AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 10, 2861–2881
https://doi.org/10.1037/xge0001342

Access to Meaning From Visual Input: Object and Word Frequency Effects in Categorization Behavior

Klara Gregorová[1], Jacopo Turini[1], Benjamin Gagl[1, 2], and Melissa Le-Hoa Võ[1]
[1] Department of Psychology and Sports Sciences, Goethe University
[2] Department of Special Education and Rehabilitation, University of Cologne

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 versus man-made categorization task (Experiment 1) and a matching–mismatching priming task (Experiments 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 uni-modal priming. Moreover, in cross-modal priming, we 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 recognition, mainly when semantic processing happens in the context of previously presented information. These findings have major implications for studies attempting to include frequency measures in investigations of access to meaning from visual inputs.

Public Significance Statement
This study highlights the role of the amount of visual and linguistic experience on the process of visual recognition, with which we make sense of the things we see.

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

Supplemental materials: https://doi.org/10.1037/xge0001342.supp

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, for example, 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, that is, the core neural substrate of high-level vision, compared the brain activation in response to these

This article was published Online First May 8, 2023.
Jacopo Turini https://orcid.org/0000-0002-2132-9150
Klara Gregorová and Jacopo Turini are the joint first authors. Benjamin Gagl and Melissa Le-Hoa Võ are the joint senior authors.

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.

The authors 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").

Data and materials are available at https://osf.io/d3j9h/files/.

Klara Gregorová served as lead for conceptualization and contributed equally to software. Jacopo Turini served as lead for visualization and contributed equally to conceptualization. Benjamin Gagl served in a supporting role for conceptualization, formal analysis, investigation, methodology, writing–original draft, and writing–review and editing. Melissa Le-Hoa Võ served as lead for funding acquisition, project administration, and resources and served in a supporting role for writing–original draft and writing–review and editing. Klara Gregorová and Jacopo Turini contributed to data curation, investigation, methodology, writing–original draft, writing–review and editing, and formal analysis equally. Benjamin Gagl and Melissa Le-Hoa Võ contributed to supervision equally.

Correspondence concerning this article should be addressed to Jacopo Turini, Department of Psychology and Sports Sciences, Goethe University, PEG Room 5.G105, Theodor-W.-Adorno Platz 6, 60323 Frankfurt/Main, Germany. Email: turini@psych.uni-frankfurt.de

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, for example, objects and words, to understand the process of accessing the same semantic representation, that is, the identical meaning (Devereux et al., 2013; Fairhall & Caramazza, 2013; Shelton & Caramazza, 1999; Shinkareva et al., 2011). We followed this approach and investigated how different types of visual input can access the same semantic representation. We were particularly interested in the frequency of occurrence in the world, operationalized by both word and object frequency (OF) 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 (WF) effect" is a well-established finding. The 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, for example, lexical decision tasks (Balota et al., 2004; for a review, Brysbaert et al., 2011, 2018), consistently show WF effects, that is, 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" (Coltheart et al., 2001; Engbert et al., 2005; Forster & Chambers, 1973; McClelland & Rumelhart, 1981; Morton, 1979).

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 (Criss & Malmberg, 2008; Morrison et al., 1992; 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 et al. (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 naming performance also showed frequency effects based on word-based estimates (Almeida et al., 2007; Bates et al., 2001; 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 to WF effects even when object pictures (nonverbal stimuli) are involved, since they reflect the word retrieval process underlying the object naming task rather than the object recognition process involved in processing the visual input.

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: that is, 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 SUBTLEX-DE, about 25, four 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, one has to determine how these measures estimated from subtitles might affect recognition performance beyond their specific modality (i.e., words). For example, they might reflect the experience with meaning instead of simply words (reflecting semantic domain-general processing instead of language-specific processing). 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 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 (Russell 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 well-established 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 Experiments 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 Method

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 N is below 30 participants (e.g., see Yarkoni, 2009), we gathered and tested 60 participants online (of whom 42 fit the above-mentioned criteria, forming our final sample $N = 42$) in Experiments 1 and 2, as well as the rating study judging typicality and familiarity of the used stimuli (age: $M = 23.52$, $SD = 8.11$, range = 16–56 years; gender: 34 F, seven M, one person did not report).

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 years; gender: 43 F, 10 M); and yet another 53 participants took part in the uni-modal priming task (age: $M = 22.66$, $SD = 4.81$, range = 18–39 years; gender: 35 F, 16 M, two NB). 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 between-participants design in the replication, following this formula: $Nbetween = (Nwithin \times 2)/(1 - \rho)$, 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 Experiment 2, $\rho = 0.208$). The formula was then solved for $Nbetween = (42 \times 2)/(1 - 0.208) = 106.061$ (Maxwell et al., 2017).

Additionally, two distinct groups of participants were recruited to collect further ratings regarding the stimuli used: (a) One group of 20 participants (age: $M = 21.65$, $SD = 2.66$, range = 19–29 years; gender: 10 F, 10 M) 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. (b) 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 years; gender: nine F, seven 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). For a summary, see Table 1.

Table 1
Demographic Information of the Participants

Participant group	Age	Gender	Native languages	Education
Experiment 1–Experiment 2—ratings familiarity and typicality	$M = 23.52$, $SD = 8.11$, range = 16–56 years	34 F, seven M, one did not report	37 one language (German), five two/more (one being German)	Four university degree, 34 high-school diploma, one technical school diploma, and three did not report
Experiment 3—cross-modal Priming	$M = 22.87$, $SD = 6.83$, range = 18–50 years	43 F, 10 M	43 one language (German), 10 two/more (one being German)	Five university degree, 48 high-school diploma
Experiment 3—uni-modal Priming	$M = 22.66$, $SD = 4.81$, range = 18–39 years	35 F, 16 M, two NB	40 one language (German), 13 two/more (one being German)	Four university degree, 47 high-school diploma, one technical school diploma, one did not report
Independent ratings familiarity and typicality	$M = 21.65$, $SD = 2.66$, range = 19–29 years	10 F, 10 M	16 one language (German), four two/more (one being German)	20 high-school diploma
Ratings conceptual distinctiveness	$M = 23$, $SD = 4.75$, range = 18–33 years	Nine F, seven M	20 German native speakers	No information collected

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 × 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 (OF)

OF measures represent the log-transformed (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 one count (i.e., 0 log10-counts). Density distribution of ADE20K OF and Greene OF for the set of stimuli can be found in Figure 1A and B in the online supplemental materials 1.

Word-Based Frequency Measures (WF)

WF measures are based on the number of occurrences of a word in a corpus of linguistic materials. Specifically, as for OF, 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 Figure 1C and D in the online supplemental 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 (CD; Konkle et al., 2010). For this purpose, we set up yet another rating study in which we collected ratings regarding our stimuli's CD 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 CD 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). CD ratings were obtained for every concept by averaging ratings across participants.

For information about variability and distribution of the ratings across participants, see Figure 2 in the online supplemental materials 1.

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; see "PCA procedure for visual and visuo-orthographic predictors," p. 3 in the online supplemental materials 1).

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 three orthogonal 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 "Image visual PCs," p. 3 in the online supplemental materials 1, Figure 3, Table 1 in the online supplemental 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 (see "Visuo-orthographic PCs," p. 4 in the online supplemental materials 1, Figure 4 and Table 2 in the online supplemental materials 1). Correlations between all predictors can be found in Figure 5 in the online supplemental 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: 6.7×6.7 cm; words, uppercase letter: ca. 0.7 cm). In all parts of the experiment, the screen background was grey (hexadecimal color #424242). The CD 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

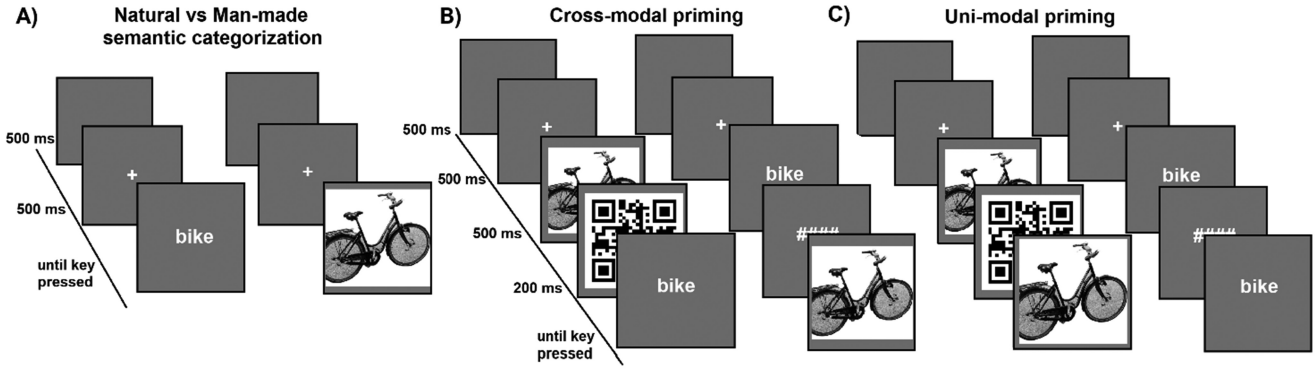
Experiment 1

Figure 1A shows an example of the natural (e.g., apple) versus 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.

Experiment 2

In the second experiment, we implemented a priming task that included uni-modal and cross-modal prime–target pairs, consisting of

Figure 1
Experimental Design



Note. (A) Experiment 1. Categorization of natural versus man-made object images and words. (B) Experiment 2. Categorization of prime–target matches versus 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. Example picture was taken and adapted from the THINGS dataset (Hebart et al., 2019). These were not part of the actual experiment.

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. We decided to have the same participants rate the same stimuli they had seen during Experiments 1 and 2 in order to capture individual participants' variance more precisely. 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 min on Day 1 (Experiment 1 and Experiment 2) and about 30 min on Day 2 (ratings).

Experiment 3

Experiment 3 was run to replicate the findings of Experiment 2 addressing potential carry-over effects in the previous study. 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 min each.

Replication Typicality and Familiarity Ratings

We collected more typicality and familiarity ratings from a new group of participants to further control for carry-over effects, which could emerge when the same participants first see the stimuli during the experiments and then also rate the same stimuli afterwards. 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. We used these ratings as covariates in the model of Experiment 3 data, but we also reanalyzed the data of Experiments 1 and 2 using this set of independent ratings. Data collection lasted about 30 min.

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 CD or high in the original investigation. By definition, concepts with high CD 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 6-point scale spanning from 1 (*very similar*) to 6 (*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 min.

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 1,500 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 cut-off at 1,500 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 response times (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 log-transformation to obtain a normal distribution to account for the ex-Gauss distribution of reaction time measures. No further preprocessing 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 factors for the experimental conditions (see "Model specifications," p. 6 in the online supplemental 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 on this in "Analysis details," p. 7 in the online supplemental 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, that is, 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 factors, 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 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 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/.

Results

Results: Experiment 1

The initial model comparison showed that, in the man-made versus 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: $\chi^2 = 29.153$, $p < .001$; dlexDB WF: $\chi^2 = 15.447$, $p = .001$), while considering OF measures did not produce a better fit (ADE20K OF: $\chi^2 = 3.228$, $p = .072$; Greene OF: $\chi^2 = 0.867$, $p = .352$). For object recognition trials, only SUBTLEX WF resulted in a significant improvement of the model fit ($\chi^2 = 6.163$, $p = .013$; dlexDB WF: $\chi^2 = 1.646$, $p = .200$; ADE20K OF: $\chi^2 = 0.051$, $p = .821$; Greene OF: $\chi^2 = 0.310$, $p = .578$; for more details, see formula and Table 3 in the online supplemental 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 frequency in both word and object trials, with no significant increase in explained variance for any of the two object-based frequencies. 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 in the online supplemental materials 3 for R-based formula; no multicollinearity detected: variance inflation factors <5). We found a significant SUBTLEX WF by Concept modality interaction ($\beta = -0.019$, $SE = 0.005$, $t = -4.160$, $p < .001$), showing a more substantial facilitatory SUBTLEX WF effect (i.e., faster RTs for high frequency items)

for words compared to objects (see Table 2 and Figure 2; for details on level of collinearity see Table 4 in the online supplemental materials 3, for similar results including participants' demographics, see Table 5 in the online supplemental materials 3).

Two post hoc models, for objects and words separately, showed significant SUBTLEX WF effects for both words ($\beta = -0.041$, $SE = 0.007$, $t = -5.794$, $p < .001$) and objects ($\beta = -0.022$, $SE = 0.009$, $t = -2.524$, $p = .012$), but the effect size for words was almost doubled (1.86 times higher; for more details on this, see Table 6 and Figure 6 in the online supplemental materials 4; for similar results using independent ratings of familiarity and typicality, see Table 7 in the online supplemental materials 5).

Since Taikh et al. (2015) failed to find a SUBTLEX WF effect for object recognition using various sets of "semantic richness" predictors (not available here), we reanalyzed the WF effect in Experiment 1. This time, we included CD as a covariate, a measure of categorical width that likely correlates with semantic richness. However, this did not change the pattern of effects (see Table 8 in the online supplemental materials 6).

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, that is, the same semantic representation being accessed from two different visual inputs. The phenomenon producing the WF effect from language experience may rely not only on modality-specific processes (i.e., WF effect in word recognition, naming, and production), but also on domain-general semantic processes (i.e., WF effect also found for object recognition). Given the unchanged results when considering the contribution of CD (as a measure of categorical width), it also seems unlikely that the observed WF effect in object recognition would have solely emerged as a confound from other unmeasured semantic dimensions, but rather WF effect reflects genuine products of language experience influencing semantic processing. Interestingly, neither OF measure improved the fit. Thus, OF is less relevant in this simple categorization task.

Table 2
Results of the Main Model of Experiment 1

Predictors	β	*SE*	*t*	*p*
(Intercept)	6.479	0.021	302.552	**<.001**
Concept modality (words − objects)	0.094	0.005	20.529	**<.001**
SUBTLEX WF	−0.031	0.007	−4.417	**<.001**
Visuo-orthographic PC	−0.006	0.007	−0.818	.413
Concept familiarity	−0.003	0.003	−0.983	.326
Image typicality	−0.004	0.003	−1.302	.193
Image visual PC1	−0.002	0.006	−0.316	.752
Image visual PC2	0.019	0.006	3.253	**.001**
Image visual PC3	0.008	0.006	1.410	.159
Target repetition	−0.011	0.002	−4.933	**<.001**
Trial accuracy (correct − incorrect)	−0.017	0.009	−1.936	.053
Concept category (natural − man-made)	0.001	0.012	0.116	.908
SUBTLEX × (Words − Objects)	−0.019	0.005	−4.160	**<.001**

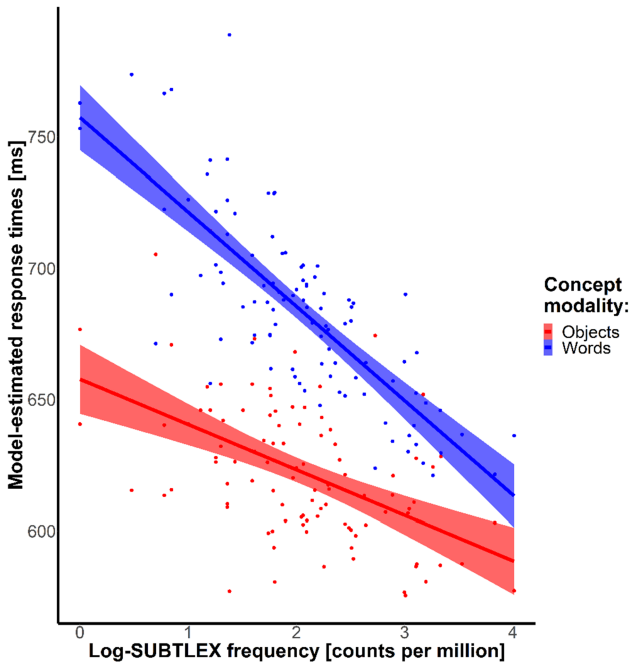
Note. WF = word frequency; PC = principal component. Significance in bold for $p < .05$.

Results: Experiment 2

In Experiment 2, we implemented a priming task to investigate the effect of the novel object-based frequency measures in a paradigm where a prime provided context. Critical here is that the prime allowed the prediction of an upcoming visual stimulus. Besides, we wanted to test the role of semantic processing on the WF and OF effects more explicitly. The critical manipulation, therefore, contrasted cross-modal and uni-modal priming (Eisenhauer et al., 2019, 2021; Scarborough et al., 1977; Tversky, 1969). As described earlier, only cross-modal priming involves conceptual/semantic information transfer from prime to target processing without shared perceptual processing. thus, strong frequency effects in cross-modal priming would indicate a substantial involvement of semantic processes in the emergence of frequency effects.

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: $\chi^2 = 10.105$, $p = .039$, words: $\chi^2 = 27.302$, $p < .001$;

Figure 2
Main Results of Experiment 1



Note. Semantic categorization response times as a function of logarithmic SUBTLEX frequency, separated for objects and words; response times were estimated based on the SUBTLEX WF × Concept Modality interaction term from the selected model. Points present participant-based mean reaction times separated for stimulus type (light gray: object stimuli; dark gray: word stimuli) in the different frequency levels. Lines represent linear fitting of points, and shaded areas represent 95% confidence interval. WF = word frequency. See the online article for the color version of this figure.

Greene OF, objects: $\chi^2 = 27.547$, $p < .001$, words: $\chi^2 = 43.409$, $p < .001$; SUBTLEX WF, objects: $\chi^2 = 52.695$, $p < .001$, words: $\chi^2 = 43.409$, $p < .001$; dlexDB WF, objects: $\chi^2 = 33.014$, $p < .001$, words: $\chi^2 = 19.105$, $p < .001$; for detailed information, see formula and Table 9 in the online supplemental materials 7). We found that both Greene and SUBTLEX frequencies had stronger fit improvements then 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 factors 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 ($\chi^2 = 5.455$, $p = .708$). 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, see in the online supplemental materials 8)

When participants had to judge whether prime and target had the same meaning, we found a significant three-way interaction between frequency, matching condition, and priming condition, for both SUBTLEX WF ($\beta = 0.017$, $SE = 0.005$, $t = 3.687$, $p < .001$; Figure 3 top) and Greene OF measures ($\beta = -0.020$, $SE = 0.005$, $t = -4.256$, $p < .001$; Figure 3 bottom; for full results, see Table 3; for details on level of collinearity, see Table 10 in the online supplemental materials 8; for similar results including participants' demographics see Table 11 in the online supplemental materials 8). 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, that is, WF and OF effects in matching and priming conditions were similar for words and objects (SUBTLEX WF: $\beta = 0.006$, $SE = 0.009$, $t = 0.612$, $p = .541$; Greene OF: $\beta = 0.002$, $SE = 0.009$, $t = 0.227$, $p = .821$).

Posthoc models showed that the frequency effects were stronger in cross-modal matching trials than in uni-modal matching trials (SUBTLEX WF: $\beta = -0.023$, $SE = 0.003$, $t = -7.094$, $p < .001$; Greene OF: $\beta = 0.018$, $SE = 0.003$, $t = 5.379$, $p < .001$), while no differential effects were found between cross-modal mismatching and uni-modal mismatching trials (SUBTLEX WF: $\beta = -0.006$, $SE = 0.003$, $t = -1.870$, $p = .062$; Greene OF: $\beta = -0.002$, $SE = 0.003$, $t = -0.644$, $p = .520$; see Table 12 and Figure 7 in the online supplemental materials 9). Besides, we only found strongly significant effects of SUBTLEX frequency ($\beta = -0.019$, $SE = 0.006$, $t = -3.230$, $p = .001$) and Greene frequency ($\beta = 0.023$, $SE = 0.005$, $t = 4.710$, $p < .001$) 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 Table 13 and Figure 8 in the online supplemental materials 9; for replication of results with independent ratings of familiarity and typicality, see Table 14 in the online supplemental materials 10). In a further exploratory analysis, we showed a substantial stability of the effects for the individual participants and individual concepts across the two modalities (for details, see "Description exploratory analysis" p. 26 in the online supplemental materials 11 and Figure 9 in the online supplemental materials 11) which again suggests *nondifferent (i.e., statistically equivalent)* processes across modalities.

Discussion: Experiment 2

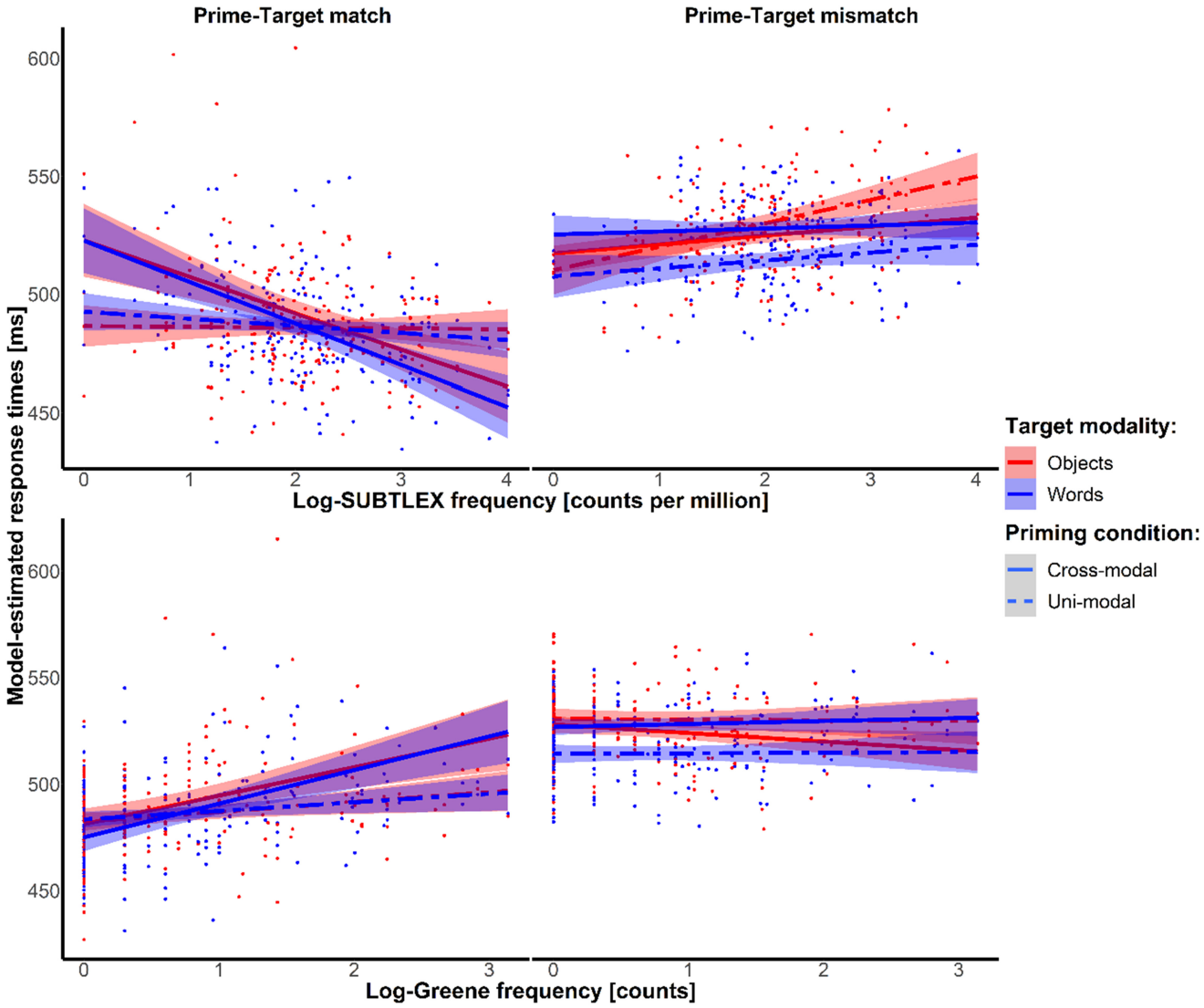
In Experiment 2, we replicated the facilitatory effect of the SUBTLEX WF found in Experiment 1 for both words and objects. It is important to note that we found the SUBTLEX WF effect only when participants categorized objects or words after seeing a semantically matched prime from the other stimulus 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, that is, 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 prime–target matching, the task difficulty for objects and words was comparable, that is, it had common perceptual and semantic matching processes across modalities. For example, considering the uni-modal priming, in both word-priming-word and object-priming-object, the task is based on similar perceptual matching processes, where the perceptual appearance of the prime predicts the perceptual appearance of the target. In the case of cross-modal priming, the processing requires extracting meaning from one modality (word or object) and forming a prediction based on the semantic identity of the input stimulus from the other modality (object or word). There seems to be no reason for these perceptual or semantic predictions being faster in one modality than the other, especially considering that the prime–target delay was long enough to cancel out any object processing advantage on word processing.

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. Semantic prediction processes are not required in uni-modal priming. In the semantic categorization task of Experiment 1, we only used single word/object trials to preclude predictions about upcoming stimuli. 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

Figure 3

Main Results of Experiment 2



Note. Response times as a function of logarithmic SUBTLEX frequency (top plots) and Greene frequency (bottom plots) in the different conditions of Experiment 2; response times were estimated based on the selected model. Points present participant-based mean response times separated for stimulus type (light gray: object stimuli; dark gray: 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 prime–target matching condition, while top-right and bottom-right plots represent the effects in prime–target mismatching condition. See the online article for the color version of this figure.

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, that is, showing facilitation for concepts encountered less often in our visual world. It seems counterintuitive that fewer occurrences could produce more efficient processing, but we can speculate on two interpretations to explain this effect found for both words and objects in Experiment 2. One possible explanation is that one can remember a concept better when presented with fewer exemplars of that category because more encounters with exemplars create interference that weakens the memory processing (Konkle et al., 2010). Based on

Table 3
Results of the Main Model From Experiment 2

Predictors	β	*SE*	*t*	*p*
(Intercept)	6.225	0.020	315.728	**<.001**
Greene OF	0.008	0.002	4.474	**<.001**
Matching condition (mismatch − match)	0.073	0.002	32.684	**<.001**
Target modality (words − objects)	−0.008	0.002	−3.618	**<.001**
Priming condition (cross-modal − uni-modal)	0.006	0.005	1.173	.241
SUBTLEX WF	−0.004	0.002	−1.812	.070
Visuo-orthographic PC	0.010	0.002	4.877	**<.001**
Concept familiarity	−0.001	0.002	−0.845	.398
Image typicality	−0.004	0.002	−2.562	**.010**
Image visual PC1	0.002	0.002	1.217	.224
Image visual PC2	0.004	0.002	2.788	**.005**
Image visual PC3	−0.001	0.002	−0.832	.405
Target repetition	−0.036	0.003	−13.357	**<.001**
Trial accuracy (correct – incorrect)	0.030	0.005	5.444	**<.001**
Greene × Matching Condition	−0.017	0.002	−7.454	**<.001**
Greene × Target Modality	0.003	0.002	1.396	.163
Matching Condition × Target Modality	−0.009	0.004	−1.939	.053
Greene × Priming Condition	0.008	0.002	3.347	**.001**
Matching Condition × Priming Condition	0.003	0.004	0.737	.461
Priming Condition × Target Modality	0.013	0.004	2.872	**.004**
SUBTLEX × Matching Condition	0.021	0.002	9.074	**<.001**
SUBTLEX × Target Modality	−0.005	0.002	−2.378	**.017**
SUBTLEX × Priming Condition	−0.015	0.002	−6.335	**<.001**
Greene × Matching Condition × Target Modality	0.003	0.005	0.668	.504
Greene × Matching Condition × Priming Condition	−0.020	0.005	−4.256	**<.001**
Greene × Priming Condition × Target Modality	0.007	0.005	1.416	.157
Matching Condition × Priming Condition × Target Modality	0.046	0.009	5.220	**<.001**
SUBTLEX × Matching Condition × Target Modality	−0.002	0.005	−0.472	.637
SUBTLEX × Matching Condition × Priming Condition	0.017	0.005	3.687	**<.001**
SUBTLEX × Priming Condition × Target Modality	0.003	0.005	0.569	.570
Greene × Matching Condition × Priming Condition × Target Modality	0.002	0.009	0.227	.821
SUBTLEX × Matching Condition × Priming Condition × Target Modality	0.006	0.009	0.612	.541

Note. OF = object frequency; WF = word frequency; PC = principal component. Significance in bold for $p < .05$.

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 processing 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 (Richter et al., 2018; Summerfield et al., 2008), 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 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 (Konkle et al., 2010; for more detailed discussions please see "Object frequency and interference" p. 29 in the online supplemental materials 12).

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 "Method" 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 CD;

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 ($r = .43$), 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 formula in the online supplemental materials 12). Despite this new model being more complex in terms of number of parameters, it showed a significantly better fit than the original model ($\chi^2 = 36.691$, $p = .004$; no multicollinearity detected: variance inflation factors <5).

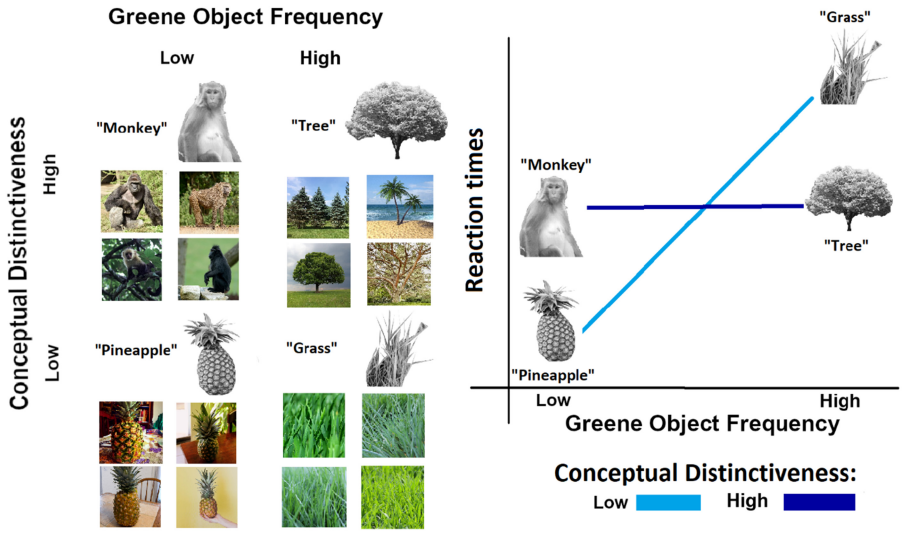
In the new model including CD, results showed that the interaction between Greene OF and CD was stronger in cross-modal matching than in uni-modal matching trials ($\beta = -0.010$, $SE = 0.003$, $t = -3.139$, $p = .002$), while no difference of the Greene OF by CD interaction was found between cross-modal mismatching and uni-modal mismatching trials ($\beta = 0.002$, $SE = 0.003$, $t = 0.495$, $p = .621$). Additionally, the Greene OF by CD interaction was found to be stronger in cross-modal matching than in cross-modal mismatching trials ($\beta = -0.012$, $SE = 0.003$, $t = -3.774$, $p < .001$; for more details, see Table 15 in the online supplemental materials 12). As shown in Figure 5, in cross-modal matching trials, higher CD was associated with weaker Greene frequency effects (the slope reduced toward 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.

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 memory processing. Crucially, this interference for higher number of exemplars of an object was reduced for object categories with higher CD (i.e., categories whose members were more easily distinguishable into subgroups of different kinds; Konkle et al., 2010). Therefore, when object categories have low CD (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 processing (many occurrences = strong interference, few occurrences = weak interference); however, when object categories have high CD (i.e., it is easy to divide their exemplars in subcategories), the number of occurrences of an object does not influence mental processing 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

Figure 4

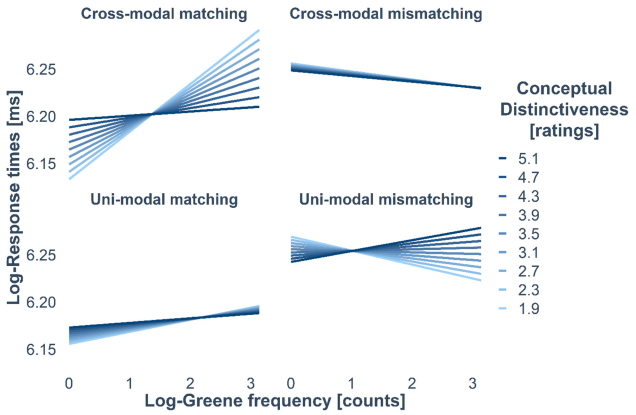
Example of Interaction Between Greene Frequency and Conceptual Distinctiveness



Note. An example of the hypothesized interaction, using object concepts from our stimulus set. Concepts are shown as black and white pictures and the associated written words. Pictures were taken and adapted from the THINGS dataset (Hebart et al., 2019). These were not part of the actual experiment. See the online article for the color version of this figure.

Figure 5

Results Interaction Between Greene Frequency and Conceptual Distinctiveness



Note. 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). Response times 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 CD (line colors: lighter = low CD, darker = high CD), in different experimental conditions (top–bottom–left–right panes). CD = conceptual distinctiveness; OF = object frequency. See the online article for the color version of this figure.

analysis of Greene OF in relation to CD of object categories might offer new, valuable insights on both these interpretations.

Similar to Konkle et al. (2010) we found impaired performance for object categories that are encountered in more exemplars (higher OF) compared to object categories that are encountered in less exemplars (lower OF). And like Konkle et al. (2010), when 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 CD 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 CD, 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.

Results: Experiment 3

Experiment 3 was a large-scale replication of the priming experiment (Experiment 2), intending to reduce potential cross-experiment carry-over effects while minimizing 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, for example, 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 reduced the number of repetitions from 18 times to eight by including two new groups of participants, each of which performed either the cross-modal or uni-modal priming tasks (between-participants design).

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.

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 ($\beta = -0.010$, $SE = 0.005$, $t = -2.165$, $p = .030$); however, the interaction of SUBTLEX, Matching condition, and Priming condition was not significant ($\beta = 0.009$, $SE = 0.005$, $t = 1.880$, $p = .060$), but qualitatively in the same direction as in Experiment 2 (see Table 4 and Figure 6). 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; no multicollinearity detected: variance inflation factors <5, for more details on the level of collinearity, see Table 16 in the online

Table 4
Results Main Model of Experiment 3

Predictors	β	*SE*	*t*	*p*
(Intercept)	6.354	0.017	374.088	**<.001**
Greene OF	0.003	0.002	1.661	.097
Matching condition (mismatch – match)	0.062	0.002	27.489	**<.001**
Target modality (words – objects)	−0.002	0.002	−0.896	.370
Priming condition (cross-modal – uni-modal)	0.014	0.033	0.413	.680
SUBTLEX WF	−0.008	0.002	−3.932	**<.001**
Visuo-orthographic PC	0.005	0.002	2.489	**.013**
Concept familiarity (replication)	−0.001	0.002	−0.347	.729
Image typicality (replication)	−0.009	0.002	−5.229	**<.001**
Image visual PC1	0.001	0.002	0.633	.527
Image visual PC2	0.003	0.002	1.929	.054
Image visual PC3	0.001	0.002	0.592	.554
Target repetition	−0.036	0.001	−31.062	**<.001**
Trial accuracy (correct – incorrect)	0.013	0.007	1.954	.051
Greene × Matching Condition	−0.008	0.002	−3.529	**<.001**
Greene × Target Modality	0.001	0.002	0.371	.711
Matching Condition × Target Modality	0.001	0.004	0.195	.845
Greene × Priming Condition	0.012	0.002	5.158	**<.001**
Priming Condition × Matching Condition	−0.016	0.004	−3.523	**<.001**
Priming Condition × Target Modality	−0.029	0.005	−6.285	**<.001**
SUBTLEX × Matching Condition	0.013	0.002	5.638	**<.001**
SUBTLEX × Target Modality	−0.011	0.002	−4.830	**<.001**
SUBTLEX × Priming Condition	−0.010	0.002	−4.291	**<.001**
Greene × Matching Condition × Target Modality	−0.004	0.005	−0.917	.359
Greene × Matching Condition × Priming Condition	−0.010	0.005	−2.165	**.030**
Greene × Priming Condition × Target Modality	0.000	0.005	0.019	.985
Matching Condition × Priming Condition × Target Modality	0.030	0.009	3.406	**.001**
SUBTLEX × Matching Condition × Target Modality	0.016	0.005	3.399	**.001**
SUBTLEX × Matching Condition × Priming Condition	0.009	0.005	1.880	.060
SUBTLEX × Priming Condition × Target Modality	−0.006	0.005	−1.230	.219
Greene × Matching Condition × Priming Condition × Target Modality	−0.011	0.009	−1.169	.242
SUBTLEX × Matching Condition × Priming Condition × Target Modality	0.021	0.009	2.269	**.023**

Note. OF = object frequency; WF = word frequency; PC = principal component. Significance in bold for $p < .05$.

supplemental materials 13; for similar results including participants' demographics see Table 17 in the online supplemental materials 13).

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: $\beta = -0.014$, $SE = 0.003$, $t = -4.324$, $p < .001$; Greene OF: $\beta = 0.017$, $SE = 0.003$, $t = 5.165$, $p < .001$). Again, no difference was found for the frequency effects in mismatching trials between cross-modal and uni-modal priming (SUBTLEX WF: $\beta = -0.006$, $SE = 0.003$, $t = -1.664$, $p = .097$; Greene OF: $\beta = 0.006$, $SE = 0.003$, $t = 1.959$, $p = .050$; see Table 18 and Figure 10 in the online supplemental materials 14). 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 Experiment 2: −0.023; Beta in Experiment 3: −0.014), while the difference of effects of the Greene OF between the two conditions was similar to Experiment 2 (Beta in Experiment 2: 0.018; Beta in Experiment 3: 0.017).

Again, the strongest frequency effects were found in cross-modal matching trials. With less trials per person, only the SUBTLEX frequency effect was significant ($\beta = -0.013$, $SE = 0.006$, $t = -2.253$, $p = .024$), while the Greene OF effect was not ($\beta = 0.010$, $SE = 0.005$, $t = 1.873$, $p = .061$). 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; see Table 19 and Figure 11 in the online supplemental materials 14).

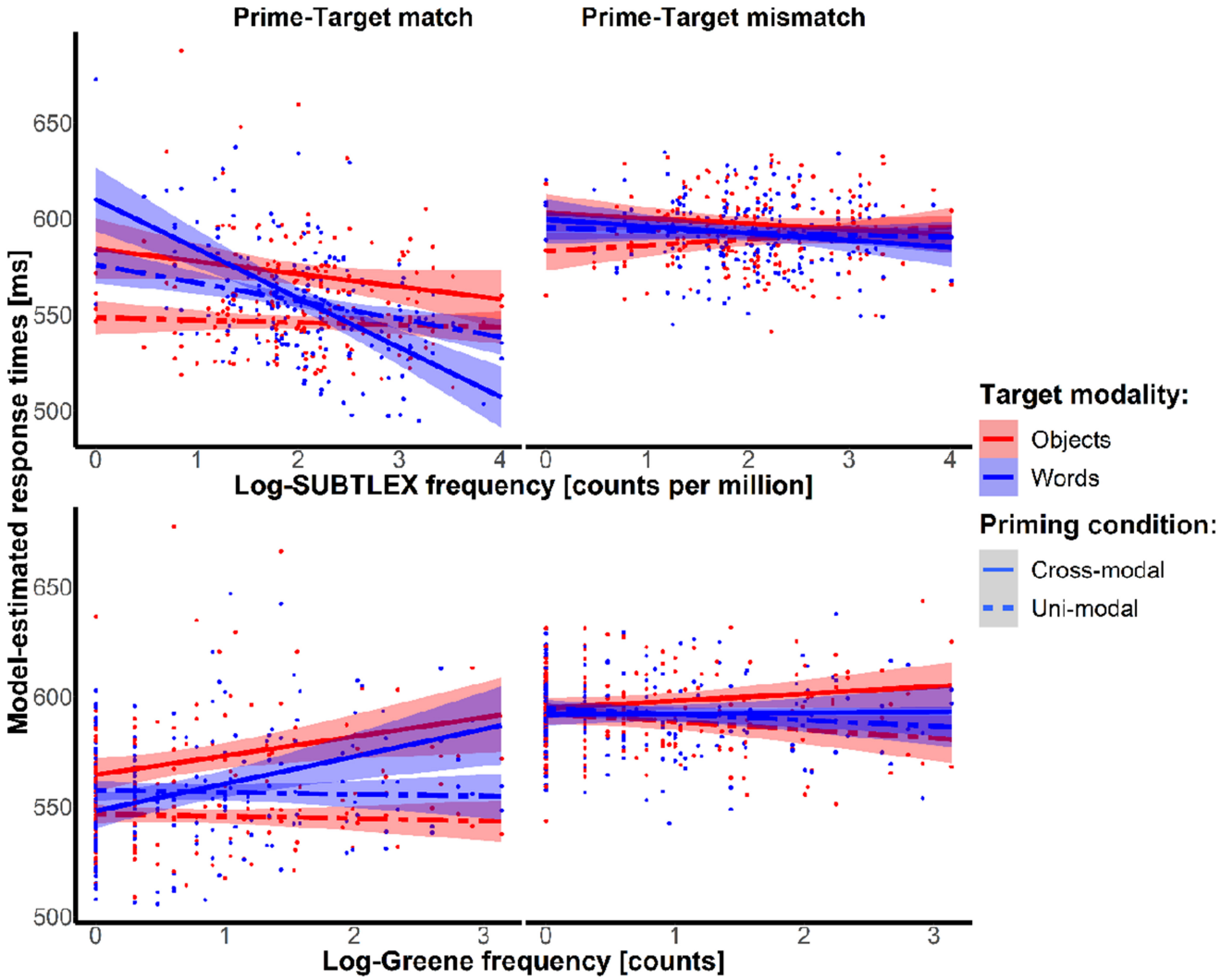
Contrary to Experiment 2, we found a significant interaction involving SUBTLEX, Matching condition, Priming condition, and Target modality ($\beta = 0.021$, $SE = 0.009$, $t = 2.269$, $p = .023$), which indicates a different modulation of words and objects as a function of SUBTLEX WF. Post hoc investigations found that the difference of SUBTLEX WF effects between cross-modal and uni-modal matching trials was stronger for words than for objects ($\beta = -0.016$, $SE = 0.007$, $t = -2.401$, $p = .016$; see Table 20 and Figure 12 in the online supplemental materials 14).

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

Figure 6
Main Results of Experiment 3



Note. Response times as a function of logarithmic SUBTLEX frequency (top plots) and Greene frequency (bottom plots) in the different conditions of Experiment 3; response times were estimated based on the selected model. Points present participant-based mean response times separated for stimulus type (light gray: object stimuli; dark gray: 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. See the online article for the color version of this figure.

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 preactivation 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 WF effect, as expected, reliably occurs across experiments, while the OF effect seems to be more volatile. With regard to the familiarity and typicality ratings, it does not seem to have been the case that having the same participants rate the stimuli they had seen during the experiment (as in Experiment 2) altered the frequency effects. When using the ratings from an independent sample to model the data of Experiment 2 instead of Experiment 3, the patterns of frequency effects or the effects of the considered covariates did not significantly change (this held truth also for data of Experiment 1).

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 uni-modal matching trials, supporting our hypothesis that frequency effects are strengthened by deeper semantic processing reflecting aspects of conceptual representation; and (b) a four-way interaction between SUBTLEX × Matching Condition × Priming Condition × 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 CD). Finally, we provide insights suggesting that these two frequency 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 WF 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; for results using this alternative measure see Tables 21 and 22 in the online supplemental materials 15). The novel aspect here is that contrary to Taikh et al. (2015), a word-based frequency measure was also found to influence object recognition. Crucially, their study 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 CD (a measure of category width 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. 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, since they better reflect linguistic experience of typical participants of most psycholinguistic studies (i.e., young university student; Brysbaert et al., 2011). Moreover, this measure is interpreted not just as capturing the strength of experience with a word (effect in word recognition), but the strength of experience with a concept/word meaning (effect in both object and word recognition) built during linguistic experience. Indeed, as we controlled for many perceptual and linguistic variables, the SUBTLEX WF effect seems to indeed reflect access to semantic representation required by the tasks.

We could speculate that this strengthening is initially established through linguistic experience and afterwards transfers to other nonlinguistic 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). This role of language in shaping the semantic changes reflected by the SUBTLEX effect might also explain why, despite the multimodality of SUBTLEX effect, it has shown stronger effects with words than with objects in Experiment 1 and Experiment 3.

The fact that dlexDB, contrary to SUBTLEX, only showed an effect in word recognition trials and not in object recognition trials (see Tables 21 and 22 in the online supplemental materials 15), suggests that the SUBTLEX WF effect for object recognition is unlikely due to the automatic co-activation of lexical representation during object processing, since otherwise we would have expected to see a dlexDB WF effect also in object recognition.

To address this issue furthermore, and discuss why SUBTLEX effect in Experiment 1 (and partially in Experiment 3) is stronger for word recognition trials than for object trials, consider the processes involved in visual word recognition and object recognition. In visual word recognition, access to meaning and semantic processing are mediated by visual processing, then orthographic processing and phonological processing (Coltheart et al., 2001). For object recognition, no such orthographic and phonological mediation is present between visual and semantic processing (Riesenhuber & Poggio, 2000). The stronger effect of SUBTLEX in word trials might therefore reflect the involvement of two processes captured by this measure: the experience with a word, reflected in

orthographic/phonological processing, and the experience with its meaning, reflected in semantic processing); for object recognition, on the other hand, the weaker SUBTLEX effect could have resulted from only capturing the involvement with semantic processing. These two aspects of linguistic experience (orthographic/phonological and semantic) which are successfully captured by SUBLTEX do not seem to be captured by dlexDB which instead mainly reflects orthographic/phonological experience, therefore resulting in weaker ecological validity as discussed above (Brysbaert et al., 2011, 2018).

Greene Object Frequency Effect and Its Relationship With Structure of Object Categories

In contrast to the SUBTLEX-based WF 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, that is, counting occurrences in a dataset and capturing properties of the world. 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, that is, 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 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 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 ($r = .43$), 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. Finally, this interaction between OF and CD, and the role of CD in predictive processing, seems to offer an explanation of why OF emerged only in Experiments 2 and 3 and not in the categorization task of Experiment 1.

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 (Gagl et al., 2020; Wang & Maurer, 2020; Zhao et al., 2019).

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 et al. (2017), where an auditory prime preactivated a representation of a previously matched visual stimulus before its presentation (Kok et al., 2017). Furthermore, in line with our results (i.e., preactivation facilitation based on frequency), they found that the preactivation 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 uni-modal 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 et al. (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 WF measure, the SUBTLEX-based 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 "Method" section). 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 OF (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 OF 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 $r = .81$ and led to similar results, see Table 23 in the online supplemental 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: (a) 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 versus uncommon labels; (b) we have shown that the Greene OF effect is present only for concepts with low CD, which have a more homogeneous set of exemplars and thus should be less prone to be biased by possible variabilities in labelling; and (c) 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 Table 23 in the online supplemental 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 WF 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 WF effect in visual word recognition is a well-established empirical finding, while there is little evidence about 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 WF measures. We, therefore, applied new metrics of OF based on state-of-the-art datasets of annotated images and evaluated them in comparison to widely used metrics of WF. 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* [Stat.]. http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4* [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), Article 4081. https://doi.org/10.1038/s41467-021-24368-2

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. https://doi.org/10.1101/2021.08.13.456197

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 Kŭera 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. https://doi.org/10.3758/BRM.41.4.977

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

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

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

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. *The 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. *Eneuro*, *6*(2), Article ENEURO.0321-18.2019. 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*, *59*(3), Article e13970. https://doi.org/10.1111/psyp.13970

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*, Article 116727. https://doi.org/10.1016/j.neuroimage.2020.116727

Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Psychology*, *4*, Article 777. 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* (Vol. 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), Article e0223792. https://doi.org/10.1371/journal.pone.0223792

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. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

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. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America*, *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), 802–823. https://doi.org/10.1037/a0030859

Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, *38*(6), 1842–1845. https://doi.org/10.1109/18.165464

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (Vol. 8). Oxford University Press on Demand.

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), 375–407. https://doi.org/10.1037/0033-295X.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 P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (pp. 259–268). Springer.

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. https://doi.org/10.1023/A:1011139631724

Olkkonen, M., Aguirre, G. K., & Epstein, R. A. (2017). Expectation modulates repetition priming under high stimulus variability. *Journal of Vision*, *17*(6), Article 10. https://doi.org/10.1167/17.6.10

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

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), Article 2. https://doi.org/10.16910/jemr.2.5.2

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *The Journal of Neuroscience*, *38*(34), 7452–7461. https://doi.org/10.1523/JNEUROSCI.3421-17.2018

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, *3*(S11), 1199–1204. https://doi.org/10.1038/81479

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*(1–3), 157–173. https://doi.org/10.1007/s11263-007-0090-8

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), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Shannon, C. E. (1948). A mathematical theory of communication. *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. https://doi.org/10.1038/nn.2163

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*. https://doi.org/10.1016/j.neuropsychologia.2019.107299

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630

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. https://doi.org/10.1111/j.1745-6924.2009.01127.x

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), Article e3000474. https://doi.org/10.1371/journal.pbio.3000474

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

Received May 17, 2021
Revision received November 2, 2022
Accepted November 9, 2022 ■