bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Conceptual knowledge affects early stages of visual mental imagery and object perception

Martin Maier[1,2], Romy Frömer[1,3], Johannes Rost[1], Werner Sommer[1,4], and Rasha Abdel Rahman[1,2]

[1]Department of Psychology, Humboldt-Universität zu Berlin

[2]Cluster of Excellence Science of Intelligence, Technische Universität Berlin

[3]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University

[4]Department of Psychology, Zhejiang Normal University

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Author Note

Martin Maier https://orcid.org/0000-0003-4564-9834

Romy Frömer https://orcid.org/0000-0002-9468-4014

Werner Sommer https://orcid.org/0000-0001-5266-3445

Rasha Abdel Rahman https://orcid.org/0000-0002-8438-1570

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – project number 390523135 and DFG grant AB 277-6 to Rasha Abdel Rahman. Martin Maier was supported by the state of Berlin and the Berlin School of Mind and Brain.

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Martin Maier, Department of Psychology, Rudower Chaussee 18, 12489 Berlin, Germany. Email: martin.maier@hu-berlin.de.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

When we imagine an object and when we actually see that object, similar brain regions become active. Yet, the time course and mechanisms with which imagery engages perceptual networks remain to be better understood. An emerging view holds that imagery and perception follow distinct dynamics during early visual processing with similarities arising only during later, high-level visual processing. However, confounds of visual stimulation and paradigms favoring observation of high-level processes associated with subjective imagery strength may have precluded evidence of earlier shared mechanisms. We therefore manipulated prior knowledge that informs early-stage top-down predictions and tracked electrophysiological brain responses while fully controlling visual stimulation. Participants saw and imagined objects associated with varying amounts of semantic knowledge. Imagery and perception were equally influenced by knowledge at an early stage (P1 component), revealing shared mechanisms that support low-level visual processing. This finding complements previous research by showing that imagery is not merely perception in reverse. Instead, in line with the predictive processing framework, both, perception and imagery are active and constructive processes that share top-down mechanisms already in early phases of processing.

Keywords: mental imagery, top-down effects, early visual processing, event-related potentials, semantic knowledge

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Conceptual knowledge affects early stages of visual mental imagery and object perception

A substantial body of research suggests that seeing something with the mind's eye—mental imagery—recruits similar neural circuits as seeing something with one's physical eyes (Albers et al., 2013; Cichy et al., 2012; Dijkstra et al., 2019; Kosslyn, 2005; Kosslyn & Thompson, 2003; O'Craven & Kanwisher, 2000; Pearson, 2019; Stokes et al., 2009). This accords well with the "inside-out" view of perception proposed in the Bayesian brain-framework: In this view, perception draws on top-down predictions about incoming input that are weighed with and updated by bottom-up sensory signals. Predictions are based on a set of priors constituting an internal generative model of the external world that incorporates previous experience and knowledge (A. Clark, 2013; Friston, 2009, 2010; Hohwy, 2013; Seth, 2015). Mental imagery is initiated and sustained by top-down processing, in the absence of relevant sensory input (Dijkstra et al., 2019; Pearson, 2019). The overlap in recruited cortical networks could be explained by the notion that mental imagery largely shares the top-down aspects of perception. Differences in neural dynamics and phenomenological experience would then arise from the fact that imagery lacks the bottom-up processes propagating sensory signals that update perceptual predictions.

While overlap with perception regarding recruited cortical areas is well established, the temporal dynamics of the neurocognitive processes supporting mental imagery are less clear. According to one hypothesis, imagery may follow a reverse hierarchy compared to perception, suggesting that perception is a bottom-up process going from simple to complex visual representations, whereas imagery starts from high-level representations (e.g., of full objects) and essentially works its way down to more visual detail (Dijkstra et al., 2019; Pearson, 2019). Accordingly, during imagery, involvement of early visual cortex would become relevant only at a later stage. Supporting this view, recent studies suggested that imagery and perception show distinct neural activation patterns during early visual processing stages, which become more similar during high-level processing

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(Dijkstra et al., 2019; Dijkstra et al., 2018; Dijkstra, Zeidman, et al., 2017; Pearson, 2019).

However, the apparent distinct dynamics during early visual processing may result from confounds that preclude unequivocal evidence in favor of or against similarities between imagery and perception at early processing stages. First, some evidence relies on direct comparisons between imagery and perception conditions, in which visual stimulation differed substantially (e.g., Dentico et al., 2014; Dijkstra et al., 2018). Without proper controls, differences in stimulation, e.g., presenting a complex scene vs. an empty frame, trivially explain distinct neural dynamics. So far, the way to solve this visual stimulation confound has been to compare neural correlates of strength of mental imagery, for instance, how vividly a mental image appears to a participant in a given trial (Albers et al., 2013; Cui et al., 2007; Dijkstra, Bosch, et al., 2017, 2019; Fulford et al., 2018; S.-H. Lee et al., 2012). In our opinion, however, this introduces a second potential confound that specifically impairs observation of early neurocognitive mechanisms supporting imagery. The initial stages of vivid and less vivid imagery should be very similar, because in both cases participants attempt to imagine an object with the same starting conditions, e.g., in terms of initial top-down predictions. The strength of the resulting conscious mental image should depend mostly on relatively late visual processing stages that are associated with the stabilization of percepts in visual awareness and involve global recurrent processing including the frontal-parietal network (Lamme & Roelfsema, 2000; Mashour et al., 2020). These two issues can explain why imagery and perception appear to follow strictly distinct dynamics during early stages and why decodable neural patterns associated with mental imagery appear to emerge only during relatively late, high-level visual processing.

Here, we aim to complement current models by exploring early stages of mental imagery in a new way, more in line with the predictive processing view of perception and the idea that imagery essentially shares the top-down aspects of perception. The Bayesian brain-framework and previous empirical evidence contradict the notion of purely "bottom-up perception" (A. Clark, 2013; Lupyan et al., 2020; Lupyan & Clark, 2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Perception is subject to top-down modulations already during early visual processing (Abdel Rahman & Sommer, 2008; Bar, 2004; Boutonnet & Lupyan, 2015; Collins & Olson, 2014; de Lange et al., 2018; Gandolfo & Downing, 2019; T. S. Lee & Mumford, 2003; Maier & Abdel Rahman, 2018, 2019; Press & Yon, 2019; Rabovsky et al., 2012; Samaha et al., 2018; Weller et al., 2019): For instance, semantic knowledge and conceptual-linguistic categories influence object perception as early as in the P1-component of the event-related potential (ERP), i.e. during low-level visual processing (Abdel Rahman & Sommer, 2008; Maier & Abdel Rahman, 2018, 2019; Maier et al., 2014; Rabovsky et al., 2012; Weller et al., 2019). If imagery shares the top-down aspects of perception, we hypothesize that this could hold for such early modulations as well.

Accordingly, in order to target early visual processing stages, we carefully manipulated the perceptual priors that go into imagery and perception. In a learning paradigm, we varied the depth of conceptual knowledge associated with visual objects, thus manipulating the quality of initial top-down predictions while keeping everything else constant (e.g., task, motivation, and bottom-up visual content). For the crucial comparisons, this allowed us to use the same type of stimulation (imagery or perception) and the same visual content (i.e. identical objects), and compare the influence of varying top-down predictions based on different depth of semantic object knowledge.

We combined this approach with recording and analyzing ERPs to test with high temporal precision whether early knowledge effects, repeatedly observed in perception, are mirrored in imagery. Based on previous findings (Abdel Rahman & Sommer, 2008; Rabovsky et al., 2012), we expected semantic knowledge to decrease the P1 component, a marker of sensory processing sensitive to low-level visual features such as luminance and perceived contrast (V. P. Clark et al., 1994; Di Russo et al., 2002; Foxe & Simpson, 2002; Haynes et al., 2003), as well as the N400 component, reflecting high-level semantic processing (Abdel Rahman & Sommer, 2008; Kutas & Federmeier, 2011; Rabovsky et al., 2012). We predicted that knowledge would influence both components similarly in

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

perception and imagery.

To integrate our results with previous work, we additionally report results obtained with the established approaches of comparing imagery with perception and leveraging the strength of imagery (Dijkstra, Bosch, et al., 2017; Pearson, 2019). In line with previous evidence, we assumed that direct comparisons of perception and imagery would result in distinct dynamics during early visual processing, with increasing similarities during high-level processing. We further assumed that successful and incomplete imagery would show similar activation patterns during low-level visual processing (P1 component), but differ in high-level configural visual processing (N1 component). Combining this approach with the study of early top-down processes informed by knowledge may lead to a more complete picture of the temporal dynamics of mental imagery.

Methods

Participants

Participants were 32 native German speakers (23 women, 9 men; mean age 24 years; age range 20-35). All were right-handed and reported normal or corrected-to-normal visual acuity. Two participants were replaced due to excessive EEG artifacts. The study was approved by the Ethics Committee of the Psychology Department at Humboldt-Universität zu Berlin. Participants gave written informed consent and received payment or course credits.

Apparatus and stimuli

Stimuli were presented on a 17” monitor using Presentation (Neurobehavioral Systems ®, Berkeley, USA) with a viewing distance of approximately 90 cm. The stimulus set comprised 40 rare objects (Abdel Rahman & Sommer, 2008) unfamiliar to all participants (Figure 1) and 20 well-known objects. All stimuli were gray-scale pictures of either entire objects or object fragments (used as cues), revealing about 20% of the objects, all displayed on a blue background frame of 3.5 × 3.5 cm (2.22° × 2.22° visual angle; see

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

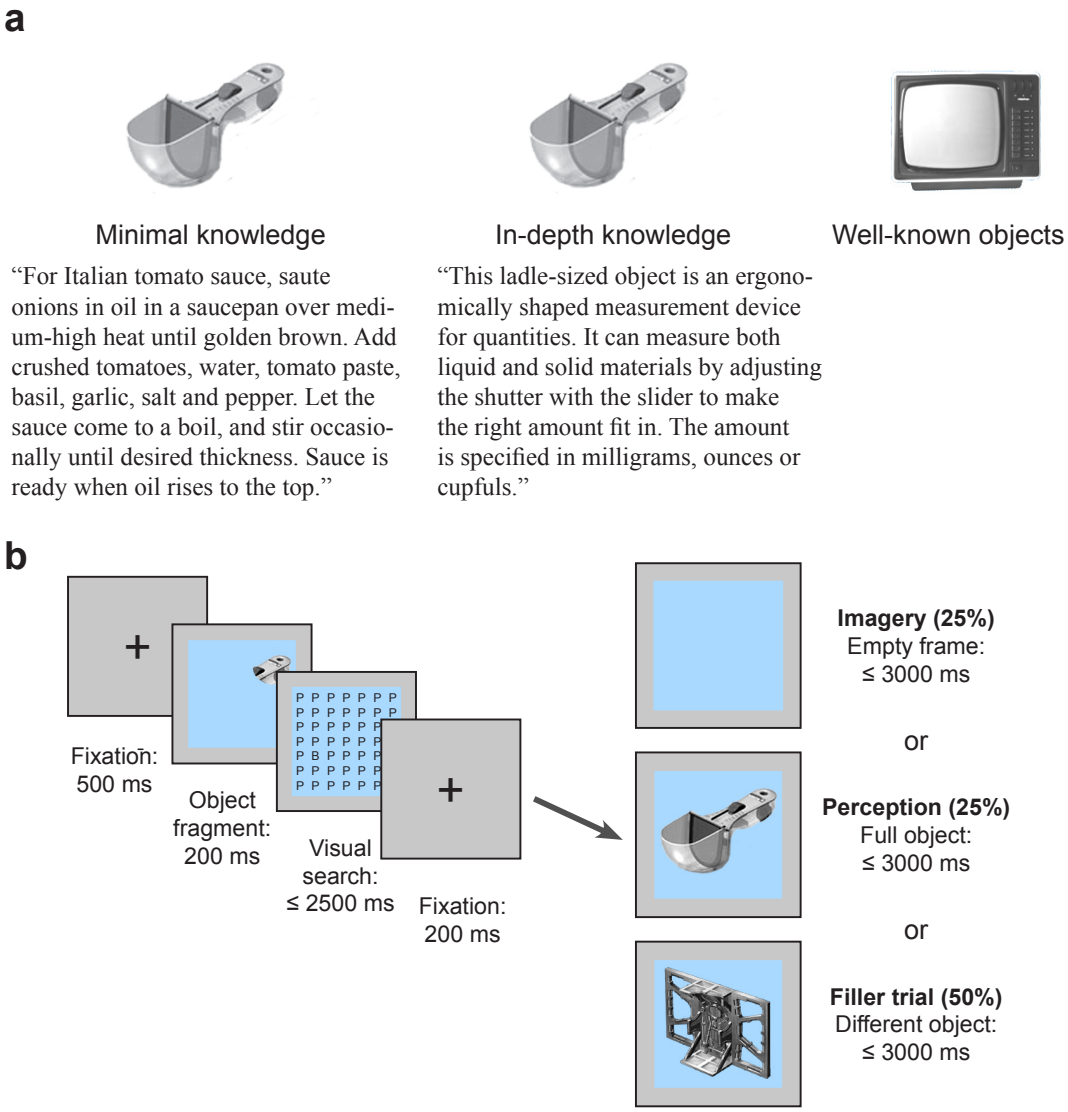

Figure 1

Study design. (a) Knowledge conditions with examples of object-unrelated information (minimal knowledge condition) and object-related information (in-depth knowledge condition). (b) Trial types and structure of the main task. All trial types occurred in all knowledge conditions (minimal, in-depth and well-known), with equal probability and in randomized order.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1). Object fragments were typical parts of the corresponding objects, allowing recognition. Fragment positions (center, left, right, top, bottom part of the object) were counterbalanced across objects. During learning, object names consisting of pseudo-nouns uninformative regarding the object's function, were presented in both written and spoken form. In addition, for each unfamiliar object, an audio description was presented containing either a short explanation of the object's function, use and origin (mean duration 18.3 s), or a cooking recipe (out of 20 recipes; mean duration 18.6 s, see Figure 1). Visual search displays consisted of a 7 by 7 matrix of uppercase letters with one single deviant letter. One of three different letter combinations (F-E, P-B, and T-L) was shown on a light blue background measuring 5 × 3.5 cm (3,17° × 2.22°). The deviant letter could appear in any position of the matrix except for the center column.

Task and procedure

All participants completed two sessions on different days: a learning session, in which they acquired semantic knowledge about unfamiliar objects, and a test session that tested imagery and perception of the learned objects along with well-known objects.

Learning Phase

The learning session consisted of two parts. In Part 1, lasting about 45 minutes, participants were presented 40 unfamiliar objects and their names (written and spoken). The first part ended with a short test (approximately 10 min), comprising verbal naming and familiarity decisions on both, well-known objects and newly learned objects. In Part 2, lasting about 75 min, participants listened to recordings that provided object-related information about origin, function and use of half of the unfamiliar objects (in-depth knowledge condition), and unrelated cooking recipes for the other half (minimal knowledge condition).

Knowledge-conditions were counterbalanced across participants, such that each unfamiliar object was equally often part of the in-depth and minimal knowledge conditions.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

All stories were presented twice. Thus, all unfamiliar objects were presented equally often and for the same duration and only object-related knowledge was manipulated. This resulted in three conditions with increasingly elaborate knowledge: newly learned objects without functional information (20 objects, minimal knowledge condition), newly learned objects with detailed information (20 objects, in-depth knowledge condition), and well-known objects, with preexisting information, visual and hands-on experience (20 objects, well-known objects condition). Part 2 ended with the same naming and familiarity test as Part 1.

Test Phase

The test session, which included EEG recordings, took place two to three days after the learning session. Before the experiment, participants filled in a knowledge questionnaire, testing recall of the pictures and related information of newly learned and well-known objects. Then, participants were familiarized with the object fragments, to make sure they could recognize the corresponding objects. Before the main task, participants performed a practice block with five well-known objects (not part of the test set), which was repeated up to two times if necessary.

In the main task, participants either imagined or saw pictures of objects. Investigating imagery with ERPs bears some timing-related difficulties: the content of imagery must be cued, but cue processing should not overlap in time with imagery, and the precise onset of imagery should be controlled. Furthermore, effects of object-knowledge on neural processing should be related to imagery, not processing of the cue.

We designed a task to control the onset and content of imagery (Figure 1). First, an object fragment was presented as cue, followed by a demanding visual search trial meant to delay the onset of imagery by taxing visual working memory (Emrich et al., 2009; Luria & Vogel, 2011; Woodman & Arita, 2011), and to avoid overlap of (semantic) processing of the cue with subsequent imagery or perception. Participants were instructed to indicate the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

position of a deviant letter in the left or right half of the display. Next, participants either saw an empty blue frame (imagery task, 180 trials, 25 %), a full picture of the cued object (perception task, 180 trials, 25 %), or a different object (filler trials, 360 trials, 50 %). Immediately after a response or if no response had been given within 3 s after stimulus onset, a blank screen of 1 s duration was presented. In the imagery task, participants were instructed to form an intact and detailed mental image of the cued object as quickly and accurately as possible. Participants indicated successful or incomplete imagery via button press. In perception and filler trials, participants indicated via button press whether the object was newly learned or well-known. In filler trials, two different non-corresponding object fragments were randomly assigned to each object per participant.

Requiring imagery only in 25% of the trials was aimed to discourage participants from initiating imagery already upon seeing the object fragment. Preparatory strategies were rendered ineffective by the frequent filler trials, in which invalid cues were shown. Response button assignments in the familiarity and mental imagery tasks were counterbalanced across participants. Trial types were presented in random order with short breaks after every 30 trials. Minimal knowledge, in-depth knowledge, and well-known object conditions were evenly distributed across tasks. At the end of the session, prototypical eye movements and blinks were recorded in a calibration procedure for ocular artifact correction.

EEG recording

The EEG was recorded from 56 Ag/AgCl electrodes placed according to the extended 10-20 system, initially referenced to the left mastoid. The vertical electrooculogram (EOG) was recorded from electrodes FP1 and IO1. The horizontal EOG was recorded from electrodes F9 and F10. Electrode impedance was kept below 5 kΩ. A band pass filter with 0.032 - 70 Hz, and a 50 Hz notch filter were applied; sampling rate was 250 Hz. Offline, the EEG was recalculated to average reference and low-pass filtered at

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

30 Hz. Eye movement and blink artifacts were removed with a spatio-temporal dipole modeling using the BESA software (Ille et al., 2002), based on the recorded prototypical eye movements and blinks. Trials with remaining artifacts and missing responses were discarded. The continuous EEG was segmented into epochs of 1.2 s locked to the stimulus of the main task (object picture or empty blue frame), including a 200 ms pre-stimulus baseline.

Experimental Design and Statistical Analysis

Statistical analyses were performed with R (Version 4.0.0. R Core Team, 2019) and the Fieldtrip toolbox (Oostenveld et al., 2011) for Matlab (Version 2016a). Data and R-code allowing to replicate the reported analyses are available at the Open Science Framework (https://osf.io/j3vt6). Trials with unsuccessful visual search or with reaction times (RTs) shorter than 150 ms or longer than 3 SDs from individual participant's means were excluded from all analyses. In addition, trials with incorrect familiarity classifications in the perception task were excluded from RT and ERP analyses. RTs were log transformed to approximate a normal distribution. Using the lme4 package (Version 1.1–23, Bates et al., 2015), accuracy and imagery success were analyzed with binomial generalized linear mixed models (GLMMs); RTs and ERPs were analyzed with linear mixed models (LMMs) (Frömer et al., 2018).

P-values were computed using the lmerTest package (Kuznetsova et al., 2016). We applied sliding difference contrasts that compare mean differences between adjacent factor levels. When indicated, we reduced models by excluding non-significant interaction terms. Model selection was performed using the anova function of the stats package in R. Along with the results of the χ^2-Test, we compared fit indices, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which are smaller for better model fit considering the number of parameters in each model. Behavioral data were analyzed using a nested model with the factor knowledge (well-known, in-depth and minimal) nested

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

within task (Imagery and Perception). Mixed model analyses included random intercepts and (if supported) random slopes for subjects and object identity, allowing for better generalization of results from the particular sample of participants and the set of object pictures used here. To specify optimal random effects structures, we used the buildmer package (Voeten, 2021), which implements an algorithm identifying the maximal random effects structure with which the model still converges and performs backward stepwise elimination in order to avoid overfitting. Following Matuschek et al. (2017), both steps were performed in a backward manner (i.e. starting with the maximal model and reducing it stepwise), and likelihood ratio tests with $\alpha = .20$ were conducted for model comparisons. LmerControl settings were specified to use the BOBYQA optimizer (bound optimization by quadratic approximation) with a set maximum of 200.000 iterations, and to turn off (time-consuming) derivative calculation after optimization is finished.

One follow-up analysis used Bayes factors to assess the evidence that knowledge effects on the P1 component are equal in imagery and perception. Following Vasishth et al. (2018), we used the brms package (Bürkner, 2017) to specify two different Bayesian linear mixed effects models: a null model M_0 not assuming an interaction and an alternative model M_1 assuming an interaction effect between task and semantic knowledge. We report the mean Bayes factor averaged from ten calculation runs, since the result can vary from run to run (Vasishth et al., 2018).

To address knowledge effects on ERPs during imagery and perception, we tested a priori hypotheses based on previous literature, that is, reduced P1 and enhanced N400 amplitudes with semantic knowledge in pre-specified regions of interest (ROIs). For the analysis of the P1 component, we averaged amplitudes within 120 to 170 ms at electrodes PO7, PO3, PO4, and PO8 (Pratt, 2011). The N400 was quantified as mean amplitude between 300 and 500 ms at electrodes PO7, PO3, PO4, PO8, O1, Oz and O2 (Abdel Rahman & Sommer, 2008; Rabovsky et al., 2012). Single trial amplitudes aggregated within ROIs and time windows were submitted to LMMs with the factors visual

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

condition (perception, imagery and incomplete imagery) and knowledge (well-known, in-depth and minimal) as fixed effects.

To relate to previous work, we also compared imagery and perception directly and leveraged differences in the strength of imagery by comparing trials with attempted but incomplete imagery and trials with successful imagery. To this end, we calculated each participant's average ERP in the perception, successful imagery, and incomplete imagery conditions across all scalp electrodes in time windows from 0 to 540 ms. Group-level statistics were based on paired-samples t-tests and corrected for multiple comparisons using cluster-based permutation tests (CBPT) across time and electrodes. The cluster forming threshold was set to $p = .05$. We report differences with corrected p-values $< .025$ as statistically significant.

Based on the hypothesis that imagery might be supported in particular by configural visual processing, we looked at the N1 component (Farah et al., 1988). N1 amplitudes were compared in a posterior ROI consisting of PO7, PO3, PO4, PO8, O1, Oz, and O2 (Pratt, 2011). To adjust for latency shifts (see Results), different time windows were used for the N1 component in perception and imagery, centered around the grand mean peak latencies: For perception, we aggregated over 170 – 210 ms and for imagery (both successful and incomplete) we aggregated over 210 – 250 ms. Frontal activity that coincided with the posterior N1 was analyzed in a ROI consisting of electrodes Fp1, Fpz, Fp2, AF3, AFz, AF4, F3, Fz, F4, FC1, FC2 (Gazzaley et al., 2008).

Results

To investigate whether perception and imagery share early top-down mechanisms and to examine their time course, we recorded EEG from 32 participants while they viewed or imagined objects with varying amounts of associated knowledge. Target objects were cued with object fragments and, following an intervening visual search task to reset visual activity, participants either made a familiarity judgment about a presented object or

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

imagined the cued object on an empty frame (see Figure 1).

Behavioral results

In the imagery task, participants were asked to form intact and detailed mental images of the cued objects. They indicated successful or incomplete imagery via button press. Overall, participants indicated successful imagery in M = 84.5 % of the trials. Imagery success rates were higher in the well-known compared to the in-depth knowledge condition (M = 89.2 % vs. 83.0 %; nested binomial GLMM: b = 0.43, z = 3.12, p = .002), but there was no difference between the in-depth and the minimal knowledge condition (M = 83.0 % vs. 81.1 %; b = 0.08, z = 0.88, p = .376; see Figure 2). Knowledge affected RTs, which gradually decreased with the depth of knowledge, indicating faster imagery for objects learned with in-depth compared to minimal knowledge (M = 1730.4 vs. 1770.2 ms; b = -0.02, t = -2.03, p = .042), and for well-known objects compared to objects with in-depth knowledge (M = 1673.7 vs. 1730.4 ms; b = -0.05, t = -2.51, p = .017).

In the perception task, participants classified object pictures as newly learned vs. well-known. Classification accuracy was lower in the well-known compared to the in-depth knowledge condition (M = 94.4 % vs. 97.8 %; b = -0.85, z = -4.02, $p < .001$) and also in the in-depth compared to the minimal knowledge condition (M = 97.8 % vs. 98.7 %; b = -0.55, z = -2.23, p = .026). RTs in the perception task (Figure 2) did not differ across knowledge conditions (in-depth - minimal: M = 819.0 ms vs. 810.9 ms; nested LMM: b < 0.01, t = 0.62, p = .534; well-known - in-depth: M = 801.0 ms vs. 819.0 ms; b = -0.02, t = -0.91, p = .371). Lower accuracy in classifying well-known objects can be explained by context effects: Participants were to classify well-known objects as “old”, but these objects had been rare during the learning session, thus in the context of the test session they were “new”. In contrast, participants were to classify newly learned objects as “new”, but in the context of the experiment these objects had been seen many times, and objects associated with richer semantic knowledge may have seemed subjectively more familiar, and thus

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

"old".

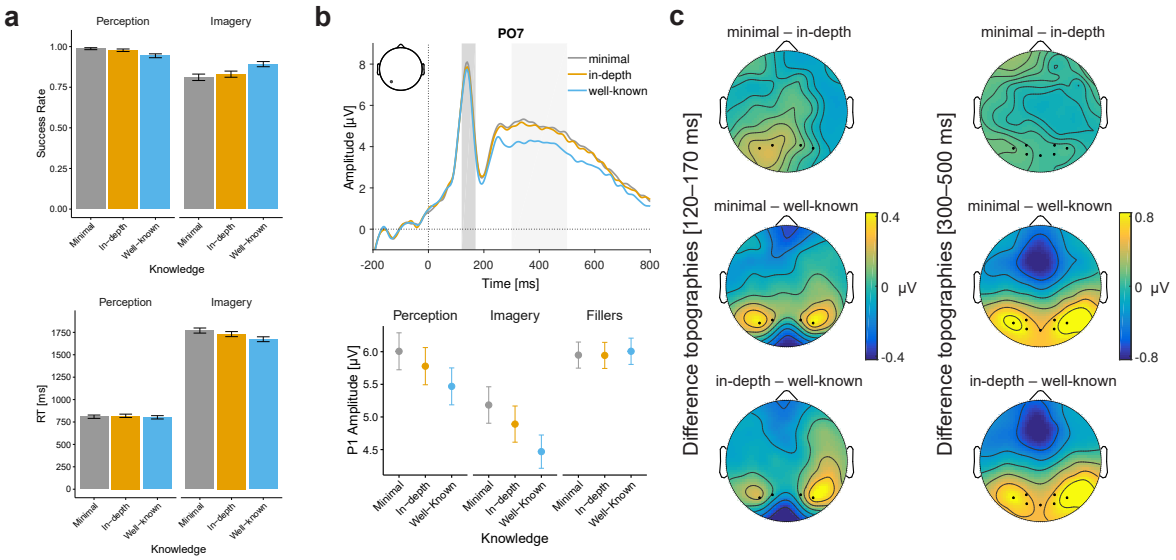

Figure 2

Semantic knowledge effects. (a) Behavioral results: Accuracy in the perception task and imagery success rate (top) and mean RTs (bottom) as a function of object knowledge. Error bars represent 95% confidence intervals. (b) Effects of object knowledge on the P1 and N400 components. Top panel: Grand average ERPs at electrode PO7, aggregated over perception and imagery. Bottom panel: mean P1 amplitudes by task, illustrating comparable knowledge effects in imagery and perception, and absence of the effect in filler trials. (c) Difference topographies comparing the knowledge conditions in the P1 and N400 time windows (120-170 ms; 300-500 ms, respectively). ROI electrodes are marked as dots.

Effects of semantic knowledge on ERPs

To test the hypothesis that imagery and perception share knowledge-related modulations of early visual activity, we analyzed the effects of semantic knowledge on the P1 component, an index of early perceptual processing. We further tested for later effects of knowledge in the N400, an indicator of semantic processing. In line with our hypothesis, across imagery and perception, P1 amplitudes decreased with semantic knowledge, yielding

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 1

Knowledge effects on the P1 and N400 components during perception and mental imagery.

	P1 component				**N400 component**			
Predictors	*Estimates*	*SE*	*t-value*	*p-value*	*Estimates*	*SE*	*t-value*	*p-value*
(Intercept)	5.18 ***	0.64	8.15	<0.001	3.70 ***	0.43	8.63	<0.001
Visual (Ima–Per)	-0.90 *	0.38	-2.35	0.024	-0.65	0.35	-1.86	0.072
Knowledge (Deep–Min)	-0.25 *	0.11	-2.20	0.031	-0.17	0.11	-1.48	0.139
Knowledge (Well–Deep)	-0.33 *	0.13	-2.56	0.013	-0.46 ***	0.13	-3.66	<0.001
Random Effects	*SD*				*SD*			
Participants	3.41				2.38			
Visual (Ima–Per)	2.13				1.90			
Knowledge (Deep–Min)	0.31				0.32			
Knowledge (Well–Deep)	0.11							
Visual × Knowledge (Deep–Min)	0.80							
Object Identity	0.26							
Visual (Ima–Per)	0.66							
Residual	4.26				4.72			
Deviance	61256.683				63314.788			
log-Likelihood	-30628.341				-31657.394			
Model	P1 ~ 1 + visual + knowledge + (1 + v_2.1 + kn_2.1 + kn_3.2 + v_2.1:kn_3.2 | subject) + (1 + v_2.1 | object)				N400 ~ 1 + visual + knowledge + (1 + v_2.1 + kn_3.2 | subject)			

Note: Visual (Ima–Per) = Perception – Imagery, Knowledge (Deep–Min) = in-depth – minimal, Knowledge (Well–Deep) = well-known – in-depth
** p<0.05 ** p<0.01 *** p<0.001*

significant reductions from minimal to in-depth, and from in-depth knowledge to well-known objects (Figure 2, Table 1). The LMM including semantic knowledge, task and their interactions revealed no significant interactions of knowledge and task, suggesting similar effects of knowledge in both conditions. Exclusion of these interaction terms did not

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

significantly decrease model fit, $\chi^2(2) < 3.32$, $p > .190$, and fit indices favored the reduced models (ΔAIC_{P1}: -3, ΔBIC_{P1}: -18; ΔAIC_{N400}: 0, ΔBIC_{N400}: -15).

Additionally, a Bayes factor analysis indicated that the observed P1-data are more likely under the assumption that knowledge effects are equal in imagery and perception. The priors going into the null model M_0 assumed main effects of semantic knowledge (b = -0.2 for each, in-depth versus minimal knowledge and well-known vs. in-depth knowledge), but no interaction term. The alternative model M_1 additionally assumed an interaction effect of task × semantic knowledge (b = 0.2), i.e. predicting the absence of a knowledge effect in the imagery task. The mean Bayes factor in ten calculation runs was $BF_{01} = 5.85$ (CI = [5.43, 6.26]), which constitutes substantial evidence in favor of M_0, i.e. equal knowledge effects in imagery and perception.

Knowledge effects on the P1 in perception have been repeatedly observed in the absence of cueing (Abdel Rahman & Sommer, 2008; Rabovsky et al., 2012; Weller et al., 2019), and any visual priming in the present study could only occur partially as we only showed object fragments followed by an intervening visual search task to reset visual activity. Nevertheless, a potential concern with the task design is that knowledge effects on the P1 may reflect spillovers from the cues, unrelated to imagery or perception of the objects themselves. If this were true, we should observe knowledge effects also in filler trials, where different, non-cued objects were shown. In a control analysis we found no evidence that the knowledge condition of the object cue influenced the P1 in filler trials (see Figure 2 c). There was no difference between the well-known and the in-depth knowledge condition ($LMM_{Fillers}$: b = 0.073, t = 0.600, $p = .551$) or between the in-depth and the minimal knowledge condition ($LMM_{Fillers}$: b = -0.008, t = -0.076, $p = .939$). Thus, knowledge effects in the P1 were specific to imagining or seeing the corresponding objects.

In the N400, well-known objects produced significantly more negative amplitudes than newly learned objects, whereas the minimal and in-depth knowledge conditions did not differ (see Table 1).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

To summarize, in line with our hypothesis we found semantic knowledge effects in early visual processes across both imagery and perception: P1 amplitudes were reduced with increasing depth of object-related knowledge. This effect replicates previous findings from visual perception (Abdel Rahman & Sommer, 2008; Rabovsky et al., 2012) and extends them to imagery. Previously reported differences between minimal and in-depth conditions in the N400, reflecting high-level semantic processes (Abdel Rahman & Sommer, 2008; Rabovsky et al., 2012), were not replicated.

Comparisons between successful and incomplete imagery

To track neural activity related to imagery strength, we compared trials in which participants had indicated successful vs. incomplete imagery. The hypothesis was that incomplete compared to successful imagery may arise from failed configural processing and should thus be associated with differences in the N1 component. Since imagery may be supported by increased frontal-posterior coupling (Dentico et al., 2014; Dijkstra, Zeidman, et al., 2017), differences in frontal activity were also expected. Even though EEG scalp distributions do not translate easily to generators of activity in the brain, we hypothesized that posterior N1 effects may coincide with mirrored effects at frontal sites (Gazzaley et al., 2008). To test for global differences between successful and incomplete imagery we compared mean amplitudes with the CBPT approach, which revealed a significant difference. Underlying this difference were two clusters across electrodes and time: a posterior cluster between 228 and 392 ms, and a frontoparietal cluster between 304 and 492 ms that was slightly lateralized to the right hemisphere (Figure 3). As expected, the beginning and topography of the posterior cluster suggested a modulation of the N1 component (Figure 3).

Follow-up LMM analyses based on single trial amplitudes in an independently determined posterior ROI (see Methods) confirmed a significant difference in the N1 component. Successful imagery was characterized by a larger N1 compared to incomplete

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

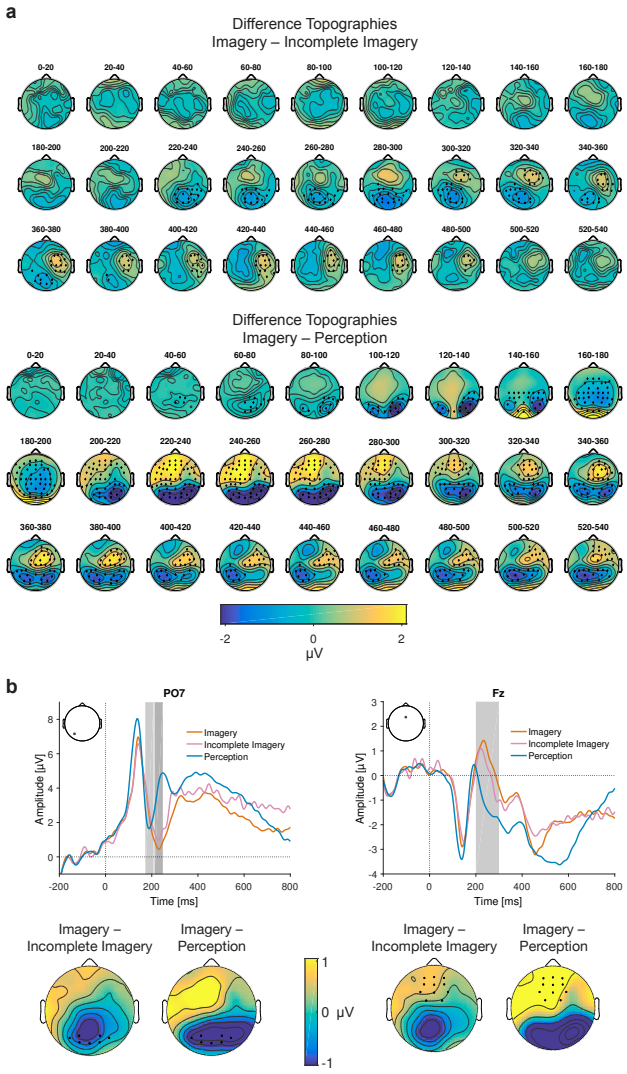

Figure 3

ERP comparisons of successful imagery, incomplete imagery, and perception. (a) Grand average difference-topographies. Highlighted electrodes are part of spatio-temporal clusters most compatible with the significant differences between successful and incomplete imagery (top), and between imagery and perception (bottom). (b) Comparisons of successful imagery, incomplete imagery, and perception in the N1 time window. Time windows entering the analysis of the posterior N1 amplitudes (left) and the simultaneous frontal positivity (right) are highlighted by grey shading. Topographies illustrate differences in the highlighted time windows with ROI electrodes marked by dots.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2

Comparisons between imagery, incomplete imagery, and perception in the N1 time window.

	N1 component				**Frontal positivity**			
Predictors	*Estimates*	*SE*	*t-value*	*p-value*	*Estimates*	*SE*	*t-value*	*p-value*
(Intercept)	1.67 ***	0.36	4.60	<0.001	0.20	0.22	0.88	0.383
Visual (Ima–Per)	-0.97 *	0.43	-2.24	0.032	0.62 **	0.21	2.92	0.006
Visual (Nima–Ima)	0.56 ***	0.15	3.78	<0.001	-0.24 *	0.12	-2.04	0.041
Centered P1	4.19 ***	0.05	79.55	<0.001				
Centered N1f					0.71 ***	0.01	81.06	<0.001
Random Effects	*SD*				*SD*			
Participants	2.03				1.22			
Visual (Ima–Per)	2.36				1.12			
Object Identity	0.21				0.12			
Visual (Ima–Per)	0.58				0.31			
Visual (Nima–Ima)	0.15							
Residual	3.94				3.12			
Deviance	59510.519				54515.483			
log-Likelihood	-29755.259				-27257.741			
Model	N1 ~ 1 + v_2.1 + v_3.2 + cP1 + (1 + v_2.1 | subject) + (1 + v_2.1 + v_3.2 | object)				P1f ~ 1 + v_2.1 + v_3.2 + cN1f + (1 + v_2.1 | subject) + (1 + v_2.1 | object)			

Note: Visual Ima–Per = Perception – Imagery, Visual Nima–Ima = Incomplete Imagery – Imagery, P1 = preceding posterior P1 component, N1f = preceding frontal N1 component
** p<0.05 ** p<0.01 *** p<0.001*

imagery (Table 2). Around the same time, successful and incomplete imagery also differed at frontal sites, with a larger positivity in the frontal ROI in successful imagery trials (Figure 3, Table 2). Thus, the comparison between successful and incomplete imagery aligns with our hypothesis that successful imagery is supported by mechanisms of configural processing indexed by the posterior N1 and potentially supported by frontal

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

top-down regulation.

To track the neural dynamics that dissociate imagery and perception, we compared these conditions using the same approach. CBPT revealed significant differences between perception and imagery. Starting with a relative negativity for imagery at parieto-occipital sites around 80 ms post stimulus, all subsequent time windows yielded significant clusters (Figure 3). As outlined above, early differences between imagery and perception are most likely due to differences in visual stimulation. Further, differences between imagery and perception could be driven by latency shifts, amplitude differences, or both. To approximate a fair comparison, we analyzed the key ERP components—P1 and N1—in the different visual conditions at their respective peak latencies (perception, successful and incomplete imagery as one factor). P1 and N1 peak latencies were detected in the average ERP at PO7 for each participant and condition.

The LMM analysis of N1 amplitudes was adjusted for latency shifts (time windows are highlighted in Figure 3). To mitigate the influence of the different visual stimulation in imagery and perception, we included centered trial-by-trial P1 amplitudes as a covariate. This can serve as a kind of baseline correction (Alday, 2019) because the P1 should capture a large portion of the variance related to differences in visual input and correct for amplitude differences resulting from evoked amplitude variance. The N1 was significantly larger (more negative) for successful imagery compared to perception (Table 2) and increased with more positive P1 amplitudes. Thus, the difference between perception and imagery found in the overall CBPT analysis appears to be driven by both, latency and amplitude differences.

Like for the comparison between successful and incomplete imagery, there was a modulation at frontal sites, where we found a larger positivity for imagery compared to perception coinciding with the posterior N1 component (Figure 3, Table 2). The frontal positivity further increased with more positive amplitudes of the preceding frontal negativity, which we controlled in order to account for earlier visually evoked differences.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

To summarize, this second approach replicated previous work by showing distinct temporal dynamics between imagery and perception. These could largely be driven by different stimulation conditions, but latency-shifts have also added to amplitude differences. Again in line with previous studies, successful and incomplete imagery started to differ only during high-level visual processing stages. We found a larger posterior N1 for successful compared to incomplete imagery and for imagery compared to perception. These effects were accompanied by modulations of a frontal positivity in the approximate time range of the N1, which was significantly enhanced for successful compared to incomplete imagery, as well as for imagery compared to perception. Taken together these findings indicate increased demands on configural processing in imagery compared to perception, potentially supported by increased recruitment of frontal top-down processing, and that imagery fails if these increased demands are not met.

Discussion

It is now widely accepted that visual perception and mental imagery rely on shared brain circuits, including regions in early visual cortex, as well as a network of frontal, parietal and temporal regions (Kosslyn, 2005; Pearson, 2019). Yet, the time course of imagery and the timing of the involvement of early visual cortex are still open questions. In line with predictive processing accounts we assumed that perception engages top-down predictions during low-level processing (Maier & Abdel Rahman, 2018, 2019; Maier et al., 2014; Rabovsky et al., 2012; Samaha et al., 2018; Weller et al., 2019), and hypothesized that imagery shares this mechanism.

The present work extends previous accounts holding that imagery works like perception in reverse—which assumed distinct temporal dynamics and that imagery does not rely on early perceptual representations (Dentico et al., 2014; Dijkstra et al., 2019; Dijkstra et al., 2018; Pearson, 2019). This account was supported by work showing similarities of brain activity between imagery and high-level perception (Dijkstra et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2018), and by imagery-related effects at the level of configural processing, as reflected in the N1 component of the ERP (Farah, 1985; Farah et al., 1988; Ganis & Schendan, 2008; Suess & Abdel Rahman, 2015). Thus, late involvement of early visual areas has been mainly supported by a lack of evidence for early involvement. Such evidence may have not been obtained, however, due to visual input confounds when comparing imagery and perception (Dijkstra et al., 2018; S.-H. Lee et al., 2012) and imagery strength manipulations favoring observation of high-level visual dynamics related to the stabilization of conscious percepts.

To overcome these obstacles, we varied the amount of knowledge associated with objects that participants saw and imagined—that is, we manipulated initial top-down predictions while controlling for bottom-up input and visual content. Using this approach, we showed that like in perception, semantic knowledge modulates low-level visual activity also during imagery, revealing similar mechanisms at a much earlier stage than previously assumed.

In line with previous evidence, we further showed that successful imagery is characterized by increased activity during high-level configural visual processing compared to both, incomplete imagery as well as perception. This suggests that demands on configural processing are higher in the absence of supporting bottom-up input. Thus, it appears that stable, high-level visual representations need to be constructed during imagery much like during perception.

Knowledge facilitates imagery and shapes early stages of imagery and perception

In imagery, like in perception, object-related knowledge and familiarity influenced visual processing at an early stage. Deeper knowledge was associated with decreases in the amplitude of the P1 component that reflects low-level visual processing in extrastriate visual areas (V. P. Clark et al., 1994; Di Russo et al., 2002; Foxe & Simpson, 2002; Haynes et al., 2003). The influence of knowledge on the P1 in particular, with its known functional

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

significance and localization, demonstrates that some imagery-related processes recruit early visual areas already at an early latency. Object knowledge appears to inform top-down predictions that are used in both tasks. We conclude that knowledge about an object's function and its relevant parts facilitates low-level feature processing when we see or when we imagine that object. This novel finding suggests that imagery and perception rely on shared top-down mechanisms acting on the early construction of low-level visual representations.

Notably, the influence of semantic knowledge on imagery was of behavioral relevance: imagery of well-known objects was more often successful and faster than imagery of less familiar objects. Additionally, imagery was faster when participants had acquired in-depth rather than only minimal knowledge about initially unfamiliar objects. Thus, the more we know about an object, the better and more quickly we can imagine it.

While the P1 component in the imagery condition was evoked by a visual stimulus, the presentation of a light blue square, this physical stimulus was identical in all semantic knowledge conditions (across participants), and hence cannot have produced the observed knowledge effects. A potential objection is that modulations of early visual ERPs might not have been related to imagery, but to spillovers from the object fragment cue. However, this explanation can be ruled out because 1) the same semantic knowledge effects on perception have been shown in the absence of cueing (Abdel Rahman & Sommer, 2008; Rabovsky et al., 2012; Weller et al., 2019), 2) we only presented fragments of the objects to be imagined or perceived, 3) visual input was reset by an intervening visual search task, and, crucially, 4) there were no cue-related knowledge effects for filler trials. We therefore conclude that the influence of object knowledge on low-level visual processing is a meaningful part of imagery and perception alike.

At variance with our predictions, we did not observe an influence of in-depth versus minimal semantic knowledge on the N400 component. In contrast to previous studies demonstrating these effects in perception (Abdel Rahman & Sommer, 2008; Rabovsky

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

et al., 2012), here, we cued the objects, which likely triggered object recognition and semantic processing. Whereas the intervening visual search task interfered with visual working memory, higher-level semantic network activation of the current object might have been sustained, given that it was potentially relevant for the upcoming task. Since the N400 is typically smaller for expected stimuli and reflects changes in semantic network activation (Rabovsky et al., 2018), the cues in our paradigm may have muted potential N400 effects.

What distinguishes successful imagery from incomplete imagery and from perception?

In line with the notion that imagery relies strongly on high-level configural visual processing, successful and incomplete imagery started to diverge in the posterior N1 component (Busey & Vanderkolk, 2005; Costa et al., 2018; Milivojevic et al., 2008; Rossion & Jacques, 2011; Tanaka & Curran, 2001). Successful imagery was associated with larger posterior N1 amplitudes, accompanied by modulations of frontal activity at the same latency. The former finding is consistent with previous EEG and MEG studies that showed imagery-related modulations of the posterior N1 (Farah, 1985; Farah et al., 1988; Ganis & Schendan, 2008; Suess & Abdel Rahman, 2015). In terms of its functional relevance and typical latency, the N1 effect fits well with the finding that neural representations decoded from imagery using MEG match those observed in perception around 160 ms, that is, the N1 time window (Dijkstra et al., 2018). As incomplete imagery did not differ from successful imagery in the P1, it seems to share the early low-level activations but to lack (some of) the later configural processes and top-down feedback that stabilizes the percept. The reduced frontal activity associated with incomplete imagery may thus reflect insufficient involvement of frontal areas, and their connectivity to occipitotemporal visual areas, which provide crucial top-down monitoring for imagery to be maintained (Dentico et al., 2014; Dijkstra, Zeidman, et al., 2017; Pearson, 2019). Holding intact and detailed images before the mind's eye thus seems to be supported by configural visual processing

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and large-scale connectivity including frontal and occipital areas that stabilizes and maintains visual representations (Dentico et al., 2014; Dijkstra et al., 2019; Dijkstra, Zeidman, et al., 2017).

This interpretation is further supported by our findings comparing imagery and perception. We found that the posterior N1 was both delayed and increased in imagery compared to perception. Simultaneously, frontal activity was more pronounced in imagery compared to perception. These results suggest that imagery relies more strongly on configural processing than perception, and engages more top-down control. When these additional demands are not met, imagery fails. To test whether success vs failure of imagery is all or none or reflects gradual degradation in configural processing, future studies could employ trial-by-trial vividness ratings to test whether they correspond to linear decreases in frontal and posterior activity (Dijkstra, Bosch, et al., 2017).

Interestingly, the finding that successful and incomplete imagery start to diverge only in the N1 component bears similarities to observations concerning conscious access and the visual awareness negativity (Förster et al., 2020). The preceding P1 component is not typically associated with perceptual awareness (Förster et al., 2020; Maier & Abdel Rahman, 2018; Sergent et al., 2005), and also did not dissociate between successful and incomplete imagery in the present study. Conscious perception is thought to depend on "global ignition" or recurrent processing in a widespread network of brain areas (Dehaene & Changeux, 2011; Lamme, 2018). It is therefore conceivable that differences between successful and incomplete imagery starting in the N1, as well as late, high-level visual representations decodable around 500 ms (Dijkstra et al., 2018), reflect the beginning of conscious mental imagery, not the beginning of imagery-related processing per se. Earlier imagery-related processing stages, revealed here by knowledge effects on the P1, appear to be pre-conscious, just as in perception.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Conclusion

Our results provide new insights into the time course of visual mental imagery by demonstrating that top-down influences modulate imagery already at an early stage of low-level visual feature processing. This extends previous assumptions that imagery and perception share neural substrates only for high-level visual processes. Instead, they already engage common neurocognitive mechanisms during early visual processing stages—consisting in top-down predictions, informed by knowledge stored in memory. Whether in seeing or imagining objects, our brains begin to construct what we "see" with the mind's eye with the help of what we know.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

Abdel Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin & Review*, *15*(6), 1055–1063. https://doi.org/10.3758/Pbr.15.6.1055

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology*, *56*(12), e13451. https://doi.org/10.1111/psyp.13451

Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/Nrn1476

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience*, *35*(25), 9329–9335. https://doi.org/10.1523/jneurosci.5111-14.2015

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*, *45*(4), 431–448. https://doi.org/10.1016/j.visres.2004.08.021

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525x12000477

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*(3), 170–187. https://doi.org/10.1002/hbm.460020306

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-013-0564-3

Costa, T. L., Orsten-Hooge, K., Gaudêncio Rêgo, G., Wagemans, J., R Pomerantz, J., & Sérgio Boggio, P. (2018). Neural Signatures of the Configural Superiority Effect and Fundamental Emergent Features in Human Vision. *Scientific Reports*, *8*(1), 13954. https://doi.org/10.1038/s41598-018-32289-2

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, *47*(4), 474–478. https://doi.org/10.1016/j.visres.2006.11.013

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Dentico, D., Cheung, B. L., Chang, J.-Y., Guokas, J., Boly, M., Tononi, G., & Van Veen, B. (2014). Reversal of cortical information flow during visual imagery as compared to visual perception. *NeuroImage*, *100*, 237–243. https://doi.org/10.1016/j.neuroimage.2014.05.081

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*(2), 95–111. https://doi.org/10.1002/Hbm.10010

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2017). Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dijkstra, N., Mostert, P., de Lange, F. P., Bosch, S., & van Gerven, M. A. (2018). Differential temporal dynamics during visual imagery and perception. *Elife*, *7*, e33904. https://doi.org/10.7554/eLife.33904

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Scientific Reports*, *7*(1), 5677. https://doi.org/10.1038/s41598-017-05888-8

Emrich, S. M., Al-Aidroos, N., Pratt, J., & Ferber, S. (2009). Visual Search Elicits the Electrophysiological Marker of Visual Working Memory. *PLOS ONE*, *4*(11), e8042. https://doi.org/10.1371/journal.pone.0008042

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General*, *114*(1), 91–103. https://doi.org/10.1037/0096-3445.114.1.91

Farah, M. J., Peronnet, F., Gonon, M. A., & Giard, M. H. (1988). Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *Journal of Experimental Psychology: General*, *117*(3), 248–57. https://doi.org/10.1037/0096-3445.117.3.248

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Förster, J., Koivisto, M., & Revonsuo, A. (2020). ERP and MEG correlates of visual consciousness: The second decade. *Consciousness and Cognition*, *80*, 102917. https://doi.org/10.1016/j.concog.2020.102917

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, *142*(1), 139–150. https://doi.org/10.1007/s00221-001-0906-7

Friston, K. (2009). The free-energy principle: A rough guide to the brain? [Edition: 2009/06/30]. *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews: Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-Level EEG-Processing Pipeline for Flexible Single Trial-Based Analyses Including Linear Mixed Models. *Frontiers in Neuroscience*, *12*(48), 1–15. https://doi.org/10.3389/fnins.2018.00048

Fulford, J., Milton, F., Salas, D., Smith, A., Simler, A., Winlove, C., & Zeman, A. (2018). The neural correlates of visual imagery vividness – An fMRI study and literature review. *Cortex*, *105*, 26–40. https://doi.org/10.1016/j.cortex.2017.09.014

Gandolfo, M., & Downing, P. E. (2019). Causal Evidence for Expression of Perceptual Expectations in Category-Selective Extrastriate Regions. *Current Biology*, *29*(15), 2496–2500. https://doi.org/10.1016/j.cub.2019.06.024

Ganis, G., & Schendan, H. E. (2008). Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. *NeuroImage*, *42*(4), 1714–1727. https://doi.org/10.1016/j.neuroimage.2008.07.004

Gazzaley, A., Clapp, W., Kelley, J., McEvoy, K., Knight, R. T., & D'Esposito, M. (2008). Age-related top-down suppression deficit in the early stages of cortical visual memory processing. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(35), 13122–13126. https://doi.org/10.1073/pnas.0806074105

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Haynes, J. D., Roth, G., Stadler, M., & Heinze, H. J. (2003). Neuromagnetic Correlates of Perceived Contrast in Primary Visual Cortex. *Journal of Neurophysiology*, *89*(5), 2655–2666. https://doi.org/10.1152/jn.00820.2002

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Kosslyn, S. M. (2005). Mental images and the Brain. *Cognitive Neuropsychology*, *22*(3-4), 333–347. https://doi.org/10.1080/02643290442000130

Kosslyn, S. M., & Thompson, W. L. (2003). When Is Early Visual Cortex Activated during Visual Mental Imagery? *Psychological Bulletin*, *129*(5), 723–746. https://doi.org/10.1037/0033-2909.129.5.723

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-33. http://CRAN.R-project.org/package=lmerTest

Lamme, V. A. F. (2018). Challenges for theories of consciousness: Seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *373*(1755), 20170344. https://doi.org/https://doi.org/10.1098/rstb.2017.0344

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *Neuroimage*, *59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A: Optics and Image Science*, *20*(7), 1434–1448. https://doi.org/https://doi.org/10.1364/JOSAA.20.001434

Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of Language on Visual Perception. *Trends in Cognitive Sciences*, *24*(11), 930–944. https://doi.org/10.1016/j.tics.2020.08.005

Lupyan, G., & Clark, A. (2015). Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Luria, R., & Vogel, E. K. (2011). Visual Search Demands Dictate Reliance on Working Memory Storage. *Journal of Neuroscience*, *31*(16), 6199–6207. https://doi.org/10.1523/jneurosci.6453-10.2011

Maier, M., & Abdel Rahman, R. (2018). Native Language Promotes Access to Visual Consciousness. *Psychological Science*, *29*(11), 1757–1772. https://doi.org/10.1177/0956797618782181

Maier, M., & Abdel Rahman, R. (2019). No matter how: Top-down effects of verbal and semantic category knowledge on early visual perception. *Cognitive, Affective & Behavioral Neuroscience*, 859–876. https://doi.org/10.3758/s13415-018-00679-8

Maier, M., Glage, P., Hohlfeld, A., & Abdel Rahman, R. (2014). Does the semantic content of verbal categories influence categorical perception? An ERP study. *Brain and Cognition*, *91*, 1–10. https://doi.org/10.1016/j.bandc.2014.07.008

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Milivojevic, B., Corballis, M. C., & Hamm, J. P. (2008). Orientation sensitivity of the N1 evoked by letters and digits. *Journal of Vision*, *8*(10), 1–14. https://doi.org/10.1167/8.10.11

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of cognitive neuroscience*, *12*(6), 1013–1023. https://doi.org/10.1162/08989290051137549

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews: Neuroscience*, *10*, 624–634. https://doi.org/10.1038/s41583-019-0202-9

Pratt, H. (2011). Sensory ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 89–114). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0050

Press, C., & Yon, D. (2019). Perceptual Prediction: Rapidly Making Sense of a Noisy World. *Current Biology*, *29*(15), R751–R753. https://doi.org/10.1016/j.cub.2019.06.054

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

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). Depth of Conceptual Knowledge Modulates Visual Processes during Word Reading. *Journal of Cognitive Neuroscience*, *24*(4), 990–1005. https://doi.org/10.1162/jocn_a_00117

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Rossion, B., & Jacques, C. (2011). The N170: Understanding the time-course of face perception in the human brain. In S. J. Luck & Kappenman, Emily S. (Eds.), *The Oxford handbook of event-related potential components* (pp. 115–142). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0064

Samaha, J., Boutonnet, B., Postle, B. R., & Lupyan, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. *Scientific Reports*, *8*, 6606. https://doi.org/10.1038/s41598-018-25093-5

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391–1400. https://doi.org/10/suppinfo/nn1549_S1.html

Seth, A. K. (2015). The Cybernetic Bayesian Brain [Section: 35(T)]. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. MIND Group. https://doi.org/10.15502/9783958570108

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience*, *29*(5), 1565–1572. https://doi.org/10.1523/JNEUROSCI.4657-08.2009

Suess, F., & Abdel Rahman, R. (2015). Mental imagery of emotions: Electrophysiological evidence. *Neuroimage*, *114*, 147–157. https://doi.org/10.1016/j.neuroimage.2015.03.063

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*(1), 43–47. https://doi.org/10.1111/1467-9280.00308

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161. https://doi.org/10.1016/j.wocn.2018.07.008

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.14.905885; this version posted June 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Voeten, C. C. (2021). Buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression. Retrieved May 5, 2021, from https://CRAN.R-project.org/package=buildmer

Weller, P. D., Rabovsky, M., & Abdel Rahman, R. (2019). Semantic Knowledge Enhances Conscious Awareness of Visual Objects. *Journal of Cognitive Neuroscience*, *31*(8), 1216–1226. https://doi.org/10.1162/jocn_a_01404

Woodman, G. F., & Arita, J. T. (2011). Direct Electrophysiological Measurement of Attentional Templates in Visual Working Memory. *Psychological Science*, *22*(2), 212–215. https://doi.org/10.1177/0956797610395395