 F, 7 M).

All participants gave their informed consent and received course credits or monetary compensation for their participation. The Ethics Committee of the Goethe University Frankfurt approved all experimental procedures (approval # 2014-106).

Stimuli

For this study, we selected 100 noun concepts that can be depicted by a single word and an image of an object in isolation. We use the phrase "object concept" here and below, to refer to the semantic representation common to a word denoting an object (e.g., "apple") and the object itself (e.g., a physical apple, an image of it). Half of the concepts could be categorized as natural (e.g., apple) and the other half as man-made (e.g., bicycle). We restricted our search to objects with word labels in the ADE20K dataset, a set of real-world images of scenes with segmented and annotated objects (Zhou et al., 2019). After selection, we translated the English word labels to German. For presentation, we displayed German nouns with an uppercase initial-letter (i.e., correct spelling in German) and in white Arial font on a grey background (hexadecimal color

#424242; jsPsych, de Leeuw, 2015). We downloaded the object images from internet databases (e.g., https://pnghunter.com/, http://pngimg.com/, https://www.cleanpng.com/). They were pasted on a white background, grey-scaled, and resized to 392 x 392 pixels.

Object and word characteristics

For all concepts, we computed four selected frequency measures (two word-based and two object-based). In addition, we computed several stimulus characteristics identified to influence recognition behavior (i.e., to consider as covariates in the statistical analysis).

Object-based frequency measures (object frequency - OF). OF measures represent the logtransformed (base 10) number of occurrences of an object in a dataset of segmented and labeled scene images (e.g., cars on the street). Implementing the log-transformation for frequency measures reduces the skewness of the frequency distribution as typically only few objects have high frequency, while majority of objects have a low frequency (Zipf's-law-like distribution; Greene, 2013). We determined the OF based on two datasets. One used more than 20,000 scene images (from 900 categories), and objects (more than 400,000 instances grouped in more than 2,500 categories) were segmented and labeled by a single expert worker and used to train an image recognition algorithm to identify objects in scenes (*ADE20K OF*; Zhou et al., 2019). Since we based our stimulus selection on objects present in the ADE20K dataset, we tried to represent all the different levels of frequency we could find there (i.e., from few appearances to tens of thousands of appearances). The second dataset used 3,499 scene images (from 16 categories; indoors, outdoors, natural, artificial), labeled by four different workers and carefully cleaned of misspellings, synonyms, and other errors, to measure statistical regularities of objects in a scene (more than 48,000 instances grouped in more than 800 object categories; *Greene OF*; Greene, 2013). Only 78 of our 100 object labels selected from ADE20K

were present in the Greene dataset. When an object was missing in the Greene dataset, we assigned an OF value of 1 count (i.e., 0 log10-counts). Density distribution of ADE20K OF and Greene OF for the set of stimuli can be found in *Supplementary Materials 1*.

Word-based frequency measures (word frequency - WF). WF measures are based on the number of occurrences of a word in a corpus of linguistic materials. Specifically, as for object frequency, the numeric parameter was computed as the logarithm (base 10) of the number of occurrences per million words in a dataset (to turn the Zipf's-law-like distribution into a normal distribution, Li, 1992). When a word was not included in a corpus, which was the case for one concept, the WF was set to 1 count per million (i.e., 0 log10-counts per million). The WF was determined based on two corpora, one using German subtitles from films and tv-shows, *SUBTLEX-DE WF* (Brysbaert et al., 2011) and the other including a large set of German written material, such as books and newspapers, *dlexDB WF* (Heister et al., 2011). The density distributions of SUBTLEX WF and dlexDB WF for the set of stimuli can be found in *Supplementary Materials 1*.

Covariates. In order to estimate and control for the contribution of other variables, we collected subjective ratings from participants, as well as we computed object- and word-specific visual predictors.

Ratings: As part of the replication, we obtained two sets of concept familiarity and image typicality ratings: one from the participants who have taken part in the original study (i.e., Experiments 1 and 2) and one from a different group of participants who had not previously taken part in any of the experiments. We measured *concept familiarity* as the subjective familiarity with an object concept to serve as a subjective counterpart of the objective frequency measures of words and objects computed from a text or image dataset (see Kuperman & van

Dyke, 2013). *Typicality*, on the other hand, represents how an object exemplar is typical of its category. In the original study, individual ratings for each concept and each participant were used to model each participant's performance on each concept in the main tasks. In contrast, in the replication experiment, ratings were averaged across participants and used to model performance on each concept since participants of the rating study differed from those of the original study.

To substantiate the interpretation of some of our results, it became important to further investigate the relationship between OF and *Conceptual Distinctiveness* (Konkle et al., 2010). For this purpose, we set up yet another rating study in which we collected ratings regarding our stimuli's Conceptual Distinctiveness from participants who had not taken part in any of the previous experiments. The rating study followed the methodology described in Konkle et al. (2010). They defined a concept as having a high Conceptula Distinctiveness if it is relatively easy to make subdivisions among the category members it denotes and where these subdivisions are not simply based on perceptual features (e.g., color or shape). Conceptual Distinctiveness ratings were obtained for every concept by averaging ratings across participants.

Visual and visuo-orthographic predictors: In addition to the subjective ratings, we computed and included various object- and word-specific measures from which we extracted visual and visuo-orthographic predictors using a Principal Component Analysis (PCA).

To assess the *visual characteristics of object images*, we computed several measures based on pixel-level input: *Entropy* (Shannon, 1948), which measures the level of "disorder" and visual variance of an image (entropy equals zero means no variance); *Signal-to-noise ratio (SNR) of pixel values* (computed as mean of all pixel values divided by the standard deviation of all pixel values), which we used as a proxy of how the content of the image differs from the background (larger negative values indicate that the content is closer to the background, values

close to zero indicates that the content is more different than the background); *graphic-based visual saliency* (Harel et al. 2007), which measures saliency of the image based on bottom-up features (every pixel has a value between 0 and 1, where zero indicates not salient and a value of one indicates high saliency); *GIST descriptor* (Oliva & Torralba, 2001), which gives us the orientation and spatial frequency in different parts of the image; finally, *Deep Convolutional Neural Network activation from convolutional layer 1, 4 and fully-connected layer 7 of the AlexNet model* (Krizhevsky et al., 2012); they represent low-level (layer 1), mid-level (layer 4) and high-level (layer 7) visual features of our images, as processed by a deep learning algorithm trained to perform human-like object categorization. From a PCA on these visual predictors, we extracted 3 othogonal principal components (PCs) that we named *Image visual PC1*, *Image visual PC2* and *Image visual PC3* (for more info on their impact and interpretations, see *Supplementary Materials 1*)

To assess the *visual and orthographic characteristics of words*, we performed another PCA. For this, we considered two visual properties, *entropy* and *SNR*, computed as described above for object images but now applied to the images of written words. In addition, we computed two orthographic measures, *word length* (i.e., the number of letters) and *distance from orthographic neighbors* (i.e., Orthographic Levenshtein Distance, Yarkoni, et al., 2008). One PC was selected from this process and was labeled *Visuo-orthographic PC*. Correlations between all predictors and between PCs and original measures, as well as PCA loadings, can be found in *Supplementary Materials 1*.

Apparatus

Participants performed the experiments online, hosted on a web server at the Goethe University Frankfurt. We used jsPsych (de Leeuw, 2015) for stimulus presentation and response recording. Participants were instructed to ensure that they started the experiments only when seated in a

quiet environment without potential interruptions and when they had enough time to dedicate to it. Besides, they were instructed to perform the experiment only on laptops or desktop computers. To account for differences in screen size and resolution, we implemented an adaptation mechanism based on the measurement of a credit card (https://www.jspsych.org/plugins/jspsych-resize/).

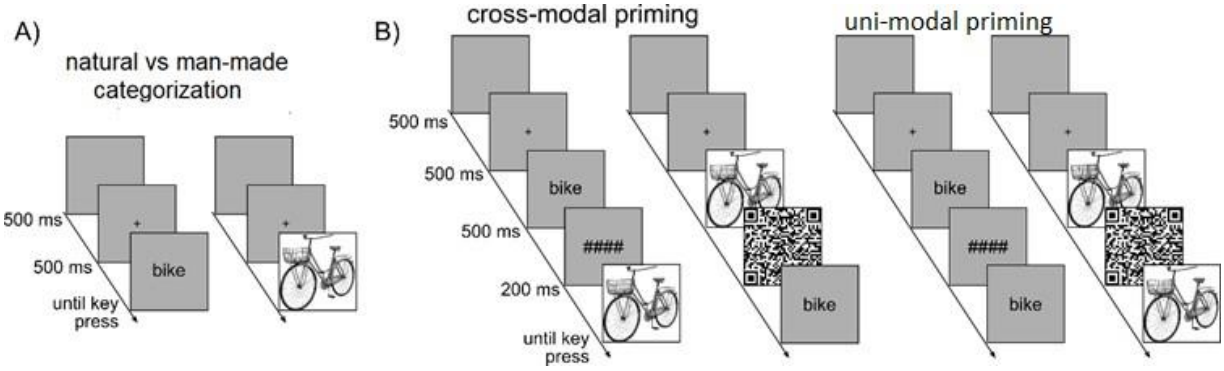
Before the experiments started, participants had to adapt a rectangle presented in the center of the screen to the size of a credit card. This information was used to ensure that the size of stimuli on screen was the same for every participant (object images: 6.7 x 6.7 cm; words, uppercase letter: circa 0.7 cm). In all parts of the experiment, the screen background was grey (hexadecimal color #424242). The Conceptual Distinctiveness rating experiment was programmed in Python using PsychoPy (version 2020.2, Builder GUI; Peirce et al., 2019) and administered online through the hosting platform Pavlovia (https://pavlovia.org/). Stimulus words were presented in black Arial text of 1.5 cm vertical size on white background.

Procedure

Experiment 1. *Figure 1A* shows an example of the natural (e.g., apple) vs. man-made (e.g., bicycle) categorization task of Experiment 1. The two stimulus modalities were presented in two separated blocks (100 stimuli each). Block order was randomized across participants, and within each block, the stimulus order was randomized for each participant. The stimulus presentation sequence started with a fixation cross at the screen center (500 ms) followed by the presentation of an object image/word. After the participants responded, the presentation was terminated. We asked participants to press a key as quickly and as accurately as possible: *j* when a "natural" stimulus was presented and *f* when a "man-made" stimulus was presented. A blank screen was presented for 500 ms between two trials (*Figure 1A*). After each block, we asked the participants to take a break.

***Figure 1*. Experimental design**

A) Experiment 1. Categorization of natural vs. man-made object images and words. B) Experiment 2. Categorization of prime-target matches vs. mismatches. Cross-modal Priming: words are primed with objects, and objects are primed with words. Uni-modal Priming: words are primed with words and objects with objects.



Experiment 2. In the second experiment, we implemented a priming task that included Uni-modal and Cross-modal prime-target pairs, consisting of object images and words. Participants evaluated if both the prime and the target had the same meaning or not. They started with two Cross-modal Priming blocks (i.e., word-priming-object, object-priming-word; see *Figure 1B*). After that, participants completed two Uni-modal Priming blocks (word-priming-word, object priming-object). Within Cross-modal and Uni-modal blocks, we randomized block order across participants. We presented all 100 object concepts twice as a target within each block (200 trials) in a randomized order. Every target was once paired with a matching and once with a mismatching prime stimulus. Mismatching pairs were randomly generated and kept constant for all blocks of each participant. We instructed the participants to evaluate whether the target and prime concepts matched or mismatched. Again, they should indicate this by pressing a key (*j* for match and *f* for mismatch) as quickly and as accurately as possible. Like in Experiment 1, trials started with a fixation cross presented in the screen center for 500 ms. After that, the prime was presented for 500 ms followed by a backward mask for 200 ms ("############" for words or QRcode-like for objects; see *Figure 1B*). The presentation of the target was

terminated by the response of the participant. Again, we asked participants to take a break in between blocks and one break halfway through every block.

Typicality and familiarity ratings. Finally, we asked participants to perform an additional session the following day to collect demographic data and stimulus ratings. This procedure was again performed online. Participants rated all stimuli on a one to six Likert scale. We assessed *concept familiarity* by presenting the concept as a written word in the screen center. In addition, we presented the question *"How familiar are you with the object that the word represents, in your everyday life?*" plus the Likert scale. *Image typicality* was assessed, presenting the object picture in the center of the screen, and the object word on top. In addition, we presented the question, "*How typical is this image in relation to the category designated by the word?*" with the Likert scale.

In total, data collection lasted for about 75 minutes on day 1 (Experiment 1 and Experiment 2) and about 30 minutes on day 2 (ratings).

Experiment 3. Experiment 3 was run to replicate the findings of Experiment 2. It therefore has the same structure of Experiment 2, except that two separate groups of new participants either performed only the two Cross-modal Priming blocks or only the two Uni-modal Priming blocks. Data collection lasted about 30 mins each.

Replication typicality and familiarity ratings. This procedure resembled that of the original typicality and familiarity rating task, with the exception of having two blocks for *concept familiarity*, one with words (*"How familiar are you with the object that the word represents, in your everyday life?")* and one with pictures (*"How familiar are you with the object that the picture represents, in your everyday life?")*, presented in counterbalanced order across participants. For the analysis, we aggregated familiarity ratings for words and objects on

concept level within each participant before averaging across participants. *Image typicality* ratings were aggregated for each concept averaging across participants. Data collection lasted about 30 minutes.

Conceptual Distinctiveness ratings. Finally, we performed a new rating study that was aimed at measuring *Conceptual Distinctiveness (CD)* as it was defined in Konkle et al. (2010). We first carefully instructed participants on the definition of CD as it was done in Konkle et al. (2010), and by presenting a set of example objects rated either as being low on Conceptual Distinctiveness or high in the original investigation. By definition, concepts with high Conceptual Distinctiveness are those whose category members can be easily divided into subgroups of different kinds, regardless of visual appearance. After this introduction, each trial presented a word from our stimulus set in the center of the screen. In addition, the question *"How distinctive are the members of the category denoted by this word?"* was presented along with a six-point scale spanning from one (very similar) to six (very distinctive). Participants responded by clicking with the mouse on a circle corresponding to the number representing their rating. Once they clicked, they saw a black fixation cross in the screen center for about 500 ms before the next word was presented. In total, participants rated all 100 object concepts. We presented the words in randomized order, and participants could take as long as they wanted to make their judgment. Data collection lasted about 15 minutes.

Analysis

Data analysis was performed using R (version 3.6.3, R Core Team, 2020). First, we excluded response times smaller than 200 ms and larger than 1500 ms from further analysis. We set a lower cut-off for excluding response times at 200 ms as typically faster response times are highly likely so-called "fast guesses" (Luce, 1986; Whelan, 2008). Since we had instructed

participants to perform the task as quickly and accurately as possible, we assumed that a cutoff at 1500 ms would prevent the inclusion of response times that did not fit this criterion. Our exclusion criteria led to the removal of only 2.7 % of collected RTs in Experiment 1 and of 1 % of collected RTs in Experiment 2 (1.4 % of the total considering the two experiments together); in Experiment 3, 2.0 % of RTs collected were removed. We implemented a logtransformation to obtain a normal distribution to account for the ex-Gauss distribution of reaction time measures. No further pre-processing was administered.

We used linear mixed-effects models (LMMs; Bates et al., 2014) for statistical analyses of log-transformed response times. Independent variables considered in the models were the four frequency measures described above (object frequencies based on the ADE20K and Greene datasets, word frequencies based on SUBTLEX and dlexDB corpora), several continuous covariates and categorical predictors for the experimental conditions (see *Supplementary Materials 1*). The main advantage of LMMs is that one can consider each trial from each participant simultaneously (i.e., estimating crossed random effects of items and participants; Baayen et al., 2008). In all our LMMs, we included intercept-only random effects for participants and object/word meanings. Note that by including random slope estimates the models did not converge, so we followed the recommendations of Bates et al. (2015).

Our analysis was divided into three steps (more details in *Supplementary Materials 1*):

1) First, we implemented a model comparison based on the Akaike Information Criterion (AIC, Akaike, 1981). This step allowed us to compare our four frequency measures and select the frequency measures with the best fit in both modalities. To implement this, we first fit one model per frequency measure (i.e., SUBTLEX, dlexDB, ADE20K, and Greene frequency) separately for the word and the object recognition trials, and then compared the four models of each modality to a "baseline" model that did not include the frequency measure, but that was estimated on the same subset of data. We selected the frequency measures following

these criteria: in the best case, we would have selected two measures, i.e., the best fitting OF and the best fitting WF measure. In the worst-case, none of the frequency measures would have explained variance in both object and word trials. While, in between, we would have selected either only an OF or a WF measure.

2) After selecting the best frequency measures, we ran a LMM estimating the effects of those selected frequencies on the entire dataset (word trials + object trials), and including all categorical factors and continuous covariates, as well as random factors for participants and concepts.

3) When we detected significant interactions between frequency measures and categorical predictors, we also ran post-hoc LMMs to understand the different effects of frequency *between different conditions* (e.g., SUBTLEX in Cross-modal trials vs. SUBTLEX in Uni-modal trials) and w*ithin each condition* (e.g., the simple effect of SUBTLEX in Crossmodal trials and simple effect of SUBTLEX in Uni-modal trials). Note that the estimation of frequency effects, given the structure of linear models, was independent (i.e., controlled for) from the effect of the other predictors/covariates included in the models.

Data, analysis scripts and stimulus materials are all available at the following link: https://osf.io/d3j9h/files/; for more details, see *Supplementary Materials 1*.

Results

Results Experiment 1

The initial model comparison showed that, in the man-made vs natural categorization task, for word recognition trials, only the SUBTLEX and dlexDB measures produced a significantly better fit when included in the models (SUBTLEX WF: $\chi 2$=29.153, p<0.001; dlexDB WF:

$\chi 2$=15.447, p=0.001), while considering OF measures did not produce a better fit (ADE20K OF: $\chi 2$=3.228, p=0.072; Greene OF: $\chi 2$=0.867, p=0.352). For object recognition trials, only SUBTLEX WF resulted in a significant improvement of the model fit ($\chi 2$=6.163, p=0.013; dlexDB WF: $\chi 2$=1.646, p=0.200; ADE20K OF: $\chi 2$=0.051, p=0.821; Greene OF: $\chi 2$=0.310, p=0.578; for more details, see *Supplementary Materials 2*; no multicollinearity was detected: variance inflation factors < 5). The result of this initial model comparison showed that the SUBTLEX measure was the best fitting parameter in both word and object trials, with no significant increase in explained variance for any of the two object-based predictors. Thus, we implemented a detailed investigation of the SUBTLEX WF effect with both word and object datasets merged.

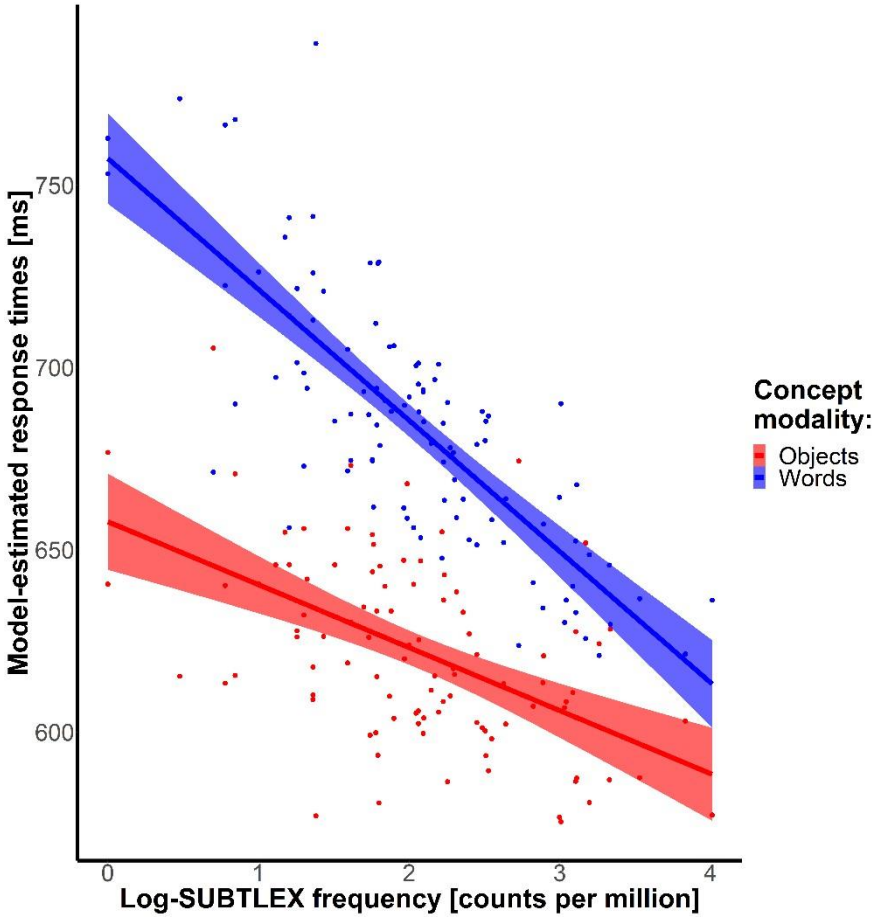
The LMM describing all response times together included a SUBTLEX WF by Concept modality (i.e., words vs. objects) interaction and nine further covariates (see *Supplementary Materials 3* for R-based formula; no multicollinearity detected: variance inflation factors < 5). We found a significant SUBTLEX WF by Concept modality interaction (β=-0.019, SE=0.005, t=-4.160, p<0.001), showing a more substantial facilitatory SUBTLEX WF effect (i.e., faster RTs for high frequency items) for words compared to objects (see Figure 2; for details see *Supplementary Materials 3*).

Two post-hoc models, for objects and words separately, showed significant SUBTLEX WF effects for both words (β=-0.041, SE=0.007, t=-5.794, p<0.001) and objects (β=-0.022, SE=0.009, t=-2.524, p=0.012), but the effect size for words was almost double (1.86 times higher; for more details, see *Supplementary Materials 4* and *5*).

***Figure 2*. Main results of Experiment 1.**

Semantic categorization response times as a function of logarithmic SUBTLEX frequency, separated for objects and words; RTs were estimated based on the SUBTLEX WF x Concept modality interaction term from the selected model. Points present participant-based mean reaction times separated for stimulus type (red: object stimuli; blue:

word stimuli) in the different frequency levels. Lines represent linear fitting of points, and shaded areas represent 95 % confidence interval.



Discussion Experiment 1

The first experiment replicated the well-established SUBTLEX WF effect in word recognition (Brysbaert et al., 2011; Gagl et al., 2020). In contrast to previous literature (Taikh et al. 2015), we also found a SUBTLEX frequency effect for object recognition performance, although the effect for object recognition was weaker than for word recognition. However, all together, findings from this experiment suggest that - given that WF has an effect on both object and word recognition - this effect might reflect processing of what word and object recognition have in common, i.e., the same semantic representation being accessed from two different visual inputs. The phenomenon producing the WF effect during language experience may not just be based on the strengthening of modality-specific representations (WF effect for words),

but also the strengthening of domain-general semantic representation (WF effect also for objects). Interestingly, neither OF measure improved the fit. Thus, OF seems to be less relevant in this simple categorization task.

In Experiment 2, we implemented a priming task to investigate the effect of the novel object-based frequency measures in a paradigm where context is given by a prime allowing prediction of an upcoming visual stimulus. Additionally, we wanted to test the role of WF effects during semantic processing of visual stimuli. The critical manipulation therefore contrasted Cross-modal and Uni-modal Priming (Tversky, 1969; Scarborough et al., 1977; Eisenhauer et al., 2019; 2021). As described earlier, Cross-modal Priming does not involve perceptual processing but rather conceptual/semantic information transfer from prime to target processing. Thus, frequency effects in Cross-modal Priming would signify an involvement of these effects with semantic rather than perceptual processing.

Results Experiment 2

First, we again implemented a model comparison procedure to determine which frequency measure should be part of a detailed analysis. Here, we found that all four frequency measures improved model fit in both stimulus modalities (words and objects; ADE20K OF, objects: χ^2=10.105, p=0.039, words: χ^2=27.302, p<0.001; Greene OF, objects: χ^2=27.547, p<0.001, words: χ^2=43.409, p<0.001; SUBTLEX WF, objects: χ^2=52.695, p<0.001, words: χ^2=43.409, p<0.001; dlexDB WF, objects: χ^2=33.014, p<0.001, words: χ^2=19.105, p<0.001; for detailed information, see *Supplementary Materials 6*). We found that both Greene and SUBTLEX frequencies had stronger fit improvements than their alternatives in both stimulus modalities. Thus, we selected the Greene and SUBTLEX frequency measures for further investigation.

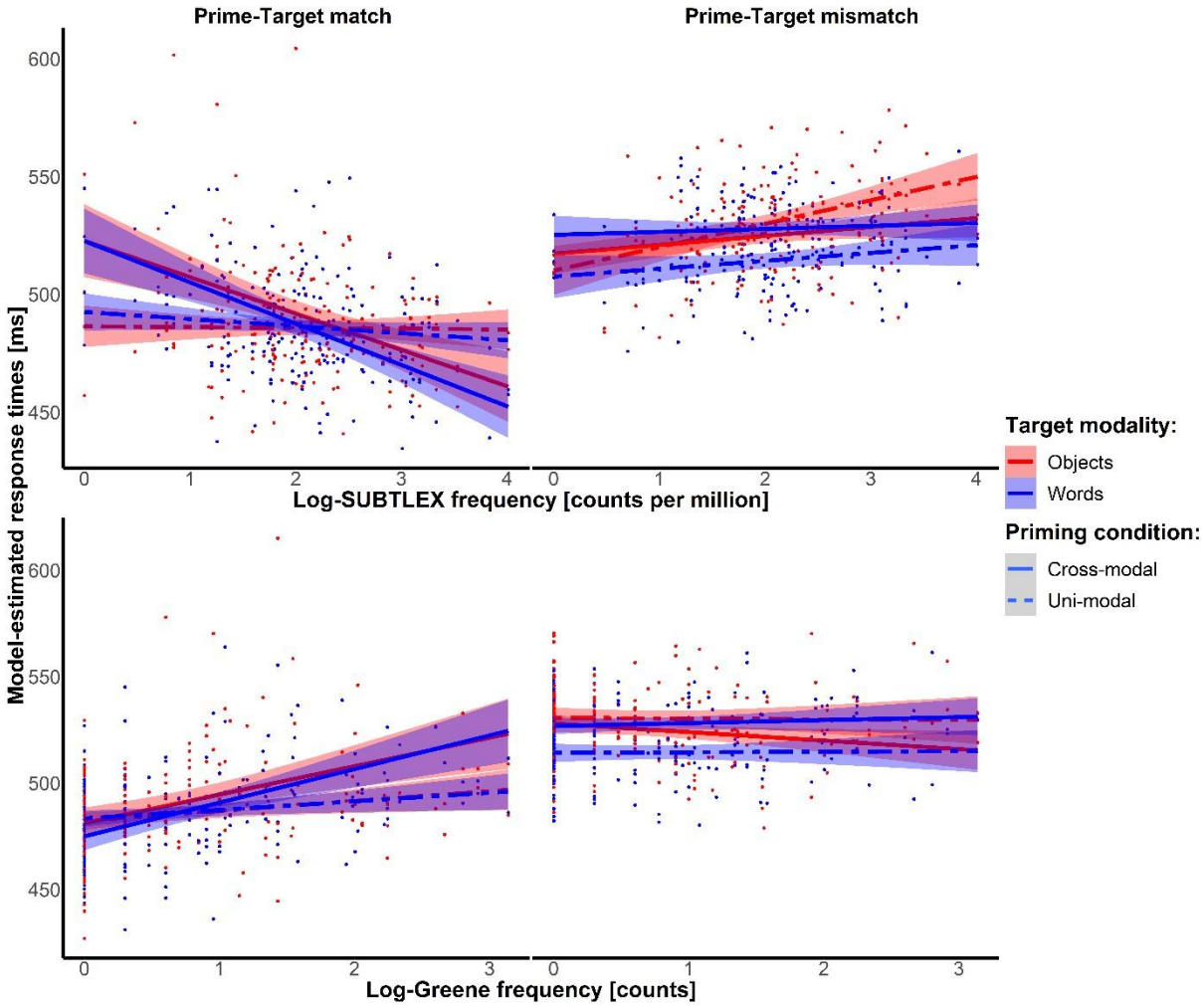
We entered the two measures into a single model, including covariates, categorical predictors and random effects, to describe the response times from the entire dataset of

Experiment 2. Further model comparisons indicated that the interaction between SUBTLEX WF and Greene OF did not improve the model fit beyond the simpler model without the interaction ($\chi 2$=5.455, p=0.708). So, the selected model included each of the two frequency measures in interaction with the experimental conditions (Priming condition: Cross-modal vs Uni-modal; Matching condition: Mismatching vs. Matching; Target modality: Words vs Objects) separately, but not in interaction with each other (for the model formula and other details, see *Supplementary Materials 7*).

When participants had to judge whether prime and target had the same meaning, we found a significant 3-way interaction between frequency, Matching condition, and Priming condition, for both SUBTLEX WF (β=0.017, SE=0.005, t=3.687, p<0.001; *Figure 3 top*) and Greene OF measures (β=-0.020, SE=0.005, t=-4.256, p<0.001; *Figure 3 bottom*, for more detailed information see *Supplementary Materials 7*). Importantly, we found that these interactions had opposite effects for Greene OF and for SUBTLEX WF. However, we found no evidence for Target modality effects, i.e., WF and OF effects in Matching and Priming conditions were similar for words and objects (SUBTLEX WF: β=0.006, SE=0.009, t=0.612, p=0.541; Greene OF: β=0.002, SE=0.009, t=0.227, p=0.821).

***Figure* 3. Main results of Experiment 2.**

Response times as a function of logarithmic SUBTLEX frequency (top plots) and Greene frequency (bottom plots) in the different conditions of Experiment 2; RTs were estimated based on the selected model. Points present participant-based mean response times separated for stimulus type (red: object stimuli; blue: word stimuli) in the different frequency levels. Lines represent linear fitting of points (solid: Cross-modal; dashed: Uni-modal), and shaded areas represent 95 % confidence interval. Top-left and bottom-left plots represent the effects in primetarget matching condition, while top-right and bottom-right plots represent the effects in prime-target mismatching condition.



Post-hoc models showed that the frequency effects were stronger in *Crossmodal Matching* trials than in *Uni-modal Matching* trials (SUBTLEX WF: β=-0.023, SE=0.003, *t*=-7.094, *p*<0.001; Greene OF: β=0.018, SE=0.003, *t*=5.379, *p*<0.001), while, no differential effects were found between *Cross-modal Mismatching* and *Uni-modal Mismatching* trials (SUBTLEX WF: β=-0.006, SE=0.003, *t*=-1.870, *p*=0.062; Greene OF: β=0.002, SE=0.003, *t*=-0.644, *p*=0.520). Besides, we only found strongly significant effects of SUBTLEX frequency (β=-0.019, SE=0.006, *t*=-3.230, *p*=0.001) and Greene frequency (β=0.023, SE=0.005, *t*=4.710, *p*<0.001) in *Cross-modal Matching* trials. The WF and OF effects went in opposite directions: while we observed faster responses for more frequent concepts when investigating the SUBTLEX WF, the Greene OF effect was characterized by faster response for more rare

concepts (*see Supplementary Materials 8* and *9*). In a further control analysis, we showed a substantial stability of the effects for the individual participants and individual concepts across the two modalities (for details, see *Supplementary Materials 10*) which again suggests non-different (i.e., statistically equivalent) processes across modalities.

Discussion Experiment 2

In Experiment 2, we replicated the facilitatory effect of the SUBTLEX WF found in Experiment 1 for both words and objects. It is important to note that we found the SUBTLEX WF effect only when participants categorized objects or words after seeing a semantically matched prime from the other stimulus modality (e.g., a bike image primed by the word "bike" and vice versa), a condition that requires the integration of semantic information from the prime in preparation for the target. The Cross-modal condition specifically includes a prediction process from one modality to the other: it requires processing both object exemplars and their verbal labels within one trial.

A novel aspect that became evident in Experiment 2 was that we also found an effect of the Greene OF in the Cross-modal Matching trials. However, the effect went in the opposite direction, i.e., better performance for low-frequency object concepts than for high-frequency concepts. Both frequency effects were stable across modalities when investigated within each participant and each concept, as shown by our exploratory analysis. Regarding the presence of these two opposite frequency effects, it is worth noting that the model that included an interaction between Greene and SUBTLEX frequency measures did not increase the model fit, implying that the two effects might represent distinct, independent processes. Also, note that the match/mismatch task of Experiment 2 did not result in a global processing advantage for objects compared to words. This was only found in Experiment 1 and replicated previous studies showing the same effect (e.g., Taikh et al., 2015). We believe that since our task in

Experiment 2 was only concerned with the prime-target matching, it resulted in this task being equally difficult for object and word target stimuli.

At this point, one might wonder why the OF improved the fit (and showed significant effect) only in the priming task of Experiment 2 (to be precise, only in Cross-modal Matching trials), and not in the semantic categorization of Experiment 1 (man-made vs. natural). Unfortunately, it is difficult to offer an easy explanation for this unexpected result. It seems that the Greene OF has an effect only when a semantic representation (i.e., concept) is part of a process to predict upcoming input. This process is not part of Uni-modal Priming and unprimed categorization (Experiment 1), where the task does not demand semantic processing (Uni-modal Priming) and it does not use semantic representations to make predictions (Experiment 1). We will now try to explain the WF and OF frequency effects and why both effects occur specifically in the Cross-modal Matching condition, which is important given the high involvement of semantic processing and the predictability of the upcoming stimulus.

The SUBTLEX WF effect is in line with results of Experiment 1 and with typically reported WF effects, reflecting how often a concept has been processed during receptive language processing. One could interpret this effect to reflect the *strength of linguistic experience* with a concept (Brysbaert et al., 2011), based on repeated experiences with that concept during regular language use. It is important to note that this frequency measure is only mildly correlated with the subjective familiarity we additionally collected via ratings, which did not show any relevant impact on reaction times either here or in Experiment 1. This would suggest that there might be a dissociation between what people experience, and therefore rate as being familiar, and how often objects truly occur in the world (Greene, 2016).

In contrast, the Greene OF effect emerged in the opposite direction, i.e., showing facilitation for concepts encountered less often in our visual world. It seems counterintuitive that fewer occurrences could strengthen mental representations, but we can speculate on two

interpretations to explain this effect that has been found for both words and object targets in our study. One possible explanation is that one can remember a concept better when presented with fewer exemplars of that category because more frequent encounters with variable exemplars create interference that weakens the memory trace (Konkle et al., 2010). Based on these findings, we could infer that the facilitation found for low Greene OF concepts (e.g., pineapple) could be due to reduced interference from fewer encounters with exemplars of that object concept during the visual perceptual experience. In contrast, more frequently encountered object categories (e.g., tree) might produce a weaker representation due to exposure to more exemplars creating the abovementioned interference.

Alternatively, the OF effect, which is only detected in congruent Cross-modal Priming, could be explained based on the predictability of the stimulus features from conceptual representations. Objects that are less frequent in the databases might be the expression of more narrow categories (less exemplars and more homogeneous), and their features would be well predictable in contrast to concepts from more broad and thus frequent categories (more exemplars and more heterogeneous). This explanation also relates to theories more deeply concerned with the neuronal preparation for highly predicted incoming stimuli, like predictive coding theories (Rao & Ballard, 1999) or sharpening (Kok et al. 2012; 2017). Evidence from similar experiments using words (Eisenhauer et al., 2019, 2021; Gagl et al., 2020), objects (Summerfield et al., 2008; Richter et al., 2018), faces (Olkkonen et al., 2017) or Cross-modal Priming paradigms (Kok et al., 2012; 2017) have provided findings that indicate feature-based prediction effects. To reevaluate this finding, we performed additional analyses on the data from Experiment 2 and collected a replication dataset in Experiment 3 to shed more light on the explanation of this initially counterintuitive effect and how it could relate to the interference process presented by Konkle and colleagues (2010), as well as to the categorical structure of the investigated concepts (see next section).

To sum up, these results suggest that when participants perform a task where contextual information (i.e., the prime) is semantically processed, different types of information in semantic memory (supposedly derived from linguistic and visual experience) are being preactivated to facilitate the processing of an upcoming input (i.e., the target). The present findings suggest that these processes seem to be at least partially domain-general and thus might depend less on the modality of the stimuli.

Results Conceptual Distinctiveness ratings

In Experiment 2, we unexpectedly found opposite effect of Greene OF on visual recognition, with less frequent concepts being recognized faster than more frequent ones. Having fewer encounters with an object may constitute an advantage in recognizing these compared to concepts for which we have experienced more exemplars, as higher frequency of occurrence has been shown to produce interference in long-term memory (LTM; Konkle et al., 2010).

To further explore this idea, we have collected ratings of *Conceptual Distinctiveness* adapting a procedure from Konkle et al., (2010; for more details, see the *Materials and Methods* section), which has been used to demonstrate how memory interference for objects presented in many exemplars (i.e., comparable to our high Greene frequency concepts) is reduced for objects whose category can easily be separated into many different subcategories (i.e., categories with a high Conceptual Distinctiveness; Konkle et al., 2010). We would expect that including CD in our model will reduce the Greene frequency effect for concepts with high CD, while the effect of Greene frequency would remain the same for concepts with low CD. To illustrate how this relates to the concepts we used in our experiment, see the examples provided in *Figure 4*.

First, we found that CD and Greene OF had a moderate correlation (r=0.43), where concepts with low Greene OF tended to also be less easily dividable in subcategories, while

concepts with high Greene OF tended to more easily dividable. Then we compared the original main LMM of Experiment 2, fitted on the data of Experiment 2, with an identical model including CD in interaction with Greene and the experimental conditions (for details, see *Supplementary Materials 11*). Despite this new model being more complex in terms of number of parameters, it showed a significantly better fit than the original model ($\chi 2$=36.691, p=0.004; no multicollinearity detected: variance inflation factors < 5). In the new model including CD, results showed that the *interaction between Greene OF and CD* was stronger in *Cross-modal Matching* than in *Uni-modal Matching trials* (β=0.010, SE=0.003, t=-3.139, p=0.002), while no difference of the *Greene OF by CD interaction* was found between *Cross-modal Mismatching* and *Uni-modal Mismatching trials* (β=0.002, SE=0.003, t=0.495, p=0.621). Additionally, the *Greene OF by CD interaction* was found to be stronger in *Cross-modal Matching* than in *Cross-modal Mismatching trials* (β=-0.012, SE=0.003, t=-3.774, p<0.001; for more details, see *Supplementary Materials 11*). As shown in *Figure 5*, in Cross-modal Matching trials, higher Conceptual Distinctiveness was associated with weaker Greene frequency effects (the slope reduced towards zero). Cross-modal Matching trials, the condition with strongest semantic processing and predictable semantic context, was also the only condition that had previously shown strong Greene OF effects and, as hypothesized based on findings by Konkle et al. (2010), the condition with the strongest modulation of Greene OF by CD.

***Figure* 4. Example of interaction between Greene frequency and Conceptual Distinctiveness.**

An example of the hypothesized interaction, using object concepts from our stimulus set. Concepts are shown as the black and white pictures used in the experiments and the associated written words (in the English translation). Exemplar pictures to show different levels of conceptual distinctiveness were taken from the THINGS dataset (Hebart et al., 2019). These were not part of the actual experiment.

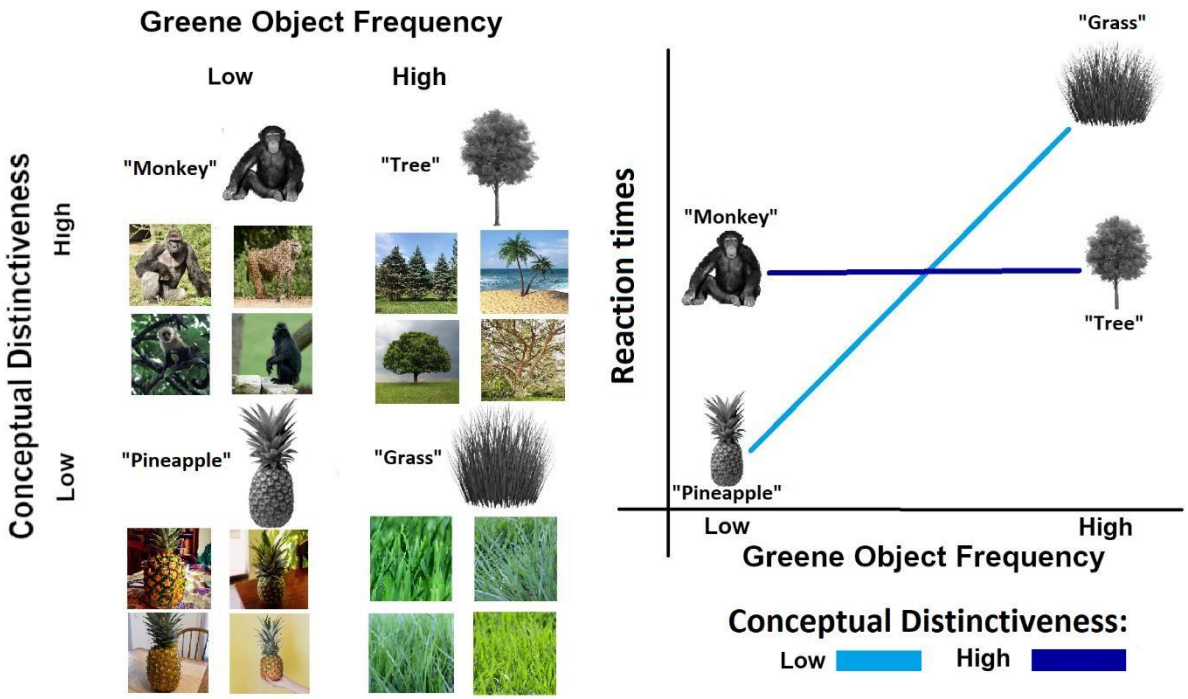
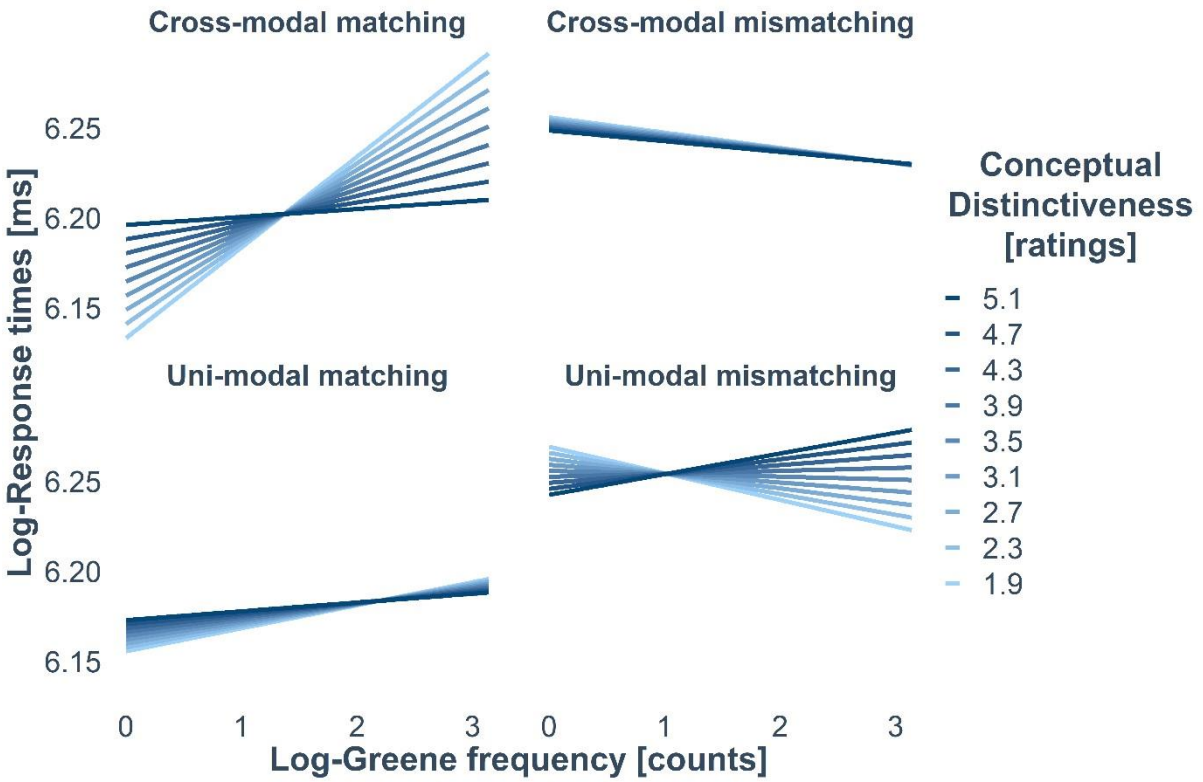


Figure 5. Results interaction between Greene frequency and Conceptual Distinctiveness.

Response times as a function of logarithmic Greene OF in interaction with Conceptual Distinctiveness across Matching conditions and Priming conditions (Cross-modal Matching vs Uni-modal Matching; Cross-modal Mismatching vs Uni-modal Mismatching; Cross-modal Matching vs Cross-modal Mismatching). RTs were estimated based on the selected model. Lines represent linear fitting of log Response times (y axis) by Greene frequency (x axis) for different values of Conceptual Distinctiveness (line colours: lighter = low CD, darker = high CD), in different experimental conditions (top-bottom-left-right panes).



Discussion Conceptual Distinctiveness ratings

When the unexpected processing facilitation for more rare concepts found in the Greene dataset (i.e., a frequency effect with opposite direction) first emerged in Experiment 2, we speculated that the Greene OF measure may reflect memory interference linked to perceptual experience with exemplars of an object concept (Konkle et al., 2010). Konkle et al. (2010) showed that memorability of an object depends on how many exemplars of that category were previously encountered, with more encounters creating a stronger interference that weakened the memory trace. Crucially, this interference for higher number of exemplars of an object was reduced for object categories with higher Conceptual Distinctiveness (i.e., whose members were more easily distinguishable into subgroups of different kinds; Konkle et al., 2010). Therefore, when object categories have low Conceptual Distinctiveness (i.e., it is difficult to divide their

exemplars in subcategories), the number of occurrences of an object has a strong impact on the mental representation (many occurrences = strong interference, few occurrences = weak interference); however, when object categories have high Conceptual Distinctiveness (i.e., it is easy to divide their exemplars in subcategories), the number of occurrences of an object does not influence mental representation to the same degree (few occurrences and many occurrences = similar weak interference). As stated in the Discussion of Experiment 2, this interpretation of Greene OF reflecting an interference process is only one possible explanation. One may also argue that less frequent objects reflect more narrow categories, which would offer more precise predictions of upcoming sensory input in Cross-modal Matching trials. We believe that this analysis of Greene OF in relation to Conceptual Distinctiveness of object categories might offer new, valuable insights on both these interpretations (for more detailed discussions please see *Supplementary Materials 11*).

Similar to Konkle et al. (2010) we found impaired performance for object categories that are encountered in more exemplars (higher object frequency) compared to object categories that are encountered in less exemplars (lower object frequency). And like Konkle et al. (2010), when Conceptual Distinctiveness (CD) was considered, the facilitation for more rare objects (low Greene frequency) was strongly reduced for those objects concepts that have more distinctive subgroups (high CD).

The example in *Figure 4* illustrates the influence of Conceptual Distinctiveness on the effect of the frequency of objects occurrence. *High Conceptual Distinctiveness* identifies the various visual experiences from a diverse set of exemplars (e.g., pine tree or palm tree; gorilla or macaque) that are connected to both frequently encountered (e.g., tree) and rarely encountered (e.g., monkey) objects. Instead, *Low Conceptual Distinctiveness* identifies the similar visual experiences from a homogeneous set of exemplars that are connected to both frequently encountered (e.g., grass) and rarely encountered (e.g., pineapple) objects. Following

the interference explanation of Konkle et al. (2010), the concepts that are encountered in many exemplars (high Greene OF) but have a diverse set of exemplars (high CD) are somehow privileged as the interference from other exemplars or different visual encounters is limited and counteracted for. For concepts with low CD, where it is less easy to distinguish between exemplars, an interference effect can be expected if many exemplars are encountered (high Greene OF).

These considerations also allow us to discuss the alternative explanation according to which the Greene OF effect is due to less frequent objects having more narrow categories allowing more precise predictions. This interpretation is especially interesting as we, again, found the interaction most strongly in the Cross-Modal Priming condition. CD is a way to measure if a category is narrow or wide in terms of the kinds of exemplars. We have shown that low OF concepts can have low CD (in line with this alternative explanation) but also high CD (opposing this alternative explanation). Indeed, the analysis of interactions between CD and the Greene OF measure could be used to show how the narrowness/width of a category impacts the frequency of occurrence: for narrower categories the frequency of occurrence has a strong impact on behavior, while for wider categories the frequency is less relevant. That is, when predicting an upcoming word or object from a low CD category it seems to be particularly beneficial for performance when the OF is low. However, clearly, the two dimensions (Greene OF and CD) do not overlap.

To conclude our discussion on Conceptual Distinctiveness, our analyses have shown how the effect of frequency of objects occurrence in real-world scenes is related to and dependent on the subcategorical structure of object concepts. In the next section, we present Experiment 3, a large-scale replication of the priming experiment, with the goal to reduce potential cross-experiment carry-over effects and object concept repetitions. In the original study (Experiments 1 and 2), participants performed the tasks in every condition (repeated

measures / within-participants design), which exposed them to many repetitions (18 times) of each concept (as either object picture or written word, as either prime or target). Despite the statistical advantages of within-participants designs, e.g., the reduction of variance from individual differences, potential carry-over effects could have created artificial frequency effects (especially since the Greene OF effect was unexpectedly going in the opposite direction of the WF effect). In Experiment 3, we therefore reduced the number of repetitions from 18 times to 8 by including two separate groups of new participants each of which performed either the Cross-modal or Uni-modal Priming tasks (between-participants design).

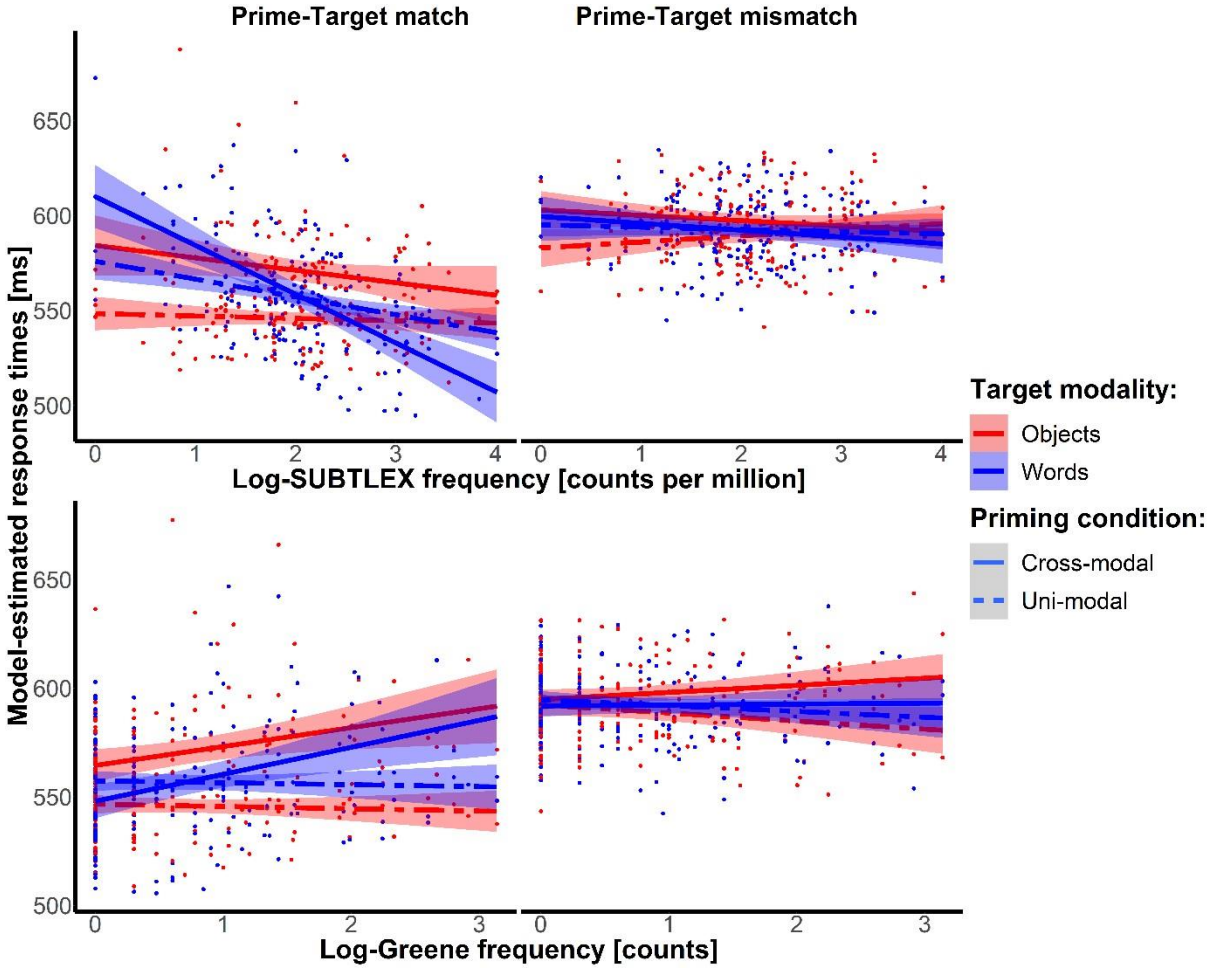
Results Experiment 3

Given that our aim was to replicate Experiment 2, we followed the same analysis, starting with the main model (i.e., no AIC-based frequency selection implemented). The only difference in the model structure was that in the current experiment (Experiment 3), we included the newly collected ratings of concept familiarity and image typicality from an independent participant sample, whereas in Experiments 1 and 2 we used ratings from the same participants who performed the task (no multicollinearity detected: variance inflation factors < 5; see *Supplementary Materials 12*).

Again, participants had to judge whether the meaning of the prime and target matched. We replicated the significant interactions between Greene OF, Matching condition, and Priming condition (β=-0.010, SE=0.005, *t*=-2.165, *p*=0.030); however, the interaction of SUBTLEX, Matching condition, and Priming condition was not significant (β=0.009, SE=0.005, *t*=1.880, *p*=0.060), but qualitatively in the same direction as in Experiment 2. Replicating Experiment 2, the interactions again revealed an effect in the opposite direction for SUBTLEX WF and Greene OF, while the two interaction effects were reduced in their effect size (i.e., about half of the effect size compared to Experiment 2).

***Figure* 6. Main results of Experiment 3.**

Response times as a function of logarithmic SUBTLEX frequency (top plots) and Greene frequency (bottom plots) in the different conditions of Experiment 3; RTs were estimated based on the selected model. Points present participant-based mean response times separated for stimulus type (red: object stimuli; blue: word stimuli) in the different frequency levels. Lines represent linear fitting of points (solid: Cross-modal; dashed: Uni-modal), and shaded areas represent 95 % confidence interval. Bottom-left and top-left plots represent the effects in prime-target matching condition, while bottom-right and top-right plots represent the effects in prime-target mismatching condition.



In a post-hoc analysis that disentangled the interaction effects, we replicated the finding that the frequency effects were stronger in *Cross-modal Matching* trials than in *Uni-modal Matching* trials (SUBTLEX WF: β=-0.014, SE=0.003, *t*=-4.324, *p*<0.001; Greene OF: β=0.017, SE=0.003, *t*=5.165, *p*<0.001). Again, no difference was found for the frequency effects in *Mismatching* trials between *Cross-modal* and *Uni-modal* Priming (SUBTLEX WF: β=-0.006, SE=0.003, *t*=-1.664, *p*=0.097; Greene OF: β=0.006, SE=0.003, *t*=1.959, *p*=0.050;

see *Supplementary Materials 13*). Compared to Experiment 2, the effect size of difference of the SUBTLEX WF effect between *Cross-modal Matching* and *Uni-modal Matching* trials was reduced by more than 1/3 (Beta in Exp. 2: -0.023; Beta in Exp. 3: -0.014), while the difference of effects of the Greene OF between the two conditions was similar to Experiment 2 (Beta in Exp. 2: 0.018; Beta in Exp. 3: 0.017; for more details, see *Supplementary Materials 13*).

Again, the strongest frequency effects were found in *Cross-modal Matching* trials. With less trials per person, only the SUBTLEX frequency effect was significant (β=-0.013, SE=0.006, *t*=-2.253, *p*=0.024), while the Greene OF effect was not (β=0.010, SE=0.005, *t*=1.873, *p*=0.061). Qualitatively, the two effects again went in opposite directions. That is, we again found facilitatory effects for more frequent concepts for the SUBTLEX WF (faster RTs for high frequency items), while facilitatory effects emerged for more rare concepts for the Greene OF (faster RTs for low frequency items).

Contrary to Experiment 2, we found a significant interaction involving SUBTLEX, Matching condition, Priming condition, and Target modality (β=0.021, SE=0.009, *t*=2.269, *p*=0.023; see the Prime-Target match pane for SUBTLEX WF in *Figure 6* and for more details *Supplementary Materials 12*), which indicates a different modulation of words and objects as a function of SUBTLEX WF. Post-hoc investigations found that the difference of SUBTLEX WF effects between Cross-modal and Uni-modal Matching trials was stronger for words than for objects (β=-0.016, SE=0.007, *t*=-2.401, *p*=0.016).

Discussion Experiment 3

Experiment 3 investigated whether the WF and OF effects would still emerge when potential carry-over effects from previous exposure to the same concepts were minimized. One of the main motivations was that multiple presentations of the same concept may alter the perceived frequency of individual concepts, causing spurious effects. To reduce the number of

presentations, we exposed one group of participants only to the Uni-modal and another group to only the Cross-modal Priming condition of Experiment 2.

In general, we largely replicated the main interaction effect found in Experiment 2: That is, SUBTLEX WF and Greene OF had opposing effects and these effects differed as a function of matching and priming condition. More specifically, we replicated the findings that suggested that frequency effects are stronger when deeper semantic processing is required (i.e., frequency effect in Cross-modal Matching trials vs. Uni-modal matching trials), and that these effects seem to reflect a pre-activation from a semantically matched stimulus. Moreover, as in Experiment 2, the WF effect qualitatively indicated faster responses to frequently occurring concepts, while the OF effect was characterized by faster responses to rare concepts.

Of note, our post-hoc analyses revealed some differences. Specifically, we found reduced effect sizes for WF and OF, which led to the Greene OF effect not reaching significance (reduction of 1/3 for the WF and > 1/2 for the OF effect). One explanation would be that fewer repetitions could reduce effect sizes. However, to account for this issue, we controlled for the number of concept repetitions using a covariate in both Experiments 2 and 3, ensuring that the confound of this variable on the frequency effects was minimal. Adding the parameter to the model increased model fit but did not affect the effect size estimates or the t-statistics. Potentially, the reduced number of occurrences of concepts in Experiment 3 compared to 2 might have resulted in less strong semantic associations of the words and object images, explicitly influencing the effects in Cross-modal Priming. Alternatively, or additionally, the between-participant design of Experiment 3 could be an explanation for this difference considering that this experiment showed a higher variance from individual differences compared to Experiment 2, which used a within-participant design. Importantly, we estimated the random effects for participants in both analyses, which should reduce the influence of the differences in design. Based on these considerations, we can summarize that the word

frequency effect, as expected, reliably occurs across experiments, while the object frequency effect seems to be more volatile.

It is also worth mentioning that two other effects emerged in the replication: A) a SUBTLEX WF effect was found for Uni-modal Matching trials with similar size and direction of the one in Cross-modal Matching trials. However, the post-hoc analysis between conditions showed that the effect in Cross-modal Matching trials remained stronger than the one in Unimodal Matching trials, supporting our hypothesis that frequency effects are strengthened by deeper semantic processing reflecting aspects of conceptual representation; B) a four-way interaction between SUBTLEX x Matching condition x Priming condition x Target modality was found, which, when explored, revealed that the effect was mainly driven by a significant SUBTLEX WF facilitation in *Cross-modal Matching trials with words* as target, while it was less pronounced for *Cross-modal Matching trials with objects* as target. Despite this difference to the original Experiment 2, the weaker influence of SUBTLEX WF on object processing resembles the one found in the semantic categorization task of Experiment 1. This stronger frequency-mediated priming effect for words might reflect the fact that words are visually more homogenous than objects, resulting in a more precise prediction of the visual aspects of the upcoming target (Gagl et al., 2020).

General Discussion

Investigating how semantic representations are accessed via different input modalities is a critical step in better understanding how humans store and organize knowledge about the world. The three experiments described in this manuscript provide evidence that high linguistic exposure to a semantic concept (i.e., how often it occurs or is used in our language) increases

recognition performance of both written words and object images (as measured by SUBTLEX WF effect). Furthermore, we present findings suggesting that semantic access might be facilitated not only when concepts are used frequently in language but also when they occur rarely in our visual world (as measured by the Greene OF effect). This phenomenon is possibly modulated by the specific categorical structure of each concept (i.e., the interaction of Greene OF effect with Conceptual Distinctiveness). Finally, we provide insights suggesting that these two effects reflect independent factors affecting visual word and object perception. All frequency effects seem to be substantially strengthened by a greater depth of semantic processing, as seen in the dependence of frequency effects on the type of task. In the following section, we will discuss the various findings in more depth.

SUBTLEX word frequency effect and strength of linguistic experience

The observed effect of subtitle-based frequency measure (SUBTLEX WF) replicated previous findings on word recognition (e.g., Brysbaert et al. 2011; Eisenhauer et al. 2021) and again showed that subtitle-based frequency estimates predict performance better than frequency estimates based on written text corpora (e.g., dlexDB, Heister et al., 2011; see *Supplementary Materials 14).* The novel aspect here is that contrary to Taikh and colleagues (2015), a wordbased frequency measure was also found to influence object recognition. Crucially, previous studies included multiple predictors of semantic richness in their regression models, which were not available for the stimulus material used here. These semantic richness measures could be of interest as they previously showed moderate correlations with the SUBTLEX WF measure (Taikh et al., 2015). However, a reanalysis of the WF effect in Experiment 1 that included Conceptual Distinctiveness (a measure that likely correlates with semantic richness) as a covariate did not change the pattern of effects described above. Thus, it is unlikely that the

observed WF effect in object recognition would have emerged as a confound (see *Supplementary Materials 15*). Nevertheless, future studies should include a larger set of semantic richness measures in order to determine the unique contributions of semantic richness on the one hand and WF on the other.

The finding that subtitle-based (i.e., SUBTLEX) but not text-based (i.e., dlexDB) WF effects were present in both stimulus modalities (i.e., words and objects) confirmed that subtitles are a more reliable source of estimation, and this measure is interpreted not just as reflecting the strength of experience with a word (effect in word recognition), but the *strength of experience with a concept* (effect in both object and word recognition). Indeed, as we control for many perceptual and linguistic variables, we suggest that this effect was modulated by the access to semantic representation required by the tasks. We could speculate that this strength is built through linguistic experience and, after that, transfers to other non-linguistic modalities. Such an interpretation would be in line with the idea that language would be not merely a means of communicating semantic information but also shaping semantic representations (Lupyan & Lewis, 2019).

Greene object frequency effect and its relationship with structure of object categories

In contrast to the SUBTLEX-based word frequency effect for objects and words, the Greene OF measure showed an opposite frequency effect: recognition performance in response to less frequent concepts was faster when compared to frequently encountered concepts. This inverted OF effect was surprising as we had computed the two measures based on a similar logic, i.e., counting occurrences in a dataset and capturing properties of the word. Furthermore, the OF effect did not emerge when we presented objects or words in isolation, but only in the matching trials of the Cross-modal Priming task, i.e., in context of a predictable prime stimulus, irrespective of modality. Note that in the same condition, we observed a substantial WF effect.

In these trials, the primed concept is retrieved from semantic memory to prepare participants for the upcoming stimulus, which is visually different but semantically matched.

This semantic memory involvement led us to reevaluate our findings based on the results reported in Konkle et al. (2010), who investigated memory interference processes when the number of exemplars belonging to a category was manipulated. They showed that we have the worse memory for the specific instance of frequently encountered objects (e.g., cars) because the increased number of exemplars creates interference. Conversely, we remember objects that we rarely encounter (e.g., pineapple) better because they suffer less from the interference of different exemplars (Konkle et al., 2010). Crucially, we found that the OF effect was only found when the objects came from a category that is not easily dividable into subgroups of different kinds, as measured by Conceptual Distinctiveness (CD). This seems to be due to the fact that when concepts can be easily divided into subgroups, this more complex division counterbalanced the interference effect produced by repeated encounters with exemplars of that category.

We want to stress that although CD and Greene OF are moderately correlated (r= 0.43), our finding of an interaction of the two measures showed that they explain different parts of variance. Thus, one should interpret the Greene OF effect beyond the effect of homogeneity/heterogeneity of object categories on the prediction of upcoming input. However, this explanation has highlighted the relevant issue of how categorical structure (more or less homogeneity) interacts with object occurrence and how this can impact the predictability of upcoming input.

Frequency effects and semantic processing

The results from the priming tasks (Experiments 2 and 3) are crucial to supporting the notion that frequency effects are also semantic. We found that they are more robust in Cross-modal

(i.e., integration of information across modalities) than Uni-modal Priming tasks (i.e., integration of information within modalities). Besides, these effects seem to reflect the processing of a corresponding prime-target combination rather than just recognizing the target, as the frequency effects were much more substantial in Cross-modal Matching trials than in Cross-modal Mismatching trials. Nevertheless, in Experiment 3, we only found a WF effect when an object picture primed a matching word but not when a word primed a matching object. It could be that the priming effect mediated by frequency is more substantial when words are the target stimulus. A potential explanation could be that words are more visually homogeneous stimuli than objects, making the upcoming word target easier to predict down to the individual pixel level (Zhao et al., 2019; Gagl et al., 2020; Wang & Maurer, 2020).

In sum, the present findings point to the semantic nature of the measured frequency effects. Moreover, these frequency effects might reflect processing common to both word and object recognition. Since they show similar patterns for word and object trials and given that what our word and object stimuli have in common is their meaning, one could speculate that this typical processing relates to accessing abstract conceptual representations.

Possible mechanisms underlying frequency effects

Regarding the mechanisms underlying the observed frequency effects in Cross-modal Matching trials, one could hypothesize that they resemble the neural processes described in Kok and colleagues (2017), where an auditory prime pre-activated a representation of a previously matched visual stimulus before its presentation (Kok et al., 2017). Furthermore, in line with our results (i.e., pre-activation facilitation based on frequency), they found that the pre-activation strength could predict behavioral responses. These findings suggest a mechanism of *sharpening* visual representation compatible with expected upcoming input, modulated by some aspects of previous experience. In our case, these aspects might be the strength built

through linguistic experience and the encounters during visual experience that are incorporated into the conceptual representations evoked by the prime.

Analogously, one could speculate that similar processes are occurring during Unimodal Priming too. The crucial difference is that what modulates sharpening is not a semantic representation but a more perceptual representation (e.g., orthographic for words, visual for objects), therefore producing hardly any frequency effects. This finding is in line with the behavioral and MEG evidence reported by Eisenhauer and colleagues (2021) who found frequency effects for words presented in isolation (i.e., as in Experiment 1 described above), but not in a Uni-modal Priming context (i.e., a word primed by the same word as in our Experiments 2 and 3). Notably, they found a modulation of neural activity by orthographic information following the prime and preceding the target word, similarly indicating a sharpening process on the neuronal level (Eisenhauer et al., 2021).

However, given the study's design and methods, we cannot yet draw firm conclusions about the nature of the mechanisms underlying our frequency effects. For example, we cannot rule out that the involved predictive processing (Rao & Ballard, 1999) functions by inhibiting the most common features of upcoming input instead of sharpening it (Gagl et al., 2020). Further investigations are needed to specify the neuronal mechanisms on representations in perceptual and or semantic processes in Cross-modal Priming. Here, electrophysiological measures (M/EEG) would allow for a more fine-grained and better temporally resolved investigation of how optimization of recognition behavior in Cross-modal Priming is implemented on the neuronal level.

Choosing the right dataset for frequency estimations

In general, any decision to use one dataset over another in order to compute frequency measures needs to be approached with great care. One problem lies in the assumption that a chosen

dataset is a good representation of the state of the world, but every dataset, even the largest available, remains an approximation. Besides, the composition of the datasets often reflects biases in the way they were composed and the sources that were used to create them. Moreover, the assumption that a dataset captures universally shared concept representations might not be valid. Factors like expertise and physical or cultural context have a different impact on the individual experience of the world (Kuperman & Van Dyke, 2013).

Of course, the quantity and variety of scene images of the datasets are lower than the corpora usually used for computing WF measures (more than 20 million words of the SUBTLEX database vs the 400,000 object annotations in the ADE20K and 48,000 object annotations in the Greene database). Concerns about the representability of selected image datasets are therefore always valid and must be considered carefully. To account for this concern - and to start somewhere - we decided to include both image datasets (the ADE20K dataset, Zhou et al., 2019, and the Greene dataset, Greene, 2013). Both datasets are widely used by computer vision scientists and cognitive psychologists working on visual cognition (Bonner & Epstein, 2021; Bracci et al., 2021). While the Greene dataset includes fewer annotations than ADE20K, it has the advantage of thoroughly cleaning up spelling mistakes, synonyms, and other issues affecting any frequency analyses based on labelled image databases. So, which frequency measure should be used?

Even though our results confirm that as a word frequency measure, the SUBTLEXbased frequency is the better predictor for categorization behavior, the situation seems less clear for object frequencies, especially given that Greene and ADE20K produce similar result patterns. In our primary analysis, Greene was preferred to ADE20K, given its more robust improvement of model fit in both words and object trials (see details of the AIC-based selection method in the *Analysis* section of *Materials and Methods*). However, also ADE20K showed a significant improvement in model fit in both modalities in Experiment 2.

The two datasets have both pros and cons, as pointed out previously: ADE20K is clearly superior when it comes to dataset size and variety of images, while the Greene dataset would be the preferred choice when looking for high quality annotations. Ideally, revising ADE20K annotations with the same approach offered by Greene (2013) would likely create the best of both worlds. However, more practical ways to decide which measure to employ would be to consider aspects like the number and types of object stimuli and the scenes they are typically found. For larger and more diverse sets (e.g., natural vs man-made, public vs private), it is more likely to find good estimates in the ADE20K dataset. For smaller and more homogeneous sets (e.g., objects found in a house), the quality of Greene's annotations could beat the quantity of ADE20K's ones. In general, one goal for the future would be a database with a high number of quality annotations that, similar to word databases, contains a sufficient number of examples for a more appropriate estimation of object frequency (i.e., at least 20 million; Brysbaert et al., 2011).

Pros and cons of using labeled image databases for cognitive studies

As previously discussed, estimating any type of frequency from databases can create unwanted biases in the frequency measures being extracted. In addition to these database dependent biases, calculating object frequency measures includes further hurdles. For instance, linguistic image databases make the evaluation of the visual domain dependent of the linguistic domain. In addition, labeling decisions must be made for each object. At times labeling decisions can be easy (e.g., pineapple), but sometimes there are very explicit decisions to make (e.g., are all types of cars simply labelled as "cars" or by their brand name/type, e.g., "Porsche" vs. "Jeep"). The problem might be more severe for highly general concepts (i.e., trees, cars, animals, and others).

Crucially, these decisions can and will have an impact on the computed frequency of occurrence, and could create differences between datasets (although ADE20K and Greene OF show strong correlation r=0.81 and led to similar results, see *Supplementary Materials 16*). The annotators of the images in the Greene database were instructed to use entry-level labels (e.g., "car", not "vehicle" or "Mercedes"), and labels were inspected and corrected for synonyms and similar confounds. We believe that the issue of biases from labelling has been addressed in our study in three ways: (i) the OF effect was always estimated independently of the WF effect, since both were included in the same model. This would allow to rule out differences arising from common vs uncommon labels; (ii) we have shown that the Greene OF effect is present only for concepts with low Conceptual Distinctiveness, which have a more homogeneous set of exemplars and thus should be less prone to be biased by possible variabilities in labelling; (iii) the OF effect was estimated independently of image typicality (included as covariate), which measures how the employed image stimuli are a typical exemplars of the categories denoted by the employed word stimuli (i.e., the labels). In this sense, it represents how strongly image-word pairs are associated and therefore predictable. Still, the labelling procedure of image datasets remains an important issue that needs to be considered in the future.

We want to stress that - to our knowledge - this is the first time that metrics such as object frequencies were computed and used to predict response times in a way traditionally done with WF measures. Even the other OF measure used in this study, ADE20K OF (Zhou et al., 2019), was found to produce similar patterns of effects in terms of size, direction, and probability when repeating the analysis of Experiment 2, substituting Greene OF with it (for more details, see *Supplementary Materials 16*). Similar fit and result patterns indicate that datasets of annotated and segmented objects capture aspects of the world and our experience with it, which are relevant for our cognitive system in general. Therefore, we hope that this first attempt at studying object-based frequency measures gives rise to broader investigations, as it

was done by some studies in cognitive neuroscience that already started with investigations in this direction (Bonner & Epstein, 2021; Bracci et al., 2021).

Conclusion

To conclude, this study aimed to expand and innovate previous investigations of semantic access from words and objects by employing new measures of object frequencies and comparing them to established word frequency measures. In a first attempt, we identified language-based and image-based frequency measures and demonstrated how they differentially influence recognition processes which might reflect two organizational principles for conceptual knowledge. Moreover, we showed that very different visual information (words vs. objects) could lead to relatively similar processing when accessing conceptual knowledge, providing further evidence for the strong interrelation between language and vision. We hope that this study will lead to further investigations of both word- and object-based frequency measures to increase our understanding of accessing meaning from visual input.

Context

The word frequency (WF) effect in visual word recognition is a well-established empirical finding, while there is little evidence about object frequency (OF)'s role in object recognition. Word and object recognition have the common goal of accessing meaning based on visual input. This similarity raises questions about whether similar parameters modulate object and word recognition. Since more frequent words are recognized more efficiently, we investigate whether the frequency of occurrence also similarly affects object recognition. This team of researchers - with expertise in visual word and object recognition - joined forces to investigate the process of accessing the meaning of objects and words using object and word frequency

measures. We, therefore, applied new metrics of object frequency based on state-of-the-art datasets of annotated images and evaluated them in comparison to widely used metrics of word frequency. Beyond this, we aimed to determine common aspects of object and word processing that would give further evidence for the strong interrelation between language and vision while providing a starting point for future investigations.

References

- Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, *16*(1), 3–14. https://doi.org/10.1016/0304-4076(81)90071-3
- Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review*, *14*(6), 1177–1182. https://doi.org/10.3758/BF03193109
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *ArXiv:1506.04967 [Stat]*. http://arxiv.org/abs/1506.04967
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823
- Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition*, *29*(7), 986–999. https://doi.org/10.3758/BF03195761
- Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature Communications*, *12*(1), 1-16.
- Bracci, S., Mraz, J., Zeman, A., Leys, G., & de Beeck, H. O. (2021). Object-scene conceptual regularities reveal fundamental differences between biological and artificial object vision. *bioRxiv*.
- Brehm, L., & Alday, P. M. (2020). *A decade of mixed models: It's past time to set your contrasts*. Talk presented at the 26th Architectures and Mechanisms for Language Processing Conference (AMLap 2020). Potsdam, Germany. 2020-09-03 - 2020-09-05.
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The Word Frequency Effect: A Review of Recent Developments and Implications for the Choice of Frequency Estimates in German. *Experimental Psychology*, *58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123
- Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, *41*(4), 977-990.
- Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of Semantic Category-specific deficits? A Critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*(3–6), 213–261. https://doi.org/10.1080/02643290244000266
- Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From Perception to Conception: How Meaningful Objects Are Processed over Time. *Cerebral Cortex*, *23*(1), 187–197. https://doi.org/10.1093/cercor/bhs002
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

- Criss, A. H., & Malmberg, K. J. (2008). Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language*, *59*(3), 331–345. https://doi.org/10.1016/j.jml.2008.05.002
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003
- Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational Similarity Analysis Reveals Commonalities and Differences in the Semantic Processing of Words and Objects. *Journal of Neuroscience*, *33*(48), 18906–18916. https://doi.org/10.1523/JNEUROSCI.3809-13.2013
- Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain Specificity in Visual Cortex. *Cerebral Cortex*, *16*(10), 1453–1461. https://doi.org/10.1093/cercor/bhj086
- Eisenhauer, S., Fiebach, C. J., & Gagl, B. (2019). Context-Based Facilitation in Visual Word Recognition: Evidence for Visual and Lexical But Not Pre-Lexical Contributions <sup/>. *Eneuro*, *6*(2), https://doi.org/10.1523/ENEURO.0321-18.2019
- Eisenhauer, S., Gagl, B., & Fiebach, C. J. (2021). Predictive pre-activation of orthographic and lexical-semantic representations facilitates visual word recognition. *Psychophysiology*, e13970.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, *112*(4), 777–813. https://doi.org/10.1037/0033-295X.112.4.777
- Fairhall, S. L., & Caramazza, A. (2013). Brain Regions That Represent Amodal Conceptual Knowledge. *Journal of Neuroscience*, *33*(25), 10552–10558. https://doi.org/10.1523/JNEUROSCI.0051-13.2013
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8
- Gagl, B., Sassenhagen, J., Haan, S., Gregorova, K., Richlan, F., & Fiebach, C. J. (2020). An orthographic prediction error as the basis for efficient visual word recognition. *NeuroImage*, *214*, 116727. https://doi.org/10.1016/j.neuroimage.2020.116727
- Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00777
- Greene, M. R. (2016). Estimations of object frequency are frequently overestimated. *Cognition*, *149*, 6–10. https://doi.org/10.1016/j.cognition.2015.12.011
- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747
- Harel, J., Koch, C., & Perona, P. (2007). Graph-Based Visual Saliency. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19* (pp. 545–552). MIT Press. http://papers.nips.cc/paper/3095-graph-based-visual-saliency.pdf
- Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS one*, *14*(10), e0223792.
- Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). DlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*(1), 10–20. https://doi.org/10.1026/0033-3042/a000029

- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302-4311.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12
- Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265-270.
- Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, *114*(39), 10473–10478. https://doi.org/10.1073/pnas.1705652114
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386
- Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 802.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13
- Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on information theory*, *38*(6), 1842-1845.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (No. 8). Oxford University Press on Demand.
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60).
- 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. https://doi.org/10.1080/23273798.2017.1404114
- Maxwell, S. E., Delaney, H. D., & Kelley, K. (2017). *Designing experiments and analyzing data: A model comparison perspective*. Routledge.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, *88*(5), 375.
- Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, *20*(6), 705–714. https://doi.org/10.3758/BF03202720
- Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In *Processing of visible language* (pp. 259-268). Springer, Boston, MA.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, *42*(3), 145-175.
- Olkkonen, M., Aguirre, G. K., & Epstein, R. A. (2017). Expectation modulates repetition priming under high stimulus variability. *Journal of vision*, *17*(6), 10-10.
- https://osf.io/d3j9h/files/
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, *51*(1), 195-203.

- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580
- Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of eye movement research*, *2*(5), 1.
- Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *Journal of Neuroscience*, *38*(34), 7452-7461.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature neuroscience*, *3*(11), 1199-1204.
- Robinson, A. P., & Froese, R. E. (2004). Model validation using equivalence tests. *Ecological Modelling*, *176*(3–4), 349–358. https://doi.org/10.1016/j.ecolmodel.2004.01.013
- Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, *77*(1), 157-173.
- Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human perception and performance*, *3*(1), 1.
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
- Shelton, J. R., & Caramazza, A. (1999). Deficits in lexical and semantic processing: Implications for models of normal language. *Psychonomic Bulletin & Review*, *6*(1), 5–27. https://doi.org/10.3758/BF03210809
- Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. *NeuroImage*, *54*(3), 2418–2425. https://doi.org/10.1016/j.neuroimage.2010.10.042
- Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience*, *11*(9), 1004-1006.
- Taikh, A., Hargreaves, I. S., Yap, M. J., & Pexman, P. M. (2015). Semantic classification of pictures and words. *Quarterly Journal of Experimental Psychology*, *68*(8), 1502–1518. https://doi.org/10.1080/17470218.2014.975728
- Tversky, B. (1969). Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics*, *6*(4), 225–233. https://doi.org/10.3758/BF03207022
- Võ, M. L.-H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205–210. https://doi.org/10.1016/j.copsyc.2019.03.009
- Wang, F., & Maurer, U. (2020). Interaction of top-down category-level expectation and bottom-up sensory input in early stages of visual-orthographic processing. Neuropsychologia, 137, 107299.
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475-482.
- Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al. (2009). *Perspectives on psychological science*, *4*(3), 294-298.
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971
- Zhao, J., Maurer, U., He, S., & Weng, X. (2019). Development of neural specialization for print: Evidence for predictive coding in visual word recognition. PLoS biology, 17(10), e3000474.

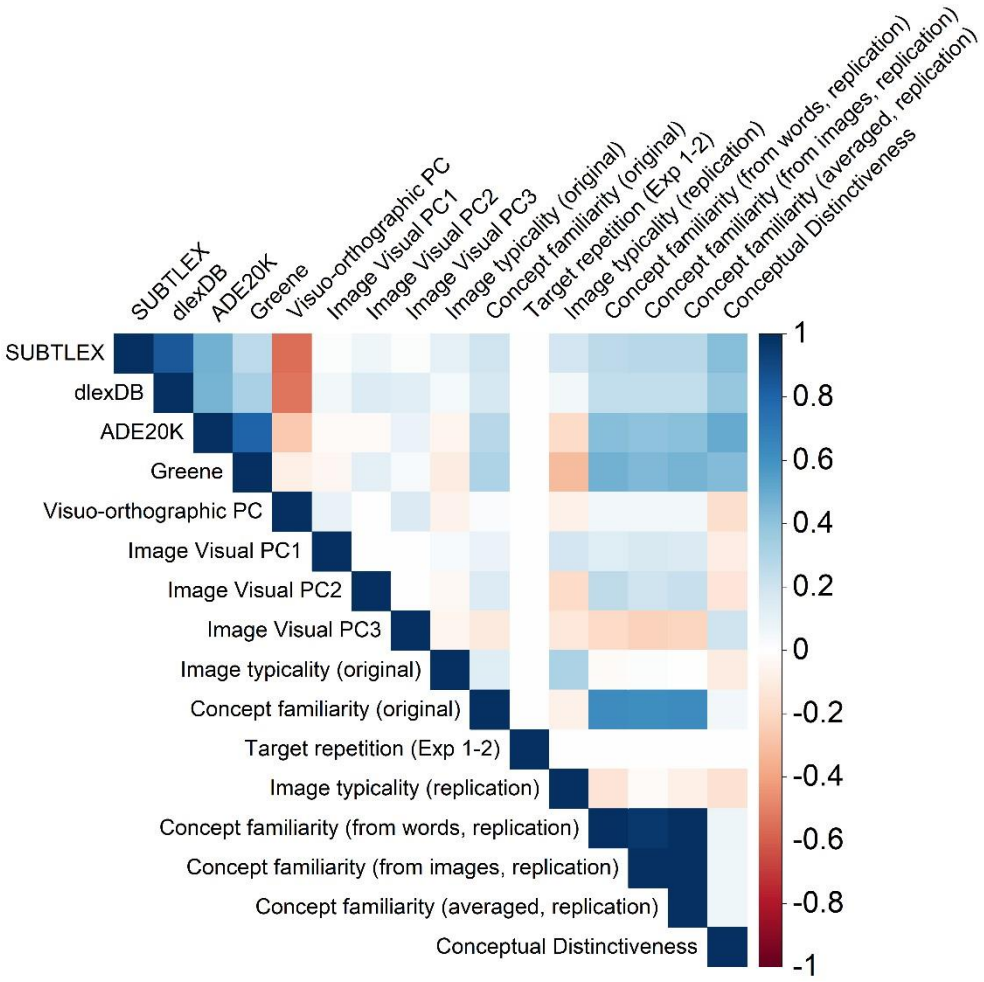
- Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2019). Semantic Understanding of Scenes Through the ADE20K Dataset. *International Journal of Computer Vision*, *127*(3), 302–321. https://doi.org/10.1007/s11263-018-1140-0

Supplementary Materials

Supplementary materials 1 – Factor correlations, distributions, analysis details

Supplementary figure 1 – Correlations between factors measured for the study

Product-Moment Correlation Coefficients for each pair of predictors used in the experiment.

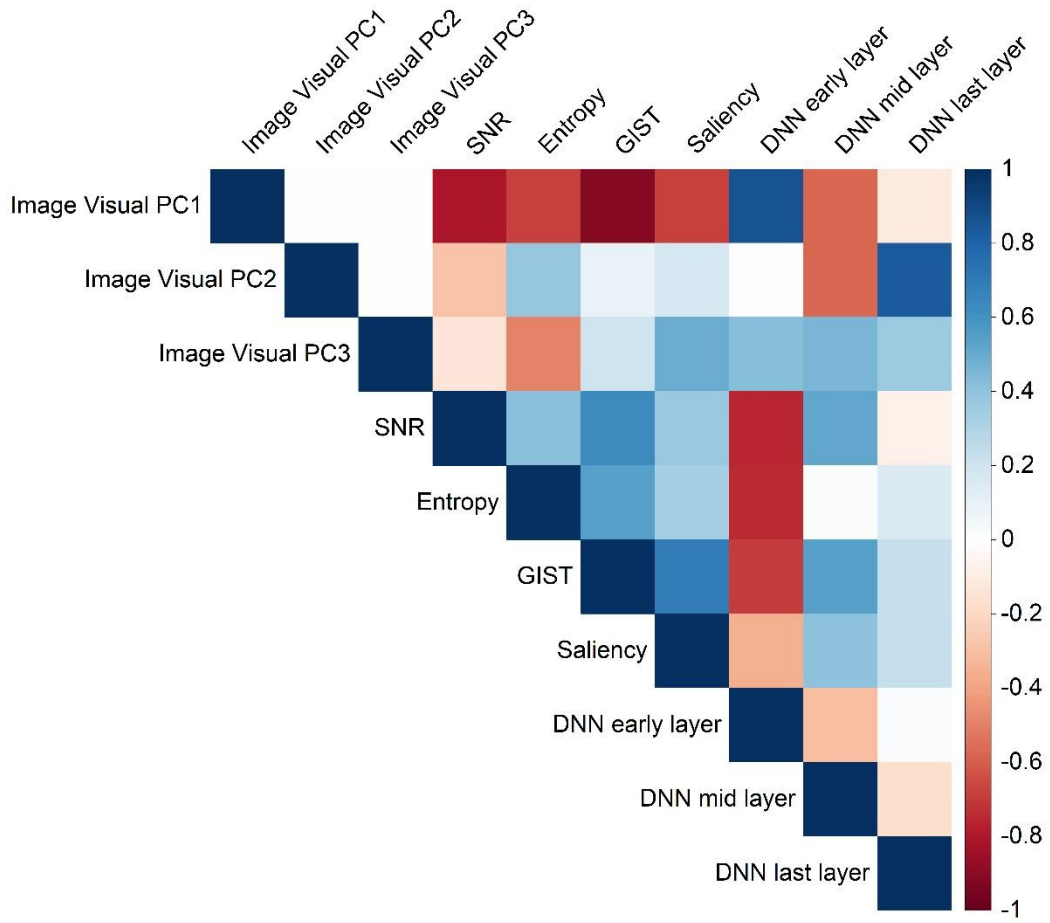


PCA procedure for visual and visuo-orthographic predictors: measures with multidimensional output were averaged to obtain a single value for every image. As a rule of thumb, we selected the Principal Components (PCs) that, alone, explained more variance than what a variable would explain if they all explained the same amount of variance.

Image visual PCs: 7 variables -> threshold: 100 / 7 = 14.29 %. We extracted three orthogonal PCs, explaining more than 84 % of the variance. We labeled the first PC *Image visual PC1* (about 50 % of variance explained, strong positive correlation with convolutional layer 1 of AlexNet, and strong negative correlation with SNR, GIST, Entropy, Saliency and AlexNet layer 4). The second PC was named *Image visual PC2* (about 18 % of variance explained, strong positive correlation with AlexNet layer 7, strong negative correlation with AlexNet layer 4). The third PC was named *Image visual PC3* (about 15 % of variance explained, medium positive correlation with AlexNet layers and Saliency, medium negative correlation with Entropy). We interpret the PC1 as an estimate of low-to-mid-level visual features of the images (stronger weights from AlexNet early layer, SNR, saliency, but also from AlexNet mid layer and GIST), while the PC2 seems to capture more complex mid-to-high-level visual features (stronger weights from AlexNet mid layer and entropy, but also from AlexNet late layer). PC3, however, has a less clear interpretation, capturing part of variance from both low-level and high-level visual features estimates (higher weights for all the three AlexNet layers).

Supplementary figure 2 – Correlations of visual predictors and extracted PCs

Product-Moment Correlation Coefficients for each pair of visual predictors of objects and the Principal Components (PCs) extracted from the PCA on those predictors.



Supplementary table 1. Object image PCA loadings for every variable in every extracted principal component (PCs). They represent the weights of every variable on the extracted PCs.

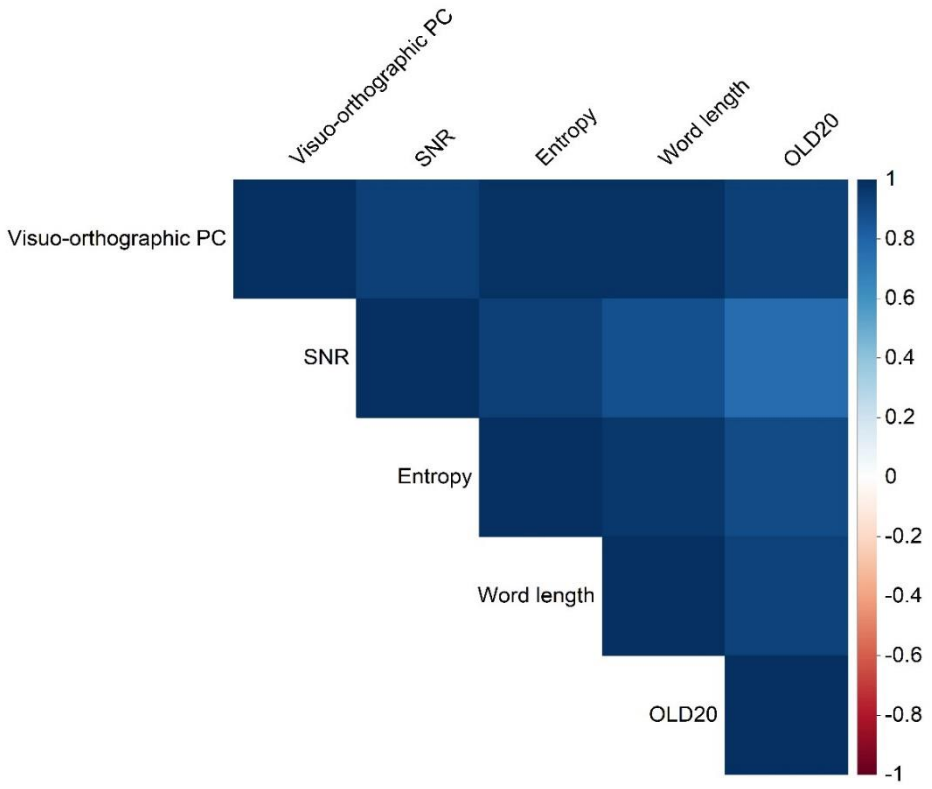
Variables	PC1 loadings	PC2 loadings	PC3 loadings
AlexNet conv1	**0.459**	0.005	**0.413**
AlexNet conv4	**-0.305**	**-0.505**	**0.436**
AlexNet fc7	-0.062	**0.735**	**0.352**
Saliency	**-0.366**	0.154	**0.477**
GIST	**-0.486**	0.080	0.194

Entropy	**-0.366**	**0.337**	**-0.482**
SNR	**-0.434**	-0.249	-0.134

Visuo-orthographic PC: 4 variables -> threshold: 100 / 4 = 25 %; one principal component (PC) was extracted and was labeled *Visuo-orthographic PC* (variance explained circa 92 %; strong positive correlation with all the original variables). Being all the variables highly correlated between them and with the PC, interpretation seems straightforward and difficult at the same time. The rationale for including many vatiables that were expected to be highly correlated was to acknowledge the different levels (visual and orthographical) from which we wanted to extract a covariate able to control for perceptual aspects of a word.

Supplementary figure 3 – Correlations between visuo-orthographic predictors and the extracted PC

Product-Moment Correlation Coefficients for each pair of visuo-orthographic predictors of words and the Principal Component (PC) extracted from the PCA on those predictors.

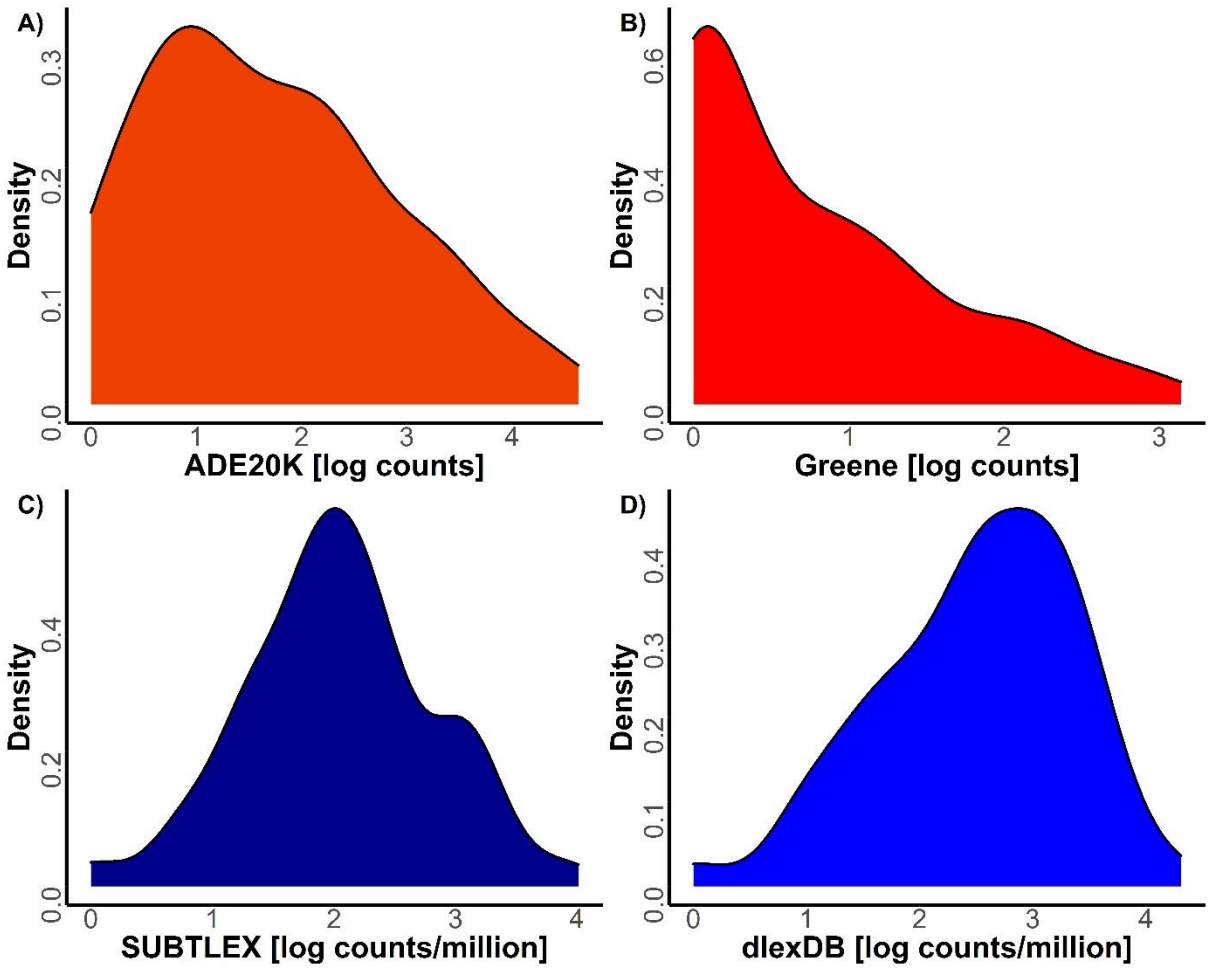


Supplementary table 2. Word image PCA loadings for every variable in the extracted principal component (PC). They represent the weights that every variable has on the extract PC.

Variables	PC1 loadings
OLD20	**0.487**
Word length	**0.512**
Entropy	**0.515**
SNR	**0.485**

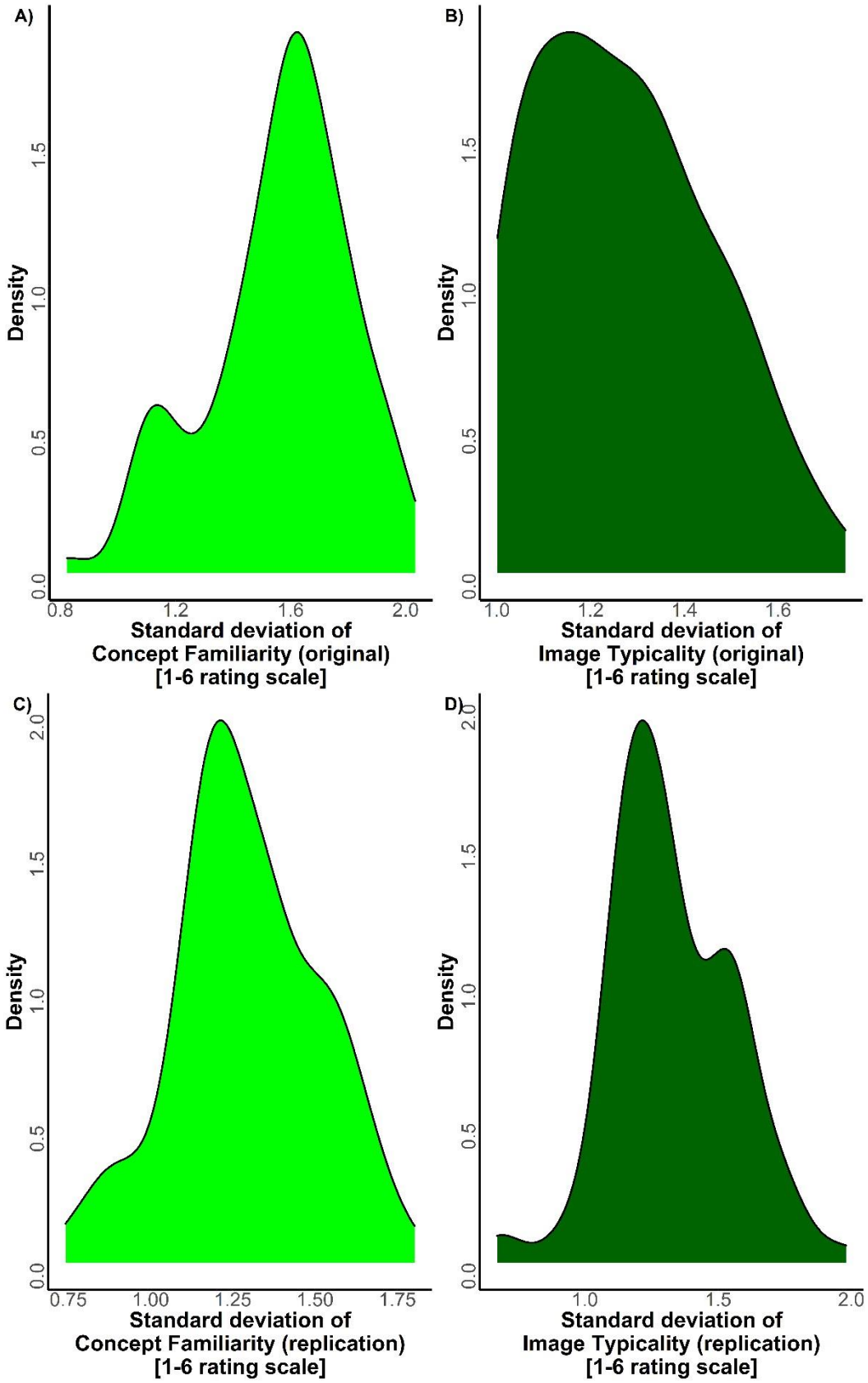
Supplementary figure 4 – Distribution of the frequency measures

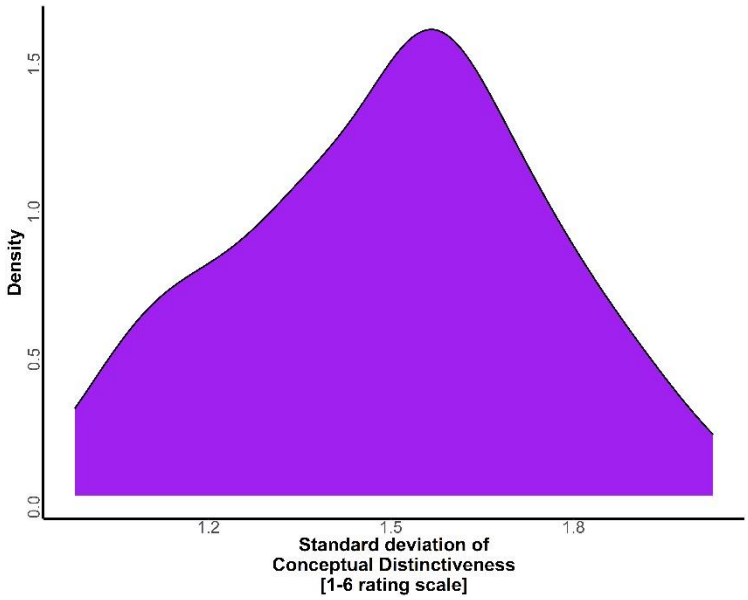
Density distribution of the frequency values for our set of stimuli.



Supplementary figure 5 – Distribution of concept variability across participants for the rating measures

Density distribution of concepts' standard deviation across participants for the collected ratings (scale 1-6)




Density
0.0
0.5
1.0
1.5
1.2
1.5
1.8
Standard deviation of
Conceptual Distinctiveness
[1-6 rating scale]

Analysis details: We fitted Linear Mixed-effects Models (LMMs) via maximum likelihood estimation, and Satterthwaite's method was used to obtain p-values (package lmerTest, Kuznetsova et al., 2017). Using the *scale()* function in R, we transformed each continuous predictor variable onto a common scale which improves model fitting procedures. These continuous predictors are the four frequency measures (SUBTLEX, dlexDB, Greene, ADE20K) and the covariates (Concept familiarity [different in Exp 1-2 and Exp 3], Image typicality [different in Exp 1-2 and Exp 3], Image visual PC1, Image visual PC2, Image visual PC3, Visual-orthographic PC, Target repetition [different in Exp 1-2 and Exp 3]).

For the coding of contrasts in categorical predictors (*Exp 1:* Concept modality: Words – Objects; Concept category: Natural – Man-made; Trial accuracy: Correct – Incorrect*; Exp 2 and Exp 3:* Target modality: Words - Objects; Priming condition: Cross-modal – Uni-modal; Matching condition: Mismatching – Matching; Trial accuracy: Correct - Incorrect), we used sum contrast coding, which in our case gave us an estimate of the difference between the two levels of each of our categorical variables, like main effects in a multi-way repeated measures ANOVA (Schad et al., 2020; Brehm & Alday, 2020).

Including trial response accuracy as a categorical covariate in the LMMs allows us to consider the variance explained by the output of the task (i.e., correct or incorrect trial), but at the same time to estimate the impact of the other variables independently from the output of the task itself. Besides, this way, we did not have to exclude further trials from the analysis, and we could exploit the flexibility offered by LMMs.

To account for the multiple repetitions of the same object concepts within participants and a potential carry-over effect that could confound frequency effects, we included in the models a numeric covariate Target repetition that represents the number of times that the

current target concept has been presented (as either a word or an object image, as either target or prime).

To prevent misinterpretation of the effects and confounds due to high correlation of the predictors, we assessed potential multicollinearity of the models by computing the variance inflation factors (VIFs) for each term in each model, using the *check_collinearity()* function in R (package performance; Lüdecke et al., 2021). When variance inflation factors are below 5, there are low correlations between predictors and therefore no predictors need to be excluded to avoid confounds in the interpretation of the results. When the variance inflation factors are higher than 5, those predictors should be excluded from the model and the analysis should be repeated.

1) We implement a model comparison based on the Akaike Information Criterion (AIC, Akaike, 1981). This step allowed us to compare our four frequency measures and select the frequency measures with the best fit. To implement this, we first fit one model per frequency measure (i.e., SUBTLEX, dlexDB, ADE20K, and Greene frequency) separately for the word and the object recognition trials (four frequency measures times two modalities: eight models in total). All models implemented the same covariates and random-effects structure. Then we compared the four models of each modality to a "baseline" model that did not include the frequency measure, but that was estimated on the same subset of data and implemented the same structure of covariates and random effects (2 baseline models in total, one for words and one for objects data). With this procedure, we could estimate the singular fit of each frequency measure in each stimulus modality. From that, we selected the frequency measures that explained a considerable amount of variance in both modalities for further analysis. A better fit was determined by a significant decrease in the AIC, which was tested by implementing the *anova()* function in R. Given the different sources from which word and object frequencies are estimated, they might provide a distinct contribution in representing the occurrence of

objects/words in the world. Therefore, we operated the AIC-based selection following these criteria: in the best case, we would have selected two measures, i.e., the best fitting OF and the best fitting WF measure. In the worst-case, none of the frequency measures would have explained variance in both object and word trials. While, in between, we would have selected either only an OF or a WF measure.

2) After selecting the best frequency measures, we ran a LMM estimating the effects of those selected frequencies on the entire dataset (word trials + object trials), and including all categorical factors and continuous covariates, as well as random factors for participants and concepts.

3) When we detected significant interactions between frequency measures and categorical predictors, we also ran post-hoc LMMs in order to understand the different effects of frequency between different conditions (e.g., SUBTLEX in Cross-modal trials vs. SUBTLEX in Uni-modal trials) and within each condition (e.g., the simple effect of SUBTLEX in Cross-modal trials and simple effect of SUBTLEX in Uni-modal trials). Note that the estimation of frequency effects, given the structure of linear models, was independent (i.e., controlled for) from the effect of the several continuous covariates included in the models.

Supplementary materials 2 – Model selection in Experiment 1

Formula of the models computed in the selection process (1 model x 4 frequency measures x 2 modalities + baseline model without frequency measures x 2 modalities = 10 models):

Exp1_logRT ~ FREQUENCY MEASURE +

Concept category + Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition + Trial accuracy +

(1|Participants) + (1/Concepts)

Supplementary table 3. Summary table of the models included in the selection process. "Frequency" indicates the frequency measure included in the model, where 'Baseline' means no measures included. "AIC" is the criterion used to evaluate the fit of the model. "Modality" indicates which subset of data was considered. "AIC difference" is the difference in AIC between every model and the baseline model of the same modality. More negative differences indicate a better fit of the model including the frequency measure; significant improvements of fit are highlighted in bold.

Frequency	**AIC**	**Modality**	**AIC difference**
Baseline	-1214.816	Objects	0
SUBTLEX WF	-1218.978	Objects	**-4.163**
ADE20K OF	-1212.867	Objects	1.949
Greene OF	-1213.126	Objects	1.690
dlexDB WF	-1214.462	Objects	0.354
Baseline	-1442.289	Words	0
SUBTLEX WF	-1469.443	Words	**-27.153**
ADE20K OF	-1443.517	Words	-1.228
dlexDB WF	-1455.736	Words	**-13.447**
Greene OF	-1441.156	Words	1.133

Supplementary materials 3 – Results of the selected model in Experiment 1

*Exp1_logRT ~ SUBTLEX WF * Concept modality +*

Concept category + Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition +

Trial accuracy + (1|Participants) + (1/Concepts)

Supplementary table 4. Results from the selected model for semantic categorization

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.479	0.021	302.552	**<0.001**
Concept modality (Words – Objects)	0.094	0.005	20.529	**<0.001**
SUBTLEX WF	-0.031	0.007	-4.417	**<0.001**
Visuo-orthographic PC	-0.006	0.007	-0.818	0.413
Concept familiarity	-0.003	0.003	-0.983	0.326
Image typicality	-0.004	0.003	-1.302	0.193
Image visual PC1	-0.002	0.006	-0.316	0.752
Image visual PC2	0.019	0.006	3.253	**0.001**
Image visual PC3	0.008	0.006	1.410	0.159
Target repetition	-0.011	0.002	-4.933	**<0.001**
Trial accuracy (Correct – Incorrect)	-0.017	0.009	-1.936	0.053
Concept category (Natural – Man-made)	0.001	0.012	0.116	0.908
SUBTLEX x (Words – Objects)	-0.019	0.005	-4.160	**<0.001**

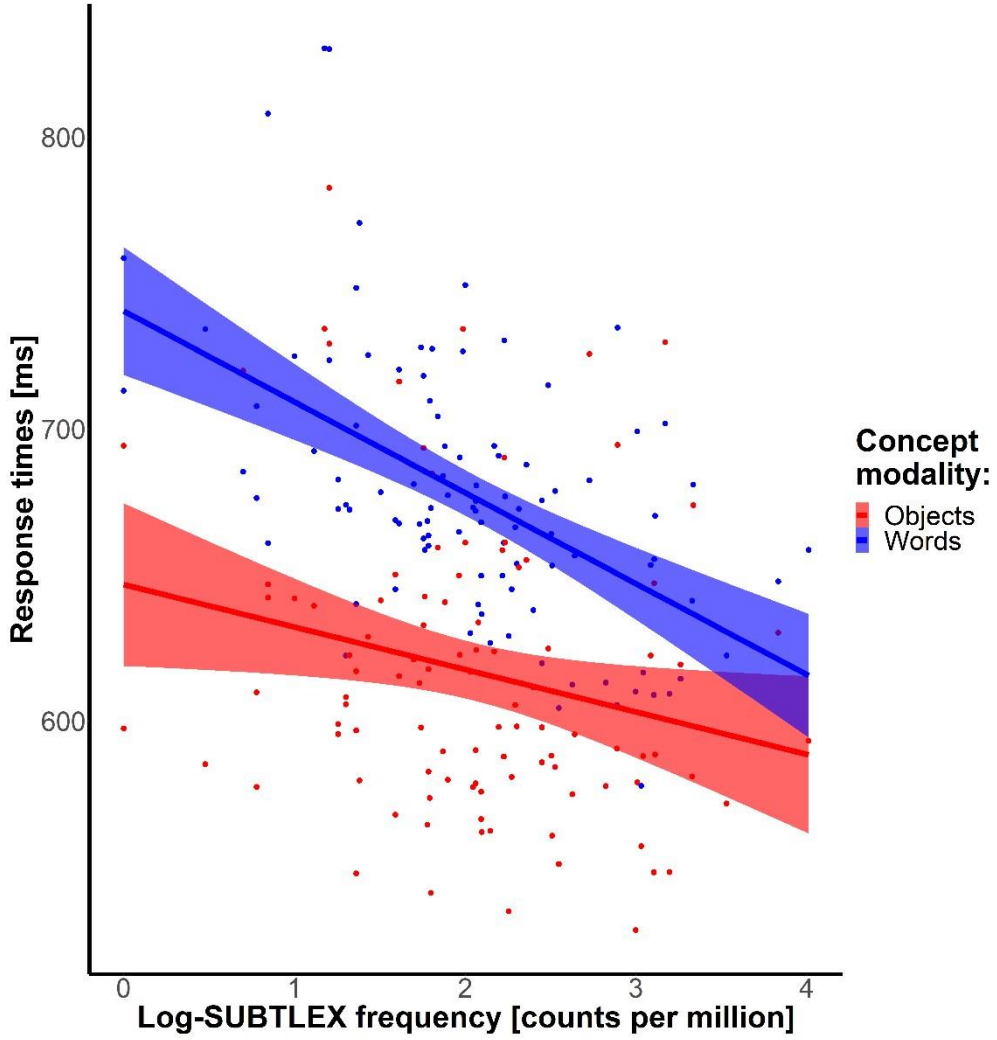
Supplementary table 5. Variance Inflation Factors for the estimated effects of the main model of Experiment 1

Term	VIF
Concept modality (Words – Objects)	1.012
SUBTLEX WF	1.535
Visuo-orthographic PC	1.698
Concept familiarity	1.043
Image typicality	1.015
Image visual PC1	1.032
Image visual PC2	1.027
Image visual PC3	1.063
Trial accuracy (Correct – Incorrect)	1.017
Concept category (Natural – Man-made)	1.180
Concept modality x SUBTLEX	1.000
Target repetition	1.010

The measured SUBTLEX WF effect was independent of visual and visuo-orthographic information of the stimuli, as well as of image typicality, subjective familiarity, concept repetition, concept category and accuracy of categorization.

Supplementary figure 6 – Raw RTs from Experiment 1

Raw response times for object (red) and word (blue) trials as a function of SUBTLEX frequency in Experiment 1. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval



Supplementary materials 4 - Post-hoc of interaction in Experiment 1

2 post-hoc models are estimated, with the same formula, but on 2 different subsest of the data (Object trials and Word trials):

Exp1_logRT ~ SUBTLEX WF +

Concept category + Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition +

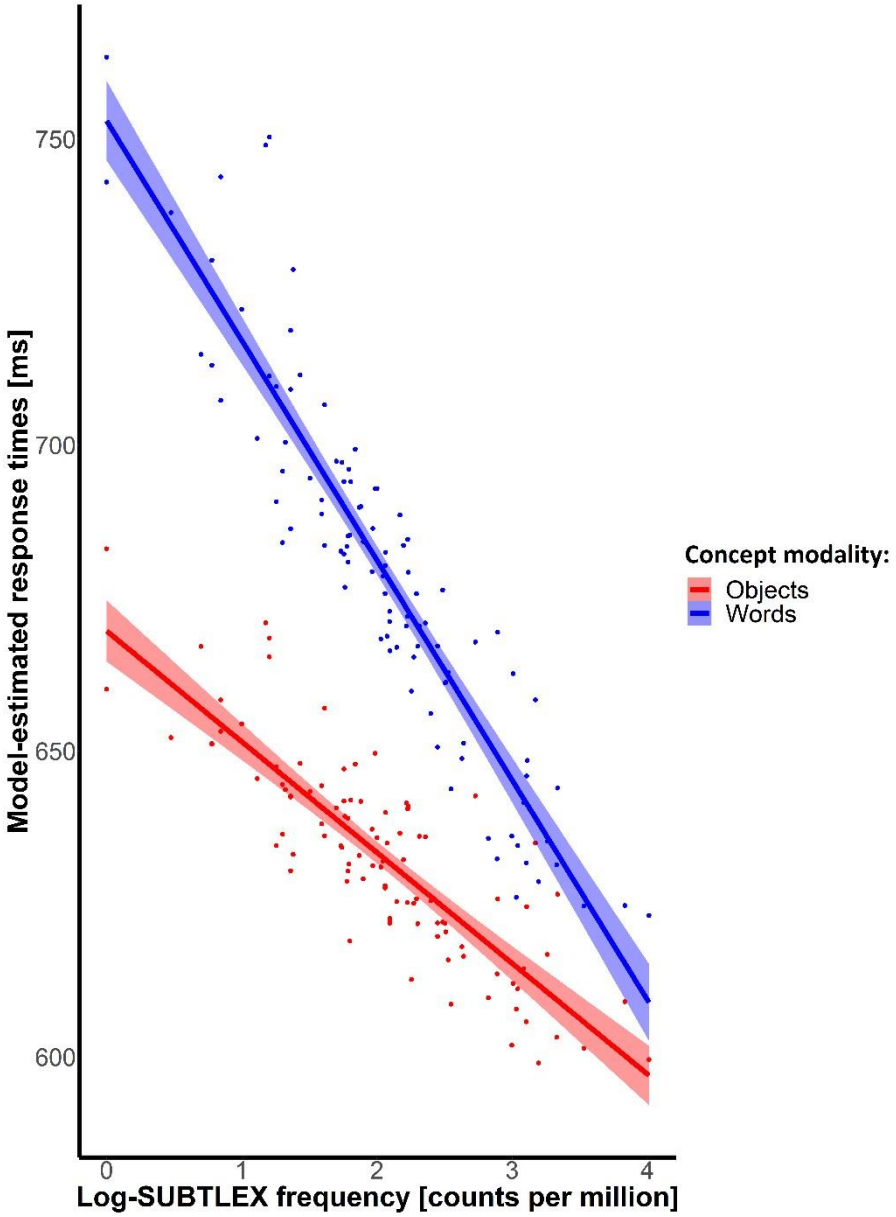
Trial accuracy + (1|Participants) + (1/Concepts)

Supplementary table 6. Results from the post-hoc models for semantic categorization

	Objects				**Words**			
Predictors	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
(Intercept)	6.449	0.022	289.694	**<0.001**	6.520	0.024	266.828	**<0.001**
SUBTLEX WF	-0.022	0.009	-2.524	**0.012**	-0.041	0.007	-5.794	**<0.001**
Concept category	0.009	0.015	0.610	0.542	-0.008	0.012	-0.626	0.531
Visuo-orthographic PC	-0.006	0.009	-0.671	0.502	-0.006	0.007	-0.823	0.411
Concept familiarity	-0.000	0.005	-0.050	0.960	-0.005	0.004	-1.151	0.250
Image typicality	-0.009	0.004	-1.977	**0.048**	-0.001	0.004	-0.258	0.797
Image visual PC1	-0.004	0.007	-0.598	0.550	0.000	0.006	0.069	0.945
Image visual PC2	0.026	0.007	3.581	**<0.001**	0.011	0.006	1.949	0.051
Image visual PC3	0.005	0.007	0.692	0.489	0.012	0.006	2.034	**0.042**
Target repetition	0.009	0.021	0.430	0.667	-0.030	0.024	-1.293	0.196
Trial accuracy	-0.059	0.013	-4.379	**<0.001**	0.005	0.011	0.483	0.629

Supplementary figure 7 – RTs estimated from post-hoc models of Experiment 1

Estimated response times from individual post-hoc models for object (red) and word (blue) trials as a function of SUBTLEX frequency in Experiment 1. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval



Supplementary materials 5 - Results of new ratings on original data of Exp 1

*Exp1_logRT ~ SUBTLEX WF * Concept modality +*

Concept category + Concept familiarity (replication) +

Image typicality (replication) +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition +

Trial accuracy + (1|Participants) + (1/Concepts)

Supplementary table 12. Results from main model of Exp 1 including ratings from replication study

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.479	0.021	303.428	**<0.001**
Concept modality (Words – Objects)	0.094	0.005	20.527	**<0.001**
SUBTLEX WF	-0.035	0.008	-4.690	**<0.001**
Visuo-orthographic PC	-0.009	0.007	-1.148	0.251
Concept familiarity (replication)	0.012	0.007	1.857	0.063
Image typicality (replication)	-0.009	0.006	-1.513	0.130
Image visual PC1	-0.001	0.006	-0.233	0.816
Image visual PC2	0.013	0.006	2.316	**0.021**
Image visual PC3	0.010	0.006	1.741	0.082
Target repetition	-0.011	0.002	-4.933	**<0.001**
Trial accuracy (Correct – Incorrect)	-0.017	0.009	-1.905	0.057
Concept category (Natural – Man-made)	0.012	0.013	0.925	0.355
SUBTLEX x (Words – Objects)	-0.019	0.005	-4.157	**<0.001**

Supplementary Materials 6 - Model selection in Experiment 2

Formula of the models computed in the selection process (1 model x 4 frequency measures x 2 modalities + baseline model without frequency measures x 2 modalities = 10 models):

*Exp2_logRT ~ FREQUENCY MEASURE * Priming condition * Matching condition +*
Concept familiarity + Image typicality +
Image visual PC1 + Image visual PC2 + Image visual PC3 +
Visuo-orthographic PC + Target repetition + Trial accuracy + 379
(1|Participants) + (1/Concepts)

Supplementary table 7. Summary table of the models included in the selection process. "Frequency" indicates the frequency measure included in the model, where 'Baseline' means no measures included. "AIC" is the criterion used to evaluate the fit of the model. "Modality" indicates which subset of data was considered. "AIC difference" is the difference in AIC between every model and the baseline model of the same modality. More negative differences indicate a better fit of the model including the frequency measure; significant improvements of fit are highlighted in bold.

Frequency	AIC	Modality	AIC difference
Baseline	-5150.153	Objects	0
SUBTLEX WF	-5194.848	Objects	**-44.695**
ADE20K OF	-5152.258	Objects	**-2.105**
DlexDB WF	-5175.167	Objects	**-25.013**
Greene OF	-5169.700	Objects	**-19.547**
Baseline	-6366.279	Words	0
SUBTLEX WF	-6401.687	Words	**-35.409**
ADE20K OF	-6385.581	Words	**-19.302**
DlexDB WF	-6377.383	Words	**-11.105**
Greene OF	-6401.694	Words	**-35.415**

Supplementary materials 7 Results selected model in Experiment 2

*Exp2_logRT ~ SUBTLEX WF * Priming condition * Matching condition * Target modality +*

*Greene OF * Priming condition * Matching condition * Target modality +*

Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition + Trial accuracy +

(1|Participants) + (1/Concepts)

Supplementary table 8. Results from the selected model for priming task

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.225	0.020	315.728	<**0.001**
Greene OF	0.008	0.002	4.474	<**0.001**
Matching condition (Mismatch – Match)	0.073	0.002	32.684	<**0.001**
Target modality (Words – Objects)	-0.008	0.002	-3.618	<**0.001**
Priming condition (Cross-modal – Uni-modal)	0.006	0.005	1.173	0.241
SUBTLEX WF	-0.004	0.002	-1.812	0.070
Visuo-orthographic PC	0.010	0.002	4.877	<**0.001**
Concept familiarity	-0.001	0.002	-0.845	0.398
Image typicality	-0.004	0.002	-2.562	**0.010**
Image visual PC1	0.002	0.002	1.217	0.224
Image visual PC2	0.004	0.002	2.788	**0.005**
Image visual PC3	-0.001	0.002	-0.832	0.405
Target repetition	-0.036	0.003	-13.357	<**0.001**
Trial accuracy (Correct – Incorrect)	0.030	0.005	5.444	<**0.001**
Greene x Matching condition	-0.017	0.002	-7.454	<**0.001**
Greene x Target modality	0.003	0.002	1.396	0.163
Matching condition x Target modality	-0.009	0.004	-1.939	0.053
Greene x Priming condition	0.008	0.002	3.347	**0.001**
Matching condition x Priming condition	0.003	0.004	0.737	0.461
Priming condition x Target modality	0.013	0.004	2.872	**0.004**

SUBTLEX x Matching condition	0.021	0.002	9.074	<**0.001**
SUBTLEX x Target modality	-0.005	0.002	-2.378	**0.017**
SUBTLEX x Priming condition	-0.015	0.002	-6.335	<**0.001**
Greene x Matching condition x Target modality	0.003	0.005	0.668	0.504
Greene x Matching condition x Priming condition	-0.020	0.005	-4.256	<**0.001**
Greene x Priming condition x Target modality	0.007	0.005	1.416	0.157
Matching condition x Priming condition x Target modality	0.046	0.009	5.220	<**0.001**
SUBTLEX x Matching condition x Target modality	-0.002	0.005	-0.472	0.637
SUBTLEX x Matching condition x Priming condition	0.017	0.005	3.687	<**0.001**
SUBTLEX x Priming condition x Target modality	0.003	0.005	0.569	0.570
Greene x Matching condition x Priming condition x Target modality	0.002	0.009	0.227	0.821
SUBTLEX x Matching condition x Priming condition x Target modality	0.006	0.009	0.612	0.541

Supplementary table 9. Variance Inflation Factors for the estimated effects of the main model of Experiment 2

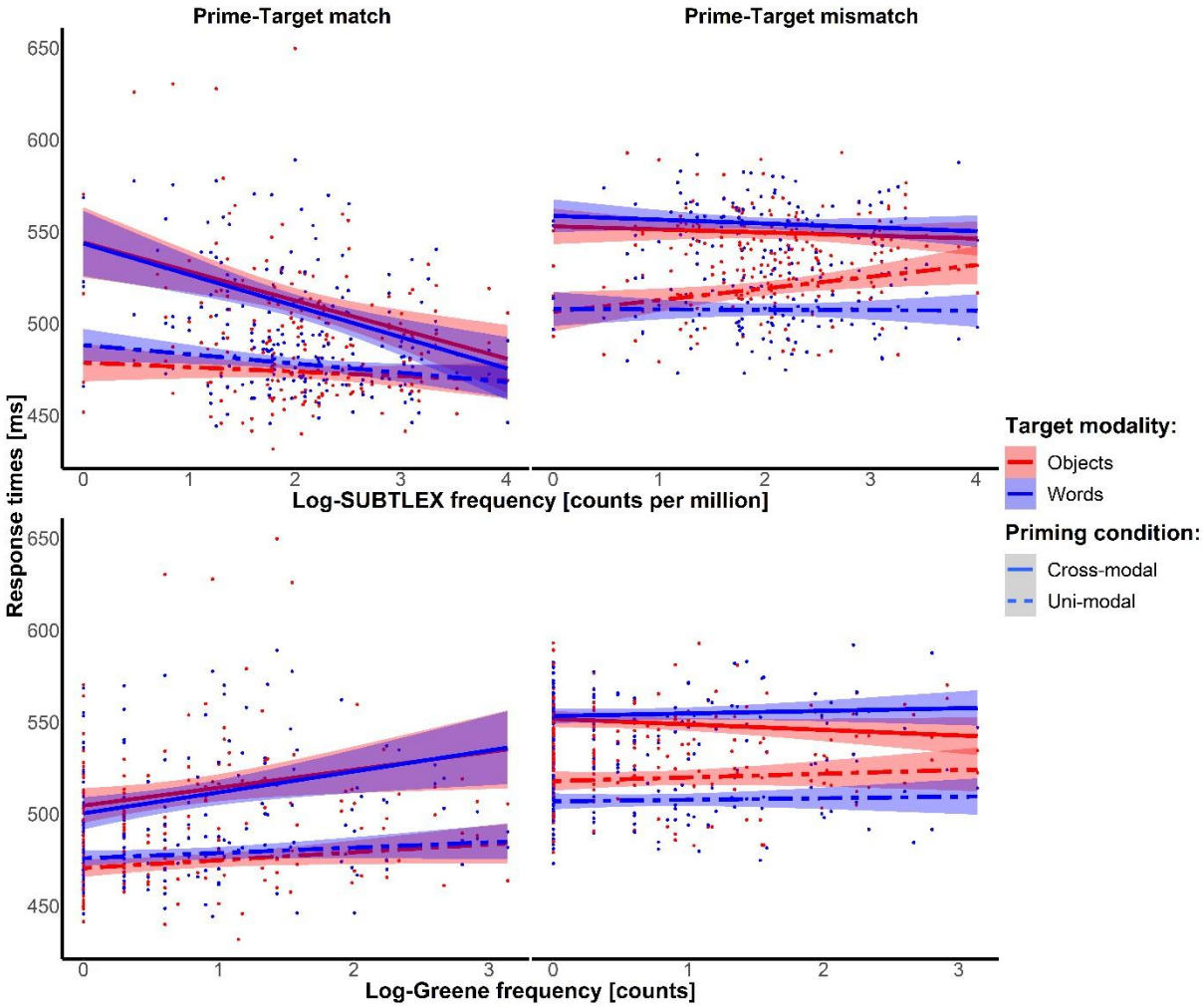
Term	**VIF**
Greene OF	1.190
Matching condition (Mismatch – Match)	1.022
Target modality (Words – Objects)	1.032
Priming condition (Cross-modal – Uni-modal)	5.799
SUBTLEX WF	1.673
Visuo-orthographic PC	1.583
Concept familiarity	1.168
Image typicality	1.037
Image visual PC1	1.026
Image visual PC2	1.026
Image visual PC3	1.069
Trial accuracy (Correct – Incorrect)	1.007
Greene x Matching condition	1.078

Greene x Target modality	1.077
Matching condition x Target modality	1.000
Greene x Priming condition	1.077
Matching condition x Priming condition	1.000
Priming condition x Target modality	1.004
SUBTLEX x Matching condition	1.078
SUBTLEX x Target modality	1.077
SUBTLEX x Priming condition	1.077
Greene x Matching condition x Target modality	1.077
Greene x Matching condition x Priming condition	1.078
Greene x Target modality x Priming condition	1.077
Matching condition x Priming condition x Target modality	1.000
SUBTLEX x Matching condition x Target modality	1.077
SUBTLEX x Matching condition x Priming condition	1.077
SUBTLEX x Target modality x Priming condition	1.077
Greene x Matching condition x Priming condition x Target modality	1.077
SUBTLEX x Matching condition x Priming condition x Target modality	1.077
Target repetition	5.857

The model showed moderate collinearity (VIFs = 5.8 and 5.9) between the Priming condition and Target repetition. This was expected because, despite counterbalancing block order for modalities (word-object or object-word) across participants, all participants performed the cross-modal blocks before the uni-modal blocks (and after Experiment 1). We kept the term for further analysis since collinearity was only just above the threshold for these terms, and because we deemed it important to account for potential carry-over effect. The measured SUBTLEX WF and Greene OF effects were independent of visual and visuo-orthographic information of the stimuli, as well as of image typicality, subjective familiarity, target repetition, and accuracy of categorization.

Supplementary figure 8 - Raw RTs from Experiment 2

Raw response times for object (red) and word (blue) trials in the priming conditions (Cross-modal solid lines, Uni-modal: dashed-dotted) and matching condition (Matching on the left, Mismatching on the right), as a function of SUBTLEX frequency (top) and Greene frequency (bottom) in Experiment 2. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval



Supplementary Materials 8 – Post-hoc of interactions in Experiment 2

Recoded factor is a factor we obtained merging *Priming condition* and *Matching condition* to explore the interaction between frequency x Priming condition x Matching condition. This new factor has 4 levels (*Cross-modal Matching, Uni-modal Matching, Cross-modal Mismatching, Uni-modal*

Mismatching) and 3 contrasts of interest are computed (*Cross-modal Matching – Uni-modal Matching, Cross-modal Mismatching – Uni-modal Mismatching, Cross-modal Matching – Uni-modal Mismatching*)

*Exp2_logRT ~ SUBTLEX WF * Recoded factor * Target modality +*
*Greene OF * Recoded factor * Target modality +*
Concept familiarity + Image typicality +
Image visual PC1 + Image visual PC2 + Image visual PC3 +
Visuo-orthographic PC + Target repetition + Trial accuracy +
(1|Participants) + (1/Concepts)

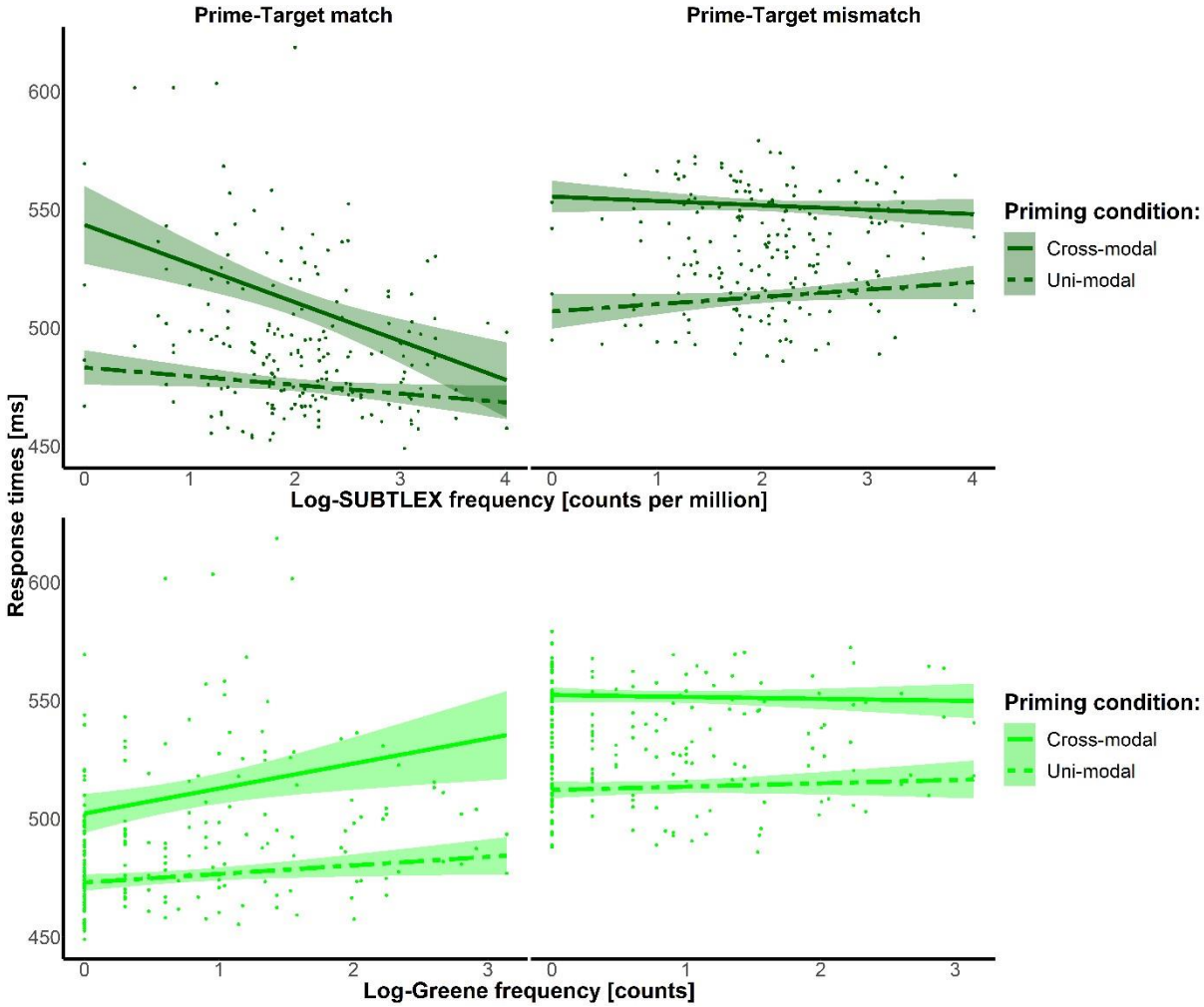
Supplementary table 10. Results from the post-hoc model with re-coded contrasts

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.225	0.020	315.730	<**0.001**
SUBTLEX WF	-0.004	0.002	-1.812	0.070
Cross-modal matching – Uni-modal matching	0.005	0.006	0.805	0.421
Cross-modal mismatching – Uni-modal mismatching	0.008	0.006	1.360	0.174
Cross-modal matching – Cross-modal mismatching	-0.075	0.003	-23.733	<**0.001**
Target modality (Words – Objects)	-0.008	0.002	-3.618	<**0.001**
Greene OF	0.008	0.002	4.474	<**0.001**
Visuo-orthographic PC	0.010	0.002	4.877	<**0.001**
Concept familiarity	-0.001	0.002	-0.845	0.398
Image typicality	-0.004	0.002	-2.562	**0.010**
Image visual PC1	0.002	0.002	1.217	0.224
Image visual PC2	0.004	0.002	2.788	**0.005**
Image visual PC3	-0.001	0.002	-0.832	0.405
Target repetition	-0.036	0.003	-13.357	<**0.001**
Trial accuracy (Correct – Incorrect)	0.030	0.005	5.444	<**0.001**

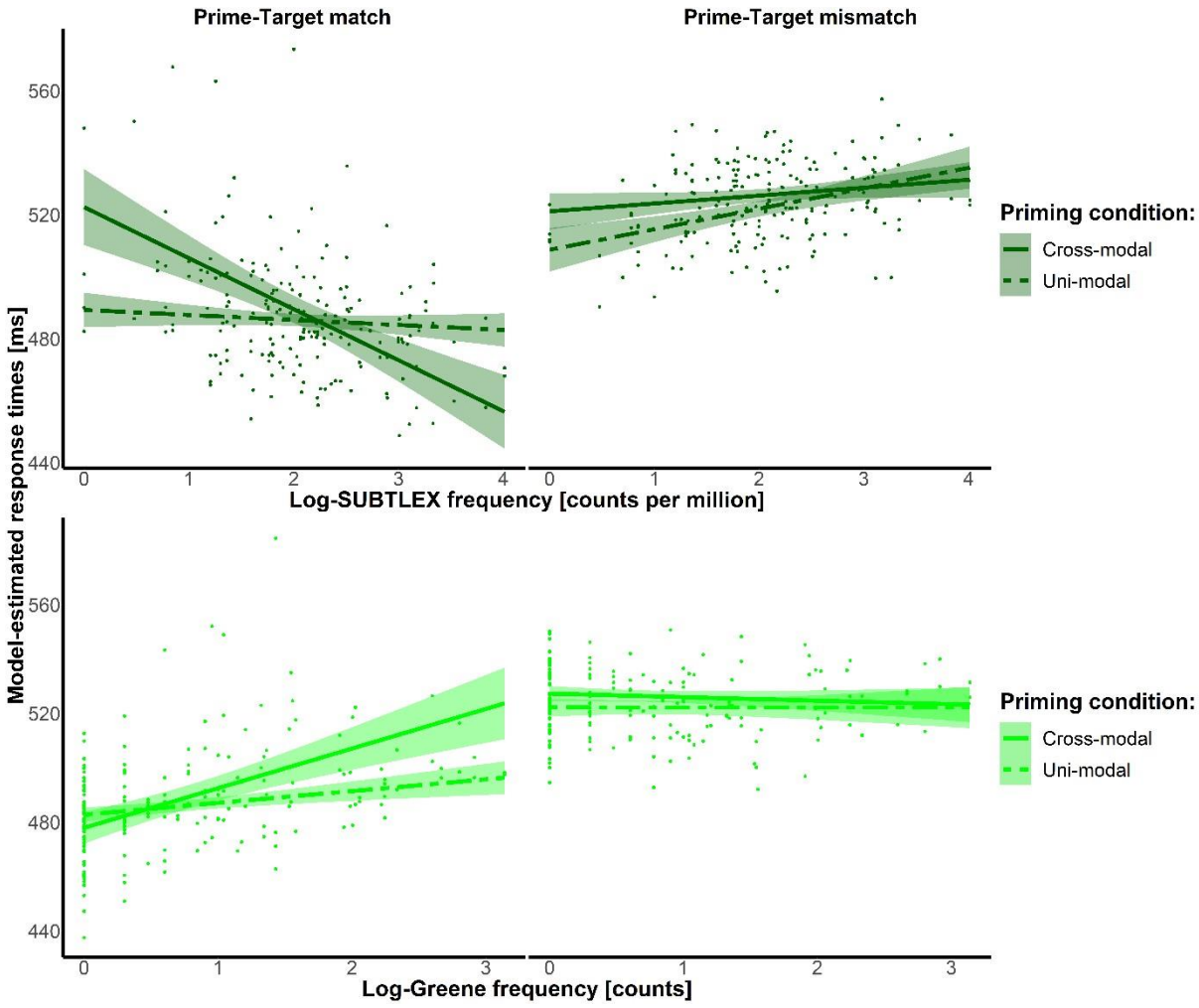
SUBTLEX x (Cross-modal matching – Uni-modal matching)	-0.023	0.003	-7.094	**<0.001**
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching)	-0.006	0.003	-1.870	0.062
SUBTLEX x (Cross-modal matching – Cross-modal mismatching)	-0.029	0.003	-9.027	**<0.001**
SUBTLEX x Target modality	-0.005	0.002	-2.378	**0.017**
(Cross-modal matching – Uni-modal matching) x Target modality	-0.010	0.006	-1.655	0.098
(Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.036	0.006	5.717	**<0.001**
(Cross-modal matching – Cross-modal mismatching) x Target modality	-0.015	0.006	-2.320	**0.020**
Greene x (Cross-modal matching – Uni-modal matching)	0.018	0.003	5.379	**<0.001**
Greene x (Cross-modal mismatching – Uni-modal mismatching)	-0.002	0.003	-0.644	0.520
Greene x (Cross-modal matching – Cross-modal mismatching)	0.027	0.003	8.276	**<0.001**
Greene x Target modality	0.003	0.002	1.396	0.163
SUBTLEX x (Cross-modal matching – Uni-modal matching) x Target modality	-0.000	0.007	-0.031	0.975
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.005	0.007	0.834	0.404
SUBTLEX x (Cross-modal matching – Cross-modal mismatching) x Target modality	-0.001	0.007	-0.099	0.921
Greene x (Cross-modal matching – Uni-modal matching) x Target modality	0.005	0.007	0.842	0.400
Greene x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.008	0.007	1.161	0.246
Greene x (Cross-modal matching – Cross-modal mismatching) x Target modality	-0.004	0.007	-0.632	0.527

Supplementary figure 9. Raw RTs from post-hoc conditions of Experiment 2

Raw response times in the priming conditions (Cross-modal: solid lines, Uni-modal: dashed-dotted) and matching condition (Matching on the left, Mismatching on the right), as a function of SUBTLEX frequency (top, dark green) and Greene frequency (bottom, light green) in Experiment 2. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval



Supplementary figure 10 - RTs estimated from post-hoc of interaction effects of Experiment 2 (between condition)
Response times as a function of logarithmic SUBTLEX frequency (top plots, dark green) and Greene frequency (bottom plots, light green) in the 3-way significant interaction with Matching condition and Priming condition (Cross-modal matching vs. Uni-modal matching; Cross-modal mismatching vs Uni-modal mismatching; Cross-modal matching vs. Cross-modal mismatching). RTs were estimated based on the selected model. Points present participant-based mean response times for concepts in the different frequency levels. Lines represent linear fitting of points (solid: cross-modal; dashed: uni-modal), and shaded areas represent 95 % confidence interval. Bottom left and top-left plots represent the effects in prime-target matching condition, while bottom-right and top-right plots represent the effects in prime-target mismatching condition



4 post-hoc models are additionally computed, one for every level of the re-coded factor (Cross-modal Matching, Uni-modal Matching, Cross-modal Mismatching, Uni-modal Mismatching)

*Exp2_logRT ~ SUBTLEX WF * Target modality + Greene OF * Target modality +*

Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition + Trial accuracy +

(1|Participants) + (1/Concepts)

Supplementary table 11. Results from the post-hoc individual models for conditions of interest

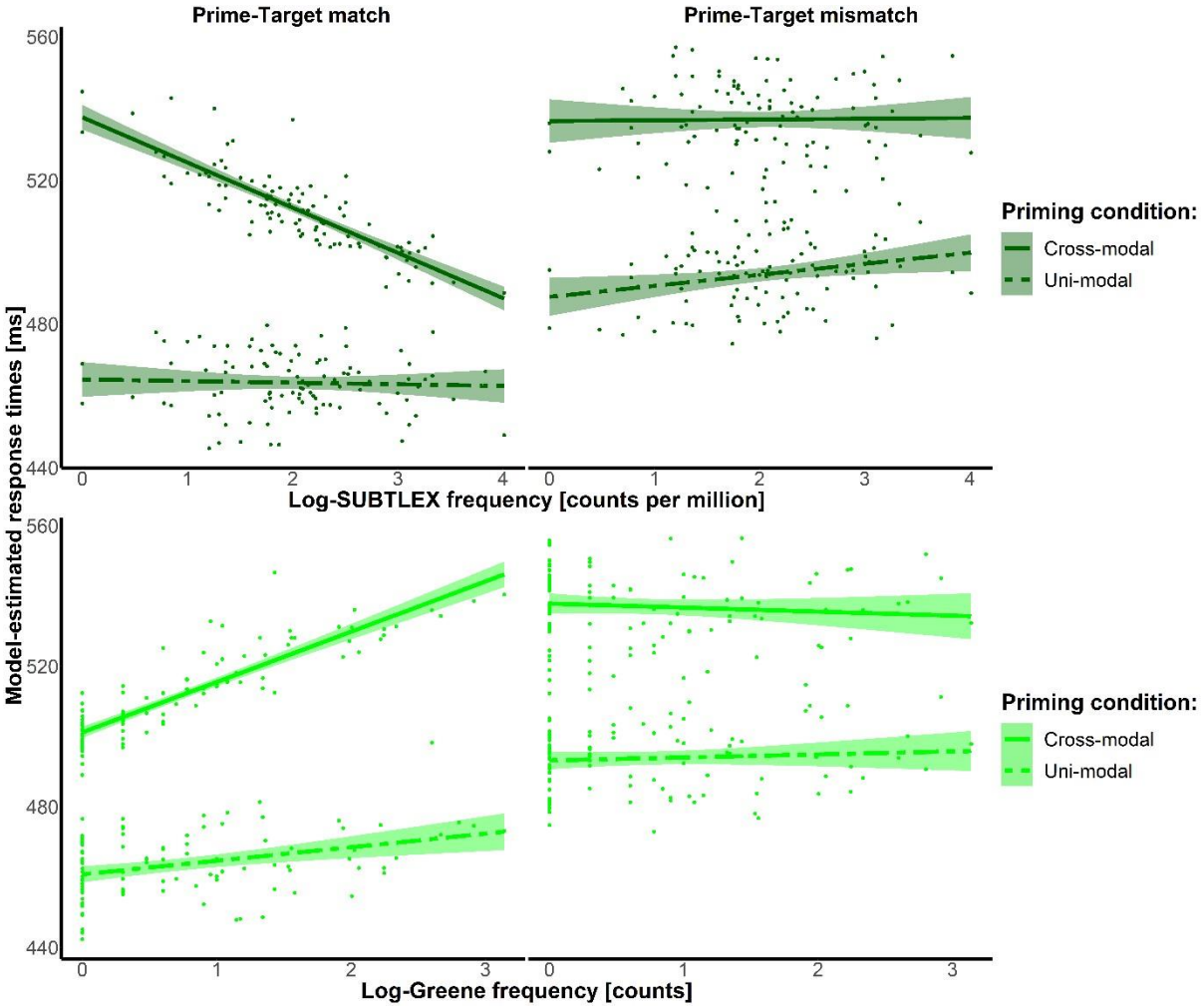
	Uni-modal Matching				**Cross-modal Matching**			
Predictors	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
(Intercept)	6.139	0.020	309.972	**<0.001**	6.238	0.021	290.842	**<0.001**
SUBTLEX WF	-0.001	0.003	-0.245	0.807	-0.019	0.006	-3.230	**0.001**
Target modality (Words – Objects)	0.008	0.004	1.790	0.073	-0.012	0.005	-2.620	**0.009**
Greene OF	0.007	0.003	2.725	**0.006**	0.023	0.005	4.710	**<0.001**
Visuo-orthographic PC	0.013	0.003	4.418	**<0.001**	0.018	0.006	3.077	**0.002**
Concept familiarity	-0.000	0.003	-0.150	0.881	-0.003	0.003	-0.858	0.391
Image typicality	-0.005	0.003	-1.835	0.067	-0.013	0.003	-3.825	**<0.001**
Image visual PC1	0.003	0.002	1.289	0.197	0.003	0.005	0.593	0.553
Image visual PC2	0.004	0.002	1.758	0.079	0.001	0.005	0.254	0.799
Image visual PC3	-0.003	0.002	-1.173	0.241	0.001	0.005	0.178	0.859
Target repetition	-0.002	0.002	-0.792	0.428	-0.028	0.002	-12.095	**<0.001**
Trial accuracy (Correct – Incorrect)	0.058	0.010	6.001	**<0.001**	-0.008	0.010	-0.760	0.448
SUBTLEX x (Words – Objects)	-0.004	0.004	-0.982	0.326	-0.005	0.005	-0.948	0.343
Greene x (Words – Objects)	-0.001	0.004	-0.304	0.761	0.005	0.005	1.012	0.311

	Uni-modal Mismatching				**Cross-modal Mismatching**			
Predictors	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
(Intercept)	6.202	0.020	312.978	**<0.001**	6.286	0.022	279.822	**<0.001**
SUBTLEX WF	0.005	0.003	1.535	0.125	0.000	0.003	0.108	0.914
Target modality (Words – Objects)	-0.026	0.004	-5.943	**<0.001**	0.004	0.004	0.863	0.388
Greene OF	0.001	0.003	0.543	0.587	-0.002	0.002	-0.733	0.463
Visuo-orthographic PC	0.001	0.003	0.407	0.684	0.005	0.003	1.978	**0.048**
Concept familiarity	-0.003	0.003	-1.225	0.221	0.002	0.003	0.586	0.558
Image typicality	0.001	0.003	0.221	0.825	0.003	0.003	1.109	0.268
Image visual PC1	0.000	0.002	0.024	0.981	0.002	0.002	0.751	0.452
Image visual PC2	0.008	0.002	3.082	**0.002**	0.005	0.002	2.194	**0.028**
Image visual PC3	-0.001	0.003	-0.322	0.748	-0.003	0.002	-1.172	0.241
Target repetition	-0.007	0.002	-3.341	**0.001**	-0.023	0.002	-10.344	**<0.001**
Trial accuracy (Correct – Incorrect)	0.083	0.012	6.712	**<0.001**	0.063	0.012	5.361	**<0.001**

SUBTLEX x (Words – Objects)	-0.009 0.004	-2.113 **0.035**	-0.004 0.005	-0.951 0.342
Greene x (Words – Objects)	0.001 0.004	0.226 0.821	0.009 0.005	2.027 **0.043**

Supplementary figure 11 - RTs estimated from post-hoc models of Experiment 2 (within conditions)

Effects of SUBTLEX WF (dark green, top) and Greene OF (light green, bottom) on reaction times estimated from the post-hoc models separately for each Priming condition (continuous and dashed-dotted line types) and Matching condition (left and right plots). Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval.



Supplementary materials 9 – Results of new ratings on original data of Exp 2

*Exp2_logRT ~ SUBTLEX WF * Priming condition * Matching condition * Target modality +*
*Greene OF * Priming condition * Matching condition * Target modality +*
Concept familiarity (replication) + Image typicality (replication) +
Image visual PC1 + Image visual PC2 + Image visual PC3 +
Visuo-orthographic PC + Target repetition + Trial accuracy +
(1|Participants) + (1/Concepts)

Supplementary table 13. Results from main model of Exp 2 including ratings from replication study

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.225	0.020	315.766	**<0.001**
Greene OF	0.005	0.002	2.978	**0.003**
Matching condition (Mismatch – Match)	0.073	0.002	32.683	**<0.001**
Target modality (Words – Objects)	-0.008	0.002	-3.613	**<0.001**
Priming condition (Cross-modal – Uni-modal)	0.006	0.005	1.175	0.240
SUBTLEX WF	-0.001	0.002	-0.576	0.565
Visuo-orthographic PC	0.010	0.002	5.485	**<0.001**
Concept familiarity (replication)	-0.002	0.002	-0.795	0.427
Image typicality (replication)	-0.008	0.002	-4.829	**<0.001**
Image visual PC1	0.003	0.002	2.077	**0.038**
Image visual PC2	0.003	0.002	2.173	**0.030**
Image visual PC3	-0.002	0.002	-1.518	0.129
Target repetition	-0.036	0.003	-13.355	**<0.001**
Trial accuracy (Correct – Incorrect)	0.030	0.005	5.453	**<0.001**
Greene x Matching condition	-0.017	0.002	-7.458	**<0.001**
Greene x Target modality	0.003	0.002	1.396	0.163
Matching condition x Target modality	-0.009	0.004	-1.938	0.053
Greene x Priming condition	0.008	0.002	3.347	**0.001**
Matching condition x Priming condition	0.003	0.004	0.736	0.462
Priming condition x Target modality	0.013	0.004	2.874	**0.004**
SUBTLEX x Matching condition	0.021	0.002	9.075	**<0.001**
SUBTLEX x Target modality	-0.005	0.002	-2.376	**0.018**
SUBTLEX x Priming condition	-0.015	0.002	-6.335	**<0.001**
Greene x Matching condition x Target modality	0.003	0.005	0.666	0.505

Greene x Matching condition x Priming condition	-0.020 0.005	-4.257	<**0.001**
Greene x Priming condition x Target modality	0.007 0.005	1.412	0.158
Matching condition x Priming condition x Target modality	0.046 0.009	5.216	<**0.001**
SUBTLEX x Matching condition x Target modality	-0.002 0.005	-0.473	0.636
SUBTLEX x Matching condition x Priming condition	0.017 0.005	3.687	<**0.001**
SUBTLEX x Priming condition x Target modality	0.003 0.005	0.569	0.569
Greene x Matching condition x Priming condition x Target modality	0.002 0.009	0.228	0.819
SUBTLEX x Matching condition x Priming condition x Target modality	0.006 0.009	0.612	0.540

Supplementary materials 10 – Exploratory analysis in Experiment 2

Since we detected similar frequency effects in both word and object modalities (i.e., no significant interaction including target modality), as well as the presence of these effects only when semantic processing is required (cross-modal trials), we decided to further explore whether the frequency effects found in Experiment 2 represented common semantic processing of objects and words. Thus, we restricted this exploratory analysis to the *Cross-modal matching trials*. Response times from this subset showed substantial word and object frequency effects and, at the same time, included predictive semantic processing to a high degree.

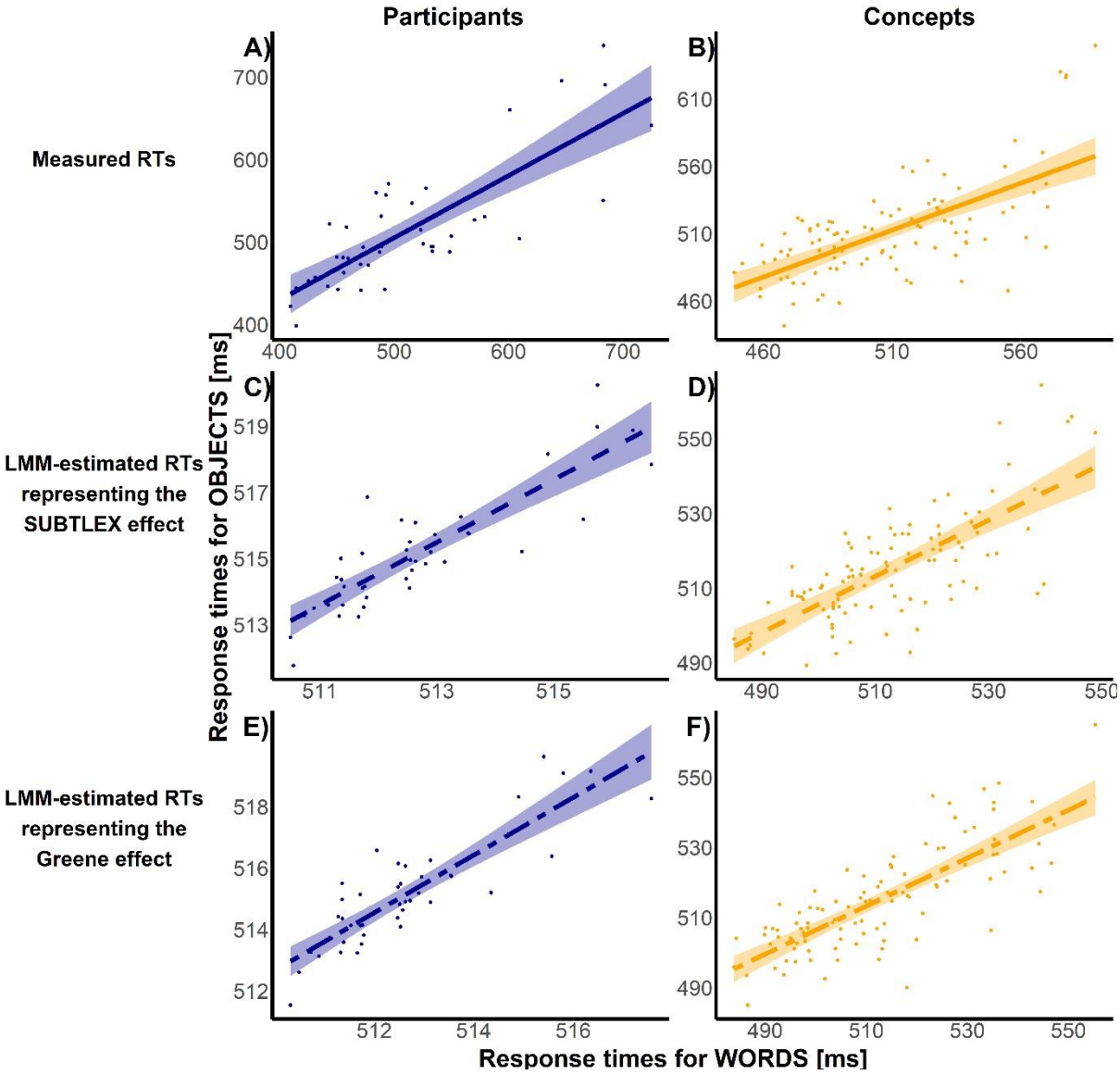
We considered three sources of data: 1) the actual response times from cross-modal matching trials of words and objects; 2) response times estimated from the effect of SUBTLEX WF in cross-modal matching trials of words and objects; 3) response times estimated from the effect of Greene OF in cross-modal matching trials of words and objects. 2) and 3) were estimated using two models (one for word trials and one for object trials) that included SUBTLEX WF, Greene OF, and all the covariates and random effects of the main model of Experiment 2 introduced before. The comparison between word and object processing was made for each of the three datasets considering response times for every participant and for every concept in both modalities. To test the similarity between frequency effects in words and objects, we implemented *paired-samples equivalence tests* and *product-moment correlation tests* between response times from word and object trials.

With the equivalence test, we can check if two samples/conditions come from the same distribution (i.e., they are equivalent). Thus, we computed test statistics for which low probability values allow us to reject the null hypothesis of statistical difference (instead of rejecting the null hypothesis of statistical equivalence of commonly used t-test). For the equivalence test, we needed to set an epsilon parameter, i.e., the maximally allowed difference to consider two conditions non-different; in our case, we used 50 % of the standard deviation of the difference between object and word trials (Robinson & Froese, 2004). The correlation of word-based vs. object-based reaction times could additionally prove whether associated entries show similar behavior.

For actual response time data, we found a significant equivalence (mean of differences = 0.005 log(ms), ε=0.045 log(ms), CI = [-0.018 0.028], p=0.003) and correlation (r=0.804, t(40)=8.554, p<0.001), between *participants' performance* in object and word trials; also, we found a significant equivalence (mean of differences = 0.006 log(ms), ε=0.028 log(ms), CI = [-0.004 0.015], p<0.001) and correlation (r=0.652, t(98)=8.503, p<0.001) between *processing of concepts* in the two different modalities (*Supplementary Figure 12A-B*). That implies high interrelation between the processing objects and words which becomes evident when comparing participants and comparing stimuli with the same semantics.

Supplementary figure 12 - Linear relationship between object trials and word trials in Cross-modal matching trials

Correlations among unique participants (dark blue: A, C, and E) and unique concepts (orange: B, D and F). A, B) Actual response times for Cross-modal matching trials (solid lines). C, D) Response times estimated from SUBTLEX WF effect in Cross-modal matching trials (dashed lines); E, F) Response times estimated from Greene OF effect in Cross-modal matching trials (dashed-dotted lines). Points represent performance of individual participants or concepts in the two tasks. Lines represent linear fitting of points, and shaded areas represent 95 % confidence interval.



To get an estimate to which degree the interrelation was driven by the WF effect, we predicted RTs that were influenced by the SUBTLEX WF effect without confounds based on the estimated models (one for cross-modal matching trials of objects, one for cross-modal matching trials for words). For *participants' performance*, we could not reject statistical difference (mean of differences = 0.0049 log(ms), ε=0.0009 log(ms), CI = [0.0044 0.0053], p=1), but we found a significant correlation (r=0.860, t(40)=10.667, p<0.001) between object trials and word trials; similarly, but *considering single concepts*, we could reject statistical difference (mean of differences = 0.005 log(ms), ε=0.010 log(ms), CI = [0.001 0.008],

p=0.004), and we found also a significant correlation (r=0.717, t(98)=10.186, p<0.001) (*Supplementary Figure 12C-D*).

We repeated the same procedure for the Greene OF effect and found the same pattern: statistical difference could not be reject for individual participants (mean of differences = 0.0048 log(ms), ε=0.0009 log(ms), CI = [0.0044 0.0053], p=1), but it was rejected for individual concepts performance (mean of differences = 0.005 log(ms), ε=0.010 log(ms), CI = [0.001 0.008], p<0.001), while both showed a strong correlation between object and word trials (participants: r=0.868, t(40)=11.067, p<0.001; concepts: r=0.779, t(98)=12.29, p<0.001) (*Supplementary Figure 12E-F*). Overall, these exploratory analyses show that even if the WF and OF do not affect object and word processing completely identically, the individual participant's frequency effects for words and object and the frequency effects for single semantic concepts are strongly associated to each other across modalities.

Supplementary Materials 11 – Effect of Greene OF with Conceptual Distinctiveness

We can draw parallels between the experimental visual experience created and tested by Konkle and colleagues (2010) (i.e., manipulating the frequency of visually presented objects in the lab) and what the Greene OF used in our study aims to represent (i.e., the frequency of visually encountered objects in the real world). This comparison might raise some concerns since the two studies seem relatively different at first sight: First, Konkle et al. artificially induced memory interference and second, they specifically measured visual LTM. That said, we believe that Konkle et al. (2010) of course aimed at measuring a phenomenon of memory interference that they think is happening intrinsically when encountering objects in the world. While Konkle's task required retrieving specific exemplars, our task required retrieving a concept (i.e., the prime meaning) from memory. For both tasks, however, interferences from other exemplars are similarly possible. In addition, to correctly perform the Cross-modal priming task in our study, participants in our study had to access representations in semantic long-term memory (LTM), which was also the locus of the memory interferences as highlighted by Konkle et al. (2010). This is in line with the observation that the Greene OF effect in our study only came into play in Cross-modal Matching trials, where semantic processing and LTM involvement was particularly high.

Recoded factor is a factor we obtained merging *Priming condition* and *Matching condition* to explore interaction between frequency x Priming condition x Matching condition. This new factor has 4 levels

(*Cross-modal Matching, Uni-modal Matching, Cross-modal Mismatching, Uni-modal Mismatching*) and 3 contrasts of interest are computed (*Cross-modal Matching – Uni-modal Matching, Cross-modal Mismatching – Uni-modal Mismatching, Cross-modal Matching – Uni-modal Mismatching*)

*Exp2_logRT ~ SUBTLEX WF * Recoded factor * Target modality +*
*Greene OF * Conceptual Distinctiveness * Recoded factor * Target modality+*
Concept familiarity + Image typicality +
Image visual PC1 + Image visual PC2 + Image visual PC3 +
Visuo-orthographic PC + Target repetition + Trial accuracy +
(1|Participants) + (1/Concepts)

Supplementary table 14. Results from model including Conceptual Distinctiveness in interaction with Greene frequency.

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.225	0.020	315.564	**<0.001**
Conceptual Distinctiveness (CD)	0.002	0.002	0.799	0.424
Greene OF	0.008	0.002	3.709	**<0.001**
Cross-modal matching – Uni-modal matching	0.009	0.006	1.533	0.125
Cross-modal mismatching – Uni-modal mismatching	0.007	0.006	1.235	0.217
Cross-modal matching – Cross-modal mismatching	-0.070	0.003	-20.276	**<0.001**
Target modality (Words – Objects)	-0.008	0.002	-3.356	**0.001**
SUBTLEX	-0.004	0.002	-1.923	0.055
Visuo-orthographic PC	0.009	0.002	4.807	**<0.001**
Concept familiarity	-0.001	0.002	-0.762	0.446
Image typicality	-0.004	0.002	-2.440	**0.015**
Image visual PC1	0.002	0.002	1.266	0.206
Image visual PC2	0.005	0.002	2.776	**0.006**
Image visual PC3	-0.002	0.002	-0.976	0.329
Target repetition	-0.036	0.003	-13.333	**<0.001**
Trial accuracy (Correct – Incorrect)	0.031	0.005	5.595	**<0.001**
Conceptual Distinctiveness (CD)x Greene	-0.001	0.002	-0.749	0.454
CD x (Cross-modal matching – Uni-modal matching)	0.004	0.004	1.051	0.293
CD x (Cross-modal mismatching – Uni-modal mismatching)	0.001	0.004	0.246	0.806
CD x (Cross-modal matching – Cross-modal mismatching)	0.007	0.004	1.937	0.053
Greene x (Cross-modal matching – Uni-modal matching)	0.021	0.004	5.364	**<0.001**
Greene x (Cross-modal mismatching – Uni-modal mismatching)	-0.003	0.004	-0.810	0.418
Greene x (Cross-modal matching – Cross-modal mismatching)	0.030	0.004	7.737	**<0.001**
Conceptual Distinctiveness (CD)x Target modality	-0.000	0.003	-0.065	0.948
Greene x Target modality	0.003	0.003	1.155	0.248
Target modality x (Cross-modal matching – Uni-modal matching)	-0.009	0.007	-1.252	0.211
Target modality x (Cross-modal mismatching – Uni-modal mismatching)	0.030	0.007	4.432	**<0.001**

Target modality x (Cross-modal matching – Cross-modal mismatching)	-0.013 0.007 -1.872	0.061
SUBTLEX x (Cross-modal matching – Uni-modal matching)	-0.023 0.004 -6.500	**<0.001**
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching)	-0.007 0.004 -1.873	0.061
SUBTLEX x (Cross-modal matching – Cross-modal mismatching)	-0.030 0.004 -8.512	**<0.001**
SUBTLEX x Target modality	-0.005 0.003 -2.181	**0.029**
CD x Greene x (Cross-modal matching – Uni-modal matching)	-0.010 0.003 -3.139	**0.002**
CD x Greene x (Cross-modal mismatching – Uni-modal mismatching)	0.002 0.003 0.495	0.621
CD x Greene x (Cross-modal matching – Cross-modal mismatching)	-0.012 0.003 -3.774	**<0.001**
CD x Greene x Target modality	0.000 0.002 0.086	0.932
CD x (Cross-modal matching – Uni-modal matching) x Target modality	-0.001 0.008 -0.072	0.942
CD x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.001 0.008 0.129	0.898
CD x (Cross-modal matching – Cross-modal mismatching) x Target modality	-0.002 0.008 -0.275	0.784
Greene x (Cross-modal matching – Uni-modal matching) x Target modality	0.008 0.008 0.984	0.325
Greene x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.001 0.008 0.143	0.886
Greene x (Cross-modal matching – Cross-modal mismatching) x Target modality	-0.002 0.008 -0.195	0.845
SUBTLEX x (Cross-modal matching – Uni-modal matching) x Target modality	0.001 0.007 0.078	0.938
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.003 0.007 0.464	0.643
SUBTLEX x (Cross-modal matching – Cross-modal mismatching) x Target modality	0.001 0.007 0.083	0.934
CD x Greene x (Cross-modal matching – Uni-modal matching) x Target modality	-0.004 0.006 -0.666	0.506
CD x Greene x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.013 0.006 2.031	**0.042**
CD x Greene x (Cross-modal matching – Cross-modal mismatching) x Target modality	-0.004 0.006 -0.631	0.528

Supplementary materials 12 – Results main model in priming task (Exp 3)

*Exp3_logRT ~ SUBTLEX WF * Priming condition * Matching condition * Target modality +*

*Greene OF * Priming condition * Matching condition * Target modality +*

Concept familiarity (replication) + Image typicality (replication) +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition + Trial accuracy +

(1|Participants) + (1/Concepts)

Supplementary table 15. Results from the main model for priming task replication

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.354	0.017	374.088	<**0.001**
Greene OF	0.003	0.002	1.661	0.097
Matching condition (Mismatch – Match)	0.062	0.002	27.489	<**0.001**
Target modality (Words – Objects)	-0.002	0.002	-0.896	0.370
Priming condition (Cross-modal – Uni-modal)	0.014	0.033	0.413	0.680
SUBTLEX WF	-0.008	0.002	-3.932	<**0.001**
Visuo-orthographic PC	0.005	0.002	2.489	**0.013**
Concept familiarity (replication)	-0.001	0.002	-0.347	0.729
Image typicality (replication)	-0.009	0.002	-5.229	<**0.001**
Image visual PC1	0.001	0.002	0.633	0.527
Image visual PC2	0.003	0.002	1.929	0.054
Image visual PC3	0.001	0.002	0.592	0.554
Target repetition	-0.036	0.001	-31.062	<**0.001**
Trial accuracy (Correct – Incorrect)	0.013	0.007	1.954	0.051
Greene x Matching condition	-0.008	0.002	-3.529	<**0.001**
Greene x Target modality	0.001	0.002	0.371	0.711
Matching condition x Target modality	0.001	0.004	0.195	0.845
Greene x Priming condition	0.012	0.002	5.158	<**0.001**
Priming condition x Matching condition	-0.016	0.004	-3.523	<**0.001**
Priming condition x Target modality	-0.029	0.005	-6.285	<**0.001**
SUBTLEX x Matching condition	0.013	0.002	5.638	<**0.001**
SUBTLEX x Target modality	-0.011	0.002	-4.830	<**0.001**
SUBTLEX x Priming condition	-0.010	0.002	-4.291	<**0.001**
Greene x Matching condition x Target modality	-0.004	0.005	-0.917	0.359
Greene x Matching condition x Priming condition	-0.010	0.005	-2.165	**0.030**
Greene x Priming condition x Target modality	0.000	0.005	0.019	0.985
Matching condition x Priming condition x Target modality	0.030	0.009	3.406	**0.001**

SUBTLEX x Matching condition x Target modality	0.016	0.005	3.399	**0.001**
SUBTLEX x Matching condition x Priming condition	0.009	0.005	1.880	0.060
SUBTLEX x Priming condition x Target modality	-0.006	0.005	-1.230	0.219
Greene x Matching condition x Priming condition x Target modality	-0.011	0.009	-1.169	0.242
SUBTLEX x Matching condition x Priming condition x Target modality	0.021	0.009	2.269	**0.023**

Supplementary table 16. Variance Inflation Factors for the effects of the main model of Replication experiment.

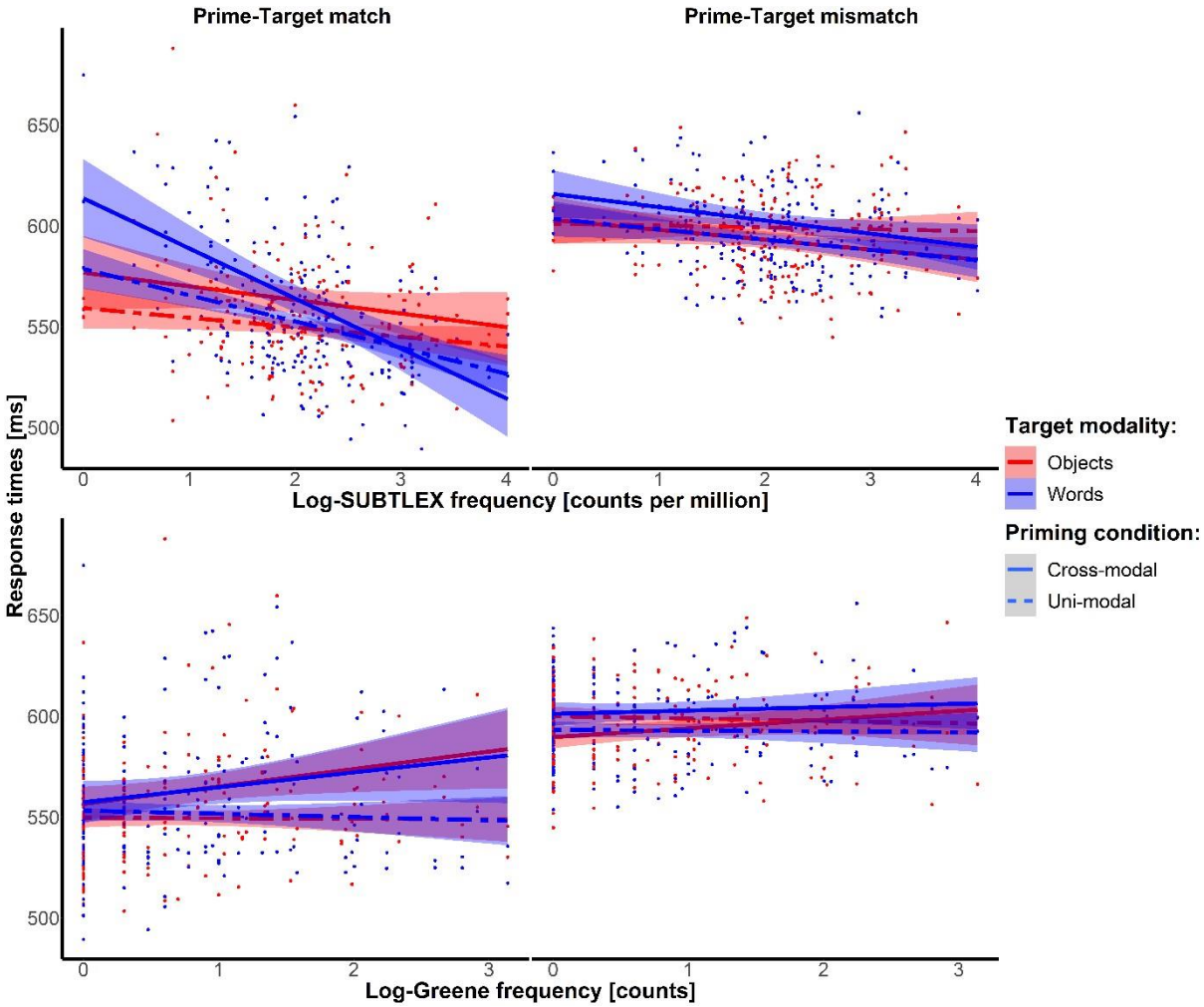
Term	**VIF**
Greene OF	1.535
Matching condition	1.022
Target modality	1.003
Priming condition	1.000
SUBTLEX WF	1.974
Visuo-orthographic PC	1.725
Concept familiarity (replication)	1.705
Image typicality (replication)	1.337
Image visual PC1	1.084
Image visual PC2	1.093
Image visual PC3	1.206
Target repetition	1.097
Trial accuracy	1.006
Greene x Matching condition	1.076
Greene x Target modality	1.076
Matching condition x Priming condition	1.000
Greene x Priming condition	1.076
Matching condition x Target modality	1.000
Priming condition x Target modality	1.077
SUBTLEX x Matching condition	1.077
SUBTLEX x Target modality	1.076
SUBTLEX x Priming condition	1.076

Greene x Matching condition x Target modality	1.076
Greene x Matching condition x Priming condition	1.076
Greene x Priming condition x Target modality	1.076
Matching condition x Priming condition x Target modality	1.000
SUBTLEX x Matching condition x Target modality	1.076
SUBTLEX x Matching condition x Priming condition	1.076
SUBTLEX x Priming condition x Target modality	1.076
Greene x Matching condition x Priming condition x Target modality	1.076
SUBTLEX x Matching condition x Priming condition x Target modality	1.076

The measured SUBTLEX WF and Greene OF effects were independent of visual and visuo-orthographic information of the stimuli, as well as of image typicality, subjective familiarity, target repetition and accuracy of categorization.

Supplementary figure 13 - Raw RTs from Experiment 3

Raw response times for object (red) and word (blue) trials in the priming conditions (Cross-modal solid lines, Uni-modal: dashed-dotted) and matching condition (Matching on the left, Mismatching on the right), as a function of SUBTLEX frequency (top) and Greene frequency (bottom) in the Replication experiment. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval



Supplementary Materials 13 – Post-hoc of interactions in priming task (Exp 3)

Recoded factor is a factor we obtained merging *Priming condition* and *Matching condition* to explore the interaction between frequency x Prining condition x Matching condition. This new factor has 4 levels (*Cross-modal Matching, Uni-modal Matching, Cross-modal Mismatching, Uni-modal Mismatching*) and 3 contrasts of interest are computed (*Cross-modal Matching – Uni-modal Matching, Cross-modal Mismatching – Uni-modal Mismatching, Cross-modal Matching – Uni-modal Mismatching*)

*Exp3_logRT ~ SUBTLEX WF * Recoded factor * Target modality +*

*Greene OF * Recoded factor * Target modality +*

Concept familiarity (replication) + Image typicality (replication) +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition + Trial accuracy +

(1|Participants) + (1/Concepts)

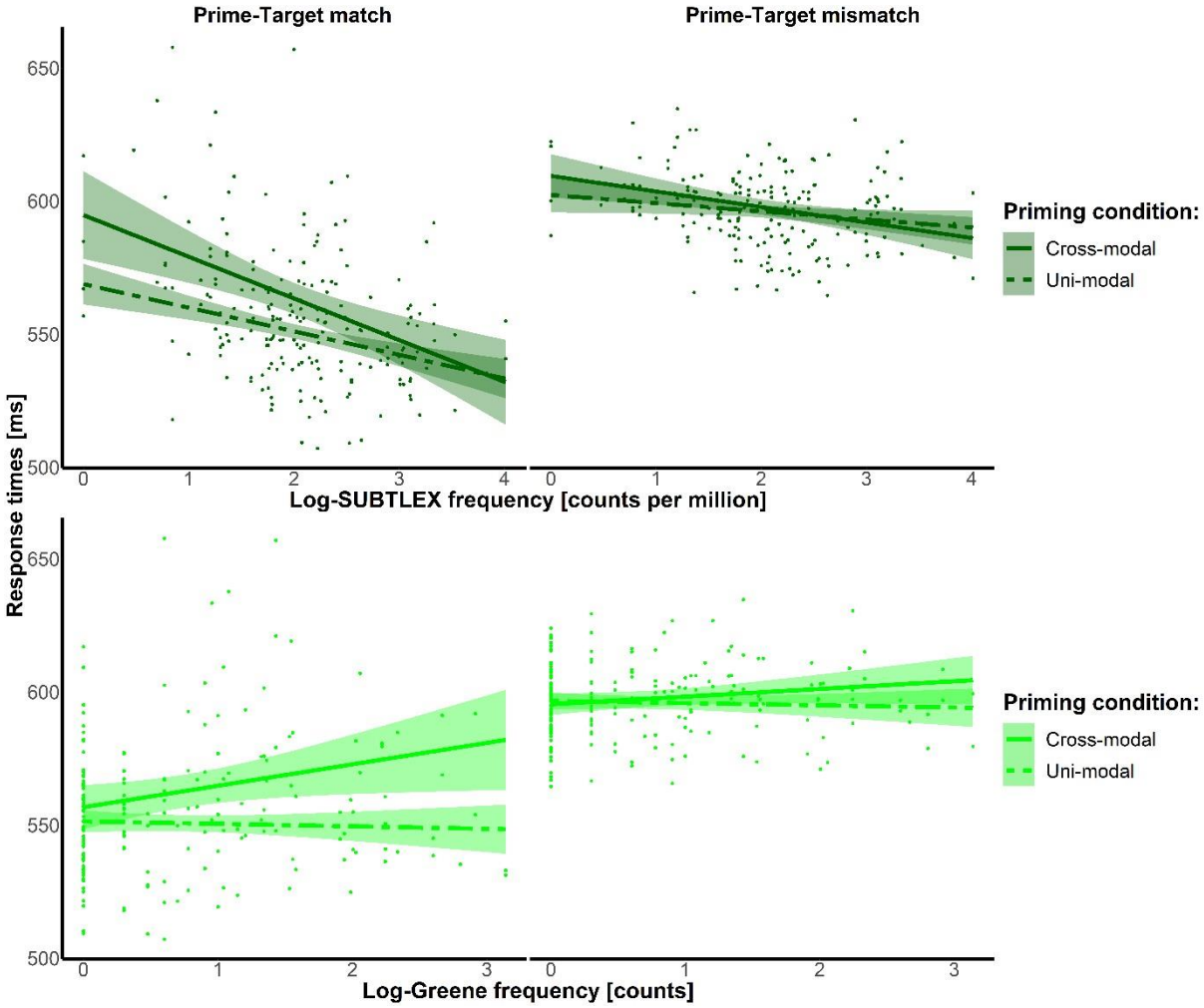
Supplementary table 17. Results from the post-hoc model with re-coded contrasts in the Replication exp

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.355	0.017	374.337	**<0.001**
SUBTLEX WF	-0.008	0.002	-3.865	**<0.001**
Cross-modal matching – Uni-modal matching	0.022	0.033	0.649	0.517
Cross-modal mismatching – Uni-modal mismatching	0.006	0.033	0.175	0.861
Cross-modal matching – Cross-modal mismatching	-0.063	0.003	-19.787	**<0.001**
Target modality (Words – Objects)	0.001	0.002	0.614	0.539
Greene OF	0.003	0.002	1.636	0.102
Visuo-orthographic PC	0.005	0.002	2.465	**0.014**
Concept familiarity (replication)	-0.001	0.002	-0.309	0.758
Image typicality (replication)	-0.009	0.002	-5.220	**<0.001**
Image visual PC1	0.001	0.002	0.557	0.577
Image visual PC2	0.003	0.002	1.881	0.060
Image visual PC3	0.001	0.002	0.618	0.537
Trial accuracy (Correct – Incorrect)	0.009	0.007	1.317	0.188

SUBTLEX x (Cross-modal matching – Uni-modal matching)	-0.014	0.003 -4.324	**<0.001**
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching)	-0.006	0.003 -1.664	0.096
SUBTLEX x (Cross-modal matching – Cross-modal mismatching)	-0.018	0.003 -5.291	**<0.001**
SUBTLEX x Target modality	-0.011	0.002 -4.786	**<0.001**
Target modality x (Cross-modal matching – Uni-modal matching)	-0.006	0.006 -0.949	0.343
Target modality x (Cross-modal mismatching – Uni-modal mismatching)	0.025	0.006 3.872	**<0.001**
Target modality x (Cross-modal matching – Cross-modal mismatching)	-0.017	0.006 -2.591	**0.010**
Greene x (Cross-modal matching – Uni-modal matching)	0.017	0.003 5.165	**<0.001**
Greene x (Cross-modal mismatching – Uni-modal mismatching)	0.007	0.003 1.959	0.050
Greene x (Cross-modal matching – Cross-modal mismatching)	0.013	0.003 3.874	**<0.001**
Greene x Target modality	0.001	0.002 0.344	0.731
SUBTLEX x (Cross-modal matching – Uni-modal matching) x Target modality	-0.016	0.007 -2.401	**0.016**
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching) x Target mod	0.005	0.007 0.794	0.427
SUBTLEX x (Cross-modal matching – Cross-modal mismatching) x Target mod	-0.027	0.007 -4.008	**<0.001**
Greene x (Cross-modal matching – Uni-modal matching) x Target modality	0.005	0.007 0.818	0.414
Greene x (Cross-modal misatching – Uni-modal mismatching) x Target modality	-0.006	0.007 -0.871	0.384
Greene x (Cross-modal matching – Cross-modal mismatching) x Target modality	0.010	0.007 1.487	0.137

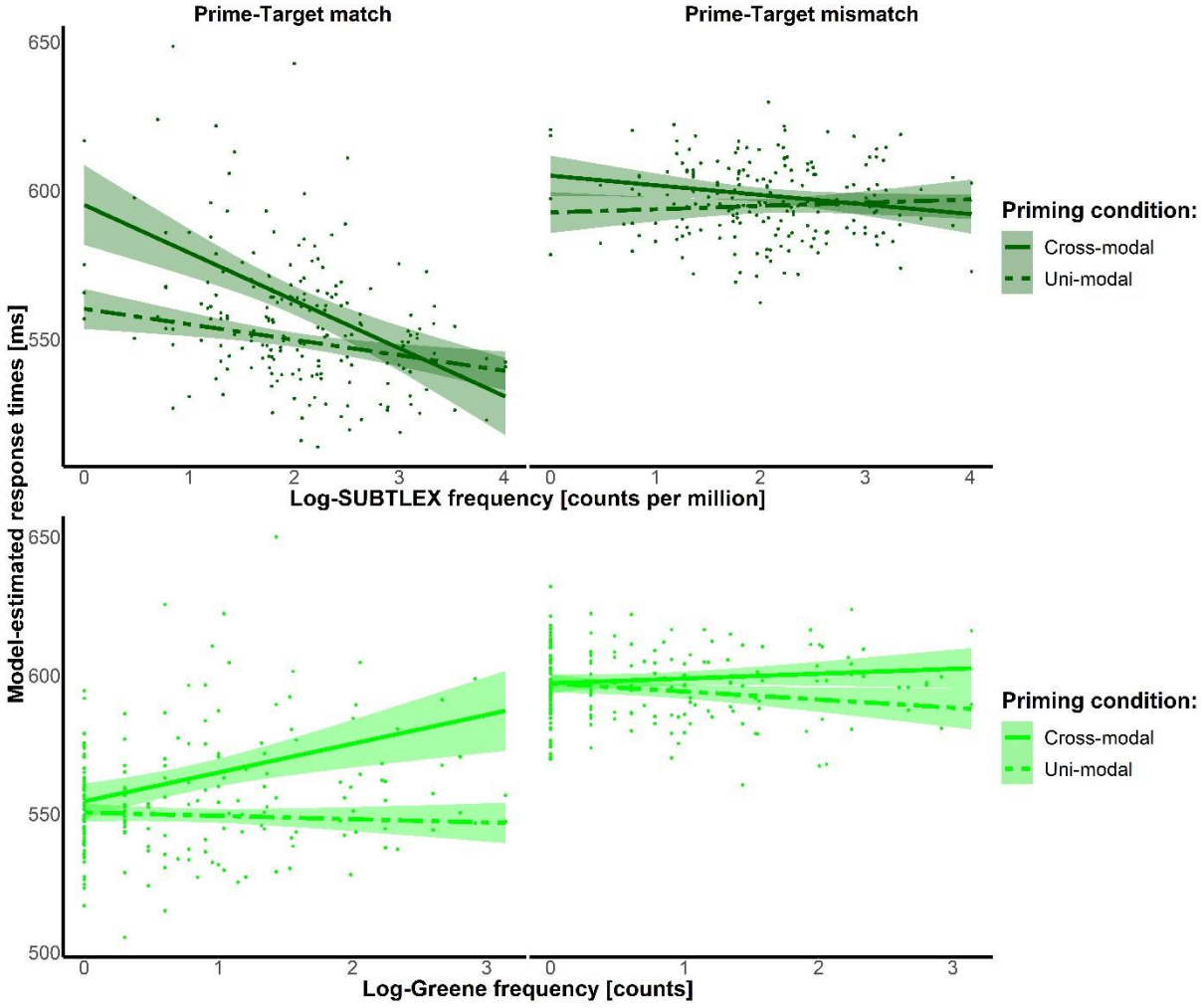
Supplementary figure 14 - Raw RTs from post-hoc conditions of Experiment 3 (between condition)

Raw response times in the priming conditions (Cross-modal solid lines, Uni-modal: dashed-dotted) and matching condition (Matching on the left, Mismatching on the right), as a function of SUBTLEX frequency (top, dark green) and Greene frequency (bottom, light green) in Replication experiment. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval



Supplementary figure 15 **- RTs estimated from post-hoc of interaction effects of Experiment 3 (between conditions)**

Response times as a function of logarithmic SUBTLEX frequency (top plots, dark green) and Greene frequency (bottom plots, light green) in the 3-way significant interaction with Matching condition and Priming condition (Cross-modal matching vs Uni-modal matching; Cross-modal mismatching vs Uni-modal mismatching; Cross-modal matching vs Cross-modal mismatching). RTs were estimated based on the selected model. Points present participant-based mean response times for concepts in the different frequency levels. Lines represent linear fitting of points (solid: cross-modal; dashed: uni-modal), and shaded areas represent 95 % confidence interval. Bottom left and top-left plots represent the effects in prime-target matching condition, while bottom-right and top-right plots represent the effects in prime-target mismatching condition



4 post-hoc models are additionally computed, one for every level of the re-coded factor (Cross-modal Matching, Uni-modal Matching, Cross-modal Mismatching, Uni-modal Mismatching)

*Rep_logRT ~ SUBTLEX WF * Target modality + Greene OF * Target modality +*
Concept familiarity (replication) + Image typicality (replication) +
Image visual PC1 + Image visual PC2 + Image visual PC3 +
Visuo-orthographic PC + Target repetition + Trial accuracy +
(1|Participants) + (1/Concepts)

Supplementary table 18. Results from the post-hoc individual models for conditions of interest in the Replication experiment.

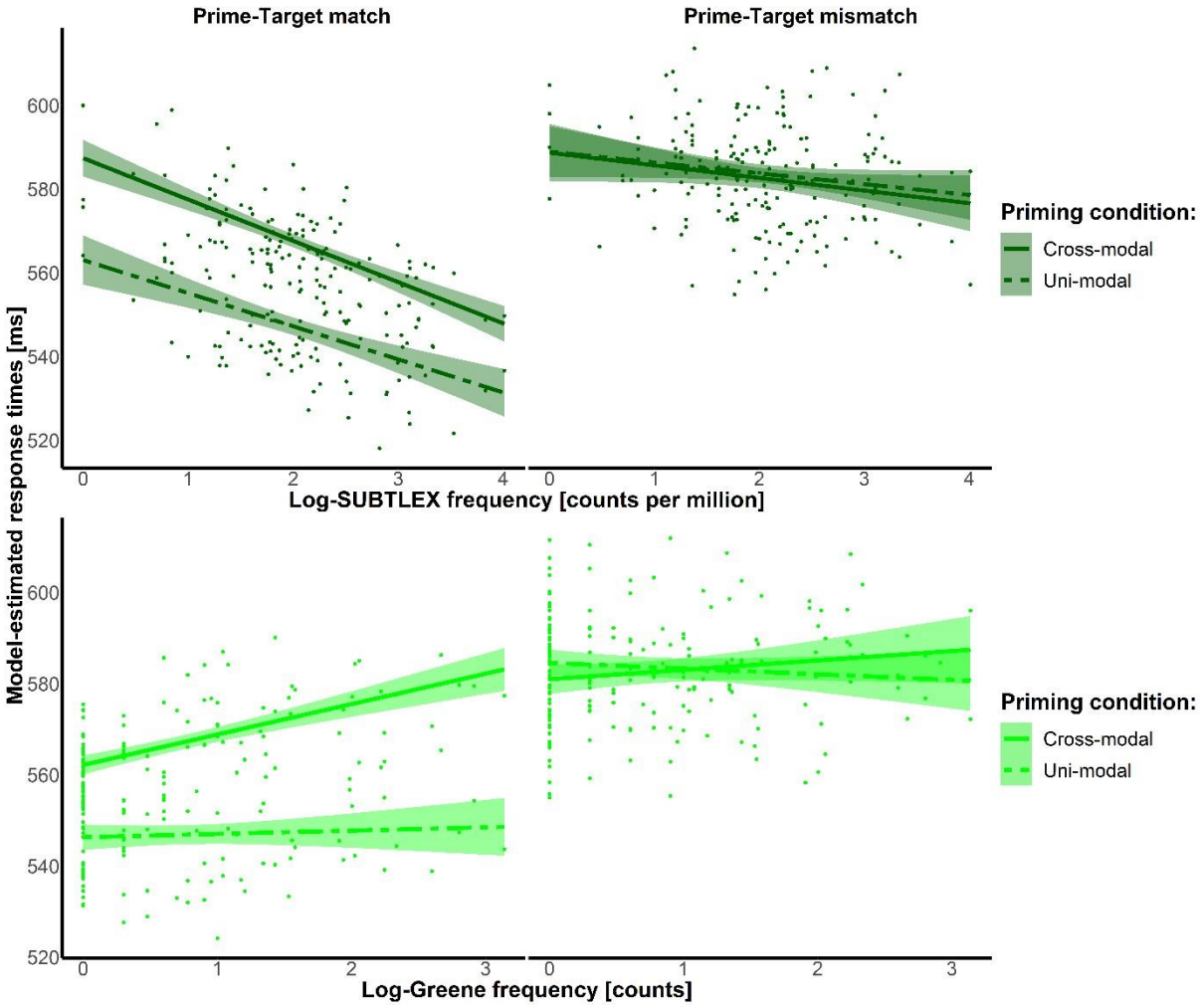
	Uni-modal Matching				**Cross-modal Matching**			
Predictors	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
(Intercept)	6.304	0.022	289.192	**<0.001**	6.340	0.027	237.811	**<0.001**
SUBTLEX WF	-0.011	0.003	-3.470	**0.001**	-0.013	0.006	-2.253	**0.024**
Target modality (Words – Objects)	0.016	0.004	3.748	**<0.001**	-0.036	0.005	-7.115	**<0.001**
Greene OF	0.001	0.003	0.398	0.691	0.010	0.005	1.873	0.061
Visuo-orthographic PC	0.002	0.003	0.818	0.413	0.009	0.006	1.613	0.107
Concept familiarity	-0.000	0.003	-0.010	0.992	-0.004	0.006	-0.754	0.451
Image typicality	-0.001	0.003	-0.463	0.643	-0.026	0.005	-5.426	**<0.001**
Image visual PC1	0.001	0.002	0.422	0.673	0.001	0.004	0.274	0.784
Image visual PC2	0.006	0.002	2.379	**0.017**	0.002	0.004	0.499	0.618
Image visual PC3	0.003	0.003	1.191	0.234	-0.000	0.005	-0.005	0.996
Target repetition	-0.026	0.002	-12.208	**<0.001**	-0.051	0.003	-20.200	**<0.001**
Trial accuracy (Correct – Incorrect)	0.019	0.011	1.781	0.075	-0.012	0.012	-1.051	0.293
SUBTLEX x (Words – Objects)	-0.011	0.004	-2.516	**0.012**	-0.027	0.005	-5.433	**<0.001**
Greene x (Words – Objects)	0.000	0.004	0.018	0.985	0.006	0.005	1.141	0.254

	Uni-modal Mismatching				**Cross-modal Mismatching**			
Predictors	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
(Intercept)	6.369	0.023	276.110	**<0.001**	6.367	0.027	233.415	**<0.001**
SUBTLEX WF	-0.003	0.003	-1.204	0.229	-0.004	0.003	-1.234	0.217
Target modality (Words – Objects)	0.001	0.004	0.337	0.736	-0.010	0.005	-2.018	**0.044**
Greene OF	-0.002	0.003	-0.699	0.484	0.003	0.003	1.044	0.297
Visuo-orthographic PC	0.002	0.003	0.565	0.572	0.006	0.003	2.018	**0.044**

Concept familiarity	0.002	0.003	0.692	0.489	-0.000	0.003	-0.094	0.925
Image typicality	-0.000	0.002	-0.205	0.838	-0.007	0.003	-2.695	**0.007**
Image visual PC1	0.004	0.002	1.896	0.058	-0.003	0.002	-1.071	0.284
Image visual PC2	0.003	0.002	1.489	0.137	0.001	0.002	0.221	0.825
Target repetition	-0.028	0.002	-13.127	**<0.001**	-0.040	0.002	-16.764	**<0.001**
Image visual PC3	-0.001	0.002	-0.538	0.591	0.002	0.002	0.733	0.464
Trial accuracy (Correct – Incorrect)	0.048	0.017	2.920	**0.004**	0.065	0.015	4.251	**<0.001**
SUBTLEX x (Words – Objects)	-0.006	0.004	-1.340	0.180	-0.000	0.005	-0.072	0.943
Greene x (Words – Objects)	0.001	0.004	0.314	0.754	-0.004	0.005	-0.796	0.426

Supplementary figure 16 – RTs estimated from post-hoc models of Experiment 3 (within conditions)

Effects of SUBTLEX WF (dark green, top) and Greene OF (light green, bottom) on reaction times estimated from the post-hoc models separately for each Priming condition (continuous and dashed-dotted line types) and Matching condition (left and right plots) in the Replication experiment. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval.

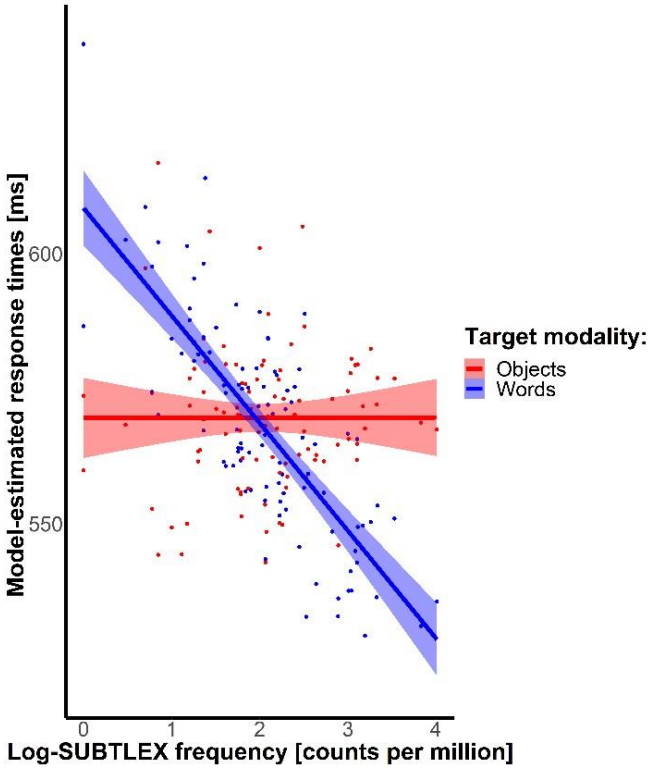


Supplementary table 19. Results from the post-hoc individual models for conditions of interest in Experiment 3.

	Cross-modal Matching Words				**Cross-modal Matching Object**			
Predictors	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
(Intercept)	6.341	0.027	239.179	**<0.001**	6.345	0.029	215.770	**<0.001**
SUBTLEX WF	-0.027	0.007	-3.699	**<0.001**	0.000	0.007	0.042	0.966
Greene OF	0.015	0.006	2.307	**0.021**	0.005	0.006	0.802	0.422
Visuo-orthographic PC	0.008	0.007	1.221	0.222	0.010	0.006	1.528	0.126
Concept familiarity	-0.008	0.007	-1.165	0.244	-0.000	0.006	-0.074	0.941
Image typicality	-0.024	0.006	-4.039	**<0.001**	-0.028	0.005	-5.162	**<0.001**
Image visual PC1	-0.001	0.005	-0.195	0.846	0.003	0.005	0.669	0.503
Image visual PC2	0.000	0.005	0.082	0.935	0.004	0.005	0.734	0.463
Image visual PC3	0.004	0.006	0.726	0.468	-0.004	0.005	-0.751	0.453
Targer repetition	-0.051	0.007	-7.415	**<0.001**	-0.055	0.007	-8.106	**<0.001**
Trial accuracy	-0.017	0.016	-1.001	0.317	-0.018	0.017	-1.065	0.287

Supplementary figure 17 – RTs estimated from post-hoc models of Experiment 3 (within modalities)

Effects of SUBTLEX WF (left) on reaction times estimated from the post-hoc models for Cross-modal matching trials of words (blue) and Cross-modal matching trials of objects (red) in the Replication experiment. Points show concepts with different level of frequency, averaged across participants; lines represent linear fitting of points and shaded areas represent 95 % confidence interval.



Supplementary Materials 14 – Effect of dlexDB on Experiment 1

*Exp1_logRT ~ dlexDB WF * Concept modality +*

Concept category + Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition +

Trial accuracy + (1|Participants) + (1/Concepts)

Supplementary table 20. Results from main model of Exp 1 including dlexDB instead of SUBTLEX

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.480	0.021	301.419	**<0.001**
Target modality (Words – Objects)	0.094	0.005	20.523	**<0.001**
dlexDB WF	-0.021	0.008	-2.801	**0.005**
Visuo-orthographic PC	-0.000	0.008	-0.050	0.960
Concept familiarity	-0.004	0.003	-1.263	0.206
Image typicality	-0.005	0.003	-1.416	0.157
Image visual PC1	-0.002	0.006	-0.261	0.794
Image visual PC2	0.020	0.006	3.272	**0.001**
Image visual PC3	0.009	0.006	1.481	0.139
Target repetition	-0.011	0.002	-4.926	**<0.001**
Trial accuracy (Correct – Incorrect)	-0.018	0.009	-2.032	**0.042**
Concept category (Natural – Man-made)	0.001	0.013	0.074	0.941
dlexDB x (Words – Objects)	-0.020	0.005	-4.394	**<0.001**

Results from 2 post-hoc models one for each stimulus modality

Exp1_logRT ~ dlexDB WF +

Concept category + Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition +

Trial accuracy + (1|Participants) + (1/Concepts)

Supplementary table 21. Results from post-hoc models of Exp 1 including dlexDB instead of SUBTLEX

	Objects trials	**Word trials**

Predictors	*β*	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
(Intercept)	6.450	0.022	289.121	<**0.001**	6.520	0.025	265.635	<**0.001**
dlexDB WF	-0.012	0.009	-1.289	0.197	-0.031	0.008	-4.081	<**0.001**
Concept category (Natural – Man-made)	0.009	0.016	0.601	0.548	-0.009	0.013	-0.664	0.506
Visuo-orthographic PC1	-0.000	0.010	-0.018	0.986	-0.001	0.008	-0.090	0.928
Concept familiarity	-0.001	0.005	-0.277	0.782	-0.007	0.004	-1.480	0.139
Image typicality	-0.009	0.004	-2.054	**0.040**	-0.002	0.004	-0.462	0.644
Image visual PC1	-0.004	0.007	-0.597	0.551	0.001	0.006	0.179	0.858
Image visual PC2	0.026	0.007	3.524	<**0.001**	0.013	0.006	2.172	**0.030**
Image visual PC3	0.005	0.008	0.665	0.506	0.014	0.006	2.192	**0.028**
Target repetition	0.009	0.021	0.428	0.668	-0.030	0.024	-1.290	0.197
Trial accuracy (Correct – Incorrect)	-0.059	0.013	-4.403	<**0.001**	0.004	0.011	0.398	0.690

Supplementary Materials 15 – Conceptual Distinctiveness in Experiment 1

*Exp1_logRT ~ SUBTLEX WF * Concept modality +*

Concept category + Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition + Conceptual Distinctiveness +

Trial accuracy + (1|Participants) + (1/Concepts)

Supplementary table 22. Results of Experiment 1 including CD as covariate

Predictors	*β*	*SE*	*t*	*p*
(Intercept)	6.479	0.021	302.570	<**0.001**
Concept modality (Words – Objects)	0.094	0.005	20.529	<**0.001**
SUBTLEX WF	-0.032	0.008	-4.150	<**0.001**
Visuo-orthographic PC	-0.006	0.007	-0.805	0.421
Concept familiarity	-0.003	0.003	-0.980	0.327
Image typicality	-0.004	0.003	-1.279	0.201
Image visual PC1	-0.002	0.006	-0.263	0.792
Image visual PC2	0.019	0.006	3.259	**0.001**
Image visual PC3	0.008	0.006	1.295	0.195
Target repetition	-0.011	0.002	-4.932	<**0.001**
Conceptual Distinctiveness (CD)	0.002	0.007	0.295	0.768

Trial accuracy (Correct – Incorrect)	-0.017	0.009	-1.940	0.052
Concept category (Natural – Man-made)	0.003	0.013	0.206	0.837
SUBTLEX WF x (Words – Objects)	-0.019	0.005	-4.160	**<0.001**

Supplementary Materials 16 – Effect of ADE20K on Experiment 2

*Exp2_logRT ~ SUBTLEX WF * Recoded factor * Target modality +*

*ADE20K OF * Recoded factor * Target modality +*

Concept familiarity + Image typicality +

Image visual PC1 + Image visual PC2 + Image visual PC3 +

Visuo-orthographic PC + Target repetition + Trial accuracy +

(1|Participants) + (1/Concepts)

Supplementary table 23. Results from main model of Exp 2 including ADE20K instead of Greene

Predictors	***β***	***SE***	***t***	***p***
(Intercept)	6.225	0.020	315.721	**<0.001**
SUBTLEX WF	-0.005	0.002	-2.420	**0.016**
Cross-modal matching – Uni-modal matching	0.005	0.006	0.829	0.407
Cross-modal mismatching – Uni-modal mismatching	0.008	0.006	1.380	0.168
Cross-modal matching – Cross-modal mismatching	-0.075	0.003	-23.729	**<0.001**
Target modality (Words – Objects)	-0.008	0.002	-3.612	**<0.001**
ADE20K OF	0.008	0.002	4.391	**<0.001**
Visuo-orthographic PC	0.010	0.002	5.256	**<0.001**
Concept familiarity	-0.001	0.002	-0.739	0.460
Image typicality	-0.004	0.002	-2.638	**0.008**
Image visual PC1	0.002	0.002	1.085	0.278
Image visual PC2	0.005	0.002	3.463	**0.001**
Image visual PC3	-0.002	0.002	-1.134	0.257
Target repetition	-0.036	0.003	-13.329	**<0.001**
Trial accuracy (Correct – Incorrect)	0.029	0.005	5.393	**<0.001**
SUBTLEX x (Cross-modal matching – Uni-modal matching)	-0.027	0.004	-7.622	**<0.001**
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching)	-0.006	0.004	-1.568	0.117
SUBTLEX x (Cross-modal matching – Cross-modal mismatching)	-0.035	0.004	-9.901	**<0.001**
SUBTLEX x Target modality	-0.007	0.003	-2.651	**0.008**
Target modality x (Cross-modal matching – Uni-modal matching)	-0.010	0.006	-1.657	0.098
Target modality x (Cross-modal mismatching – Uni-modal mismatching)	0.036	0.006	5.720	**<0.001**
Target modality x (Cross-modal matching – Cross-modal mismatching)	-0.015	0.006	-2.322	**0.020**
ADE20K x (Cross-modal matching – Uni-modal matching)	0.018	0.004	5.167	**<0.001**

ADE20K x (Cross-modal mismatching – Uni-modal mismatching)	-0.002	0.004	-0.617	0.537
ADE20K x (Cross-modal matching – Cross-modal mismatching)	0.028	0.004	7.731	<**0.001**
ADE20K x Target modality	0.004	0.003	1.726	0.084
SUBTLEX x (Cross-modal matching – Uni-modal matching) x Target modality	-0.005	0.007	-0.743	0.457
SUBTLEX x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.005	0.007	0.762	0.446
SUBTLEX x (Cross-modal matching – Cross-modal mismatching) x Target modality	-0.003	0.007	-0.448	0.654
ADE20K x (Cross-modal matching – Uni-modal matching) x Target modality	0.014	0.007	1.937	0.053
ADE20K x (Cross-modal mismatching – Uni-modal mismatching) x Target modality	0.004	0.007	0.588	0.557
ADE20K x (Cross-modal matching – Cross-modal mismatching) x Target modality	0.003	0.007	0.441	0.660

Additional Bibliography

- Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, *16*(1), 3–14. https://doi.org/10.1016/0304-4076(81)90071-3
- Brehm, L., & Alday, P. M. (2020). *A decade of mixed models: It's past time to set your contrasts*. Talk presented at the 26th Architectures and Mechanisms for Language Processing Conference (AMLap 2020). Potsdam, Germany. 2020-09-03 - 2020-09-05.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60).
- Robinson, A. P., & Froese, R. E. (2004). Model validation using equivalence tests. *Ecological Modelling*, *176*(3–4), 349–358. https://doi.org/10.1016/j.ecolmodel.2004.01.013
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038