Report

Current Biology

Strength of predicted information content in the brain biases decision behavior

Highlights

- Investigated the visual content that the brain predicts for categorization
- Showed that predicted content flows down from the ventral to the occipital cortex
- Showed that content lateralizes in the occipital cortex before stimulus is shown
- Demonstrated that per-trial strength of predicted content biases subsequent behavior

Authors

Yuening Yan, Jiayu Zhan, Oliver Garrod, Xuan Cui, Robin A.A. Ince, Philippe G. Schyns

Correspondence

philippe.schyns@glasgow.ac.uk

In brief

Yan et al. reveal the visual contents that the brain predicts before the stimulus is shown. They show that these contents flow down from ventral to occipital cortex, where they are lateralized. Critically, strength of predicted content in the brain biases subsequent perceptual categorization behavior when the stimulus is shown.

Yan et al., 2023, Current Biology *33*, 5505–5514
December 18, 2023
https://doi.org/10.1016/j.cub.2023.10.042

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

Strength of predicted information content in the brain biases decision behavior

Yuening Yan,[1] Jiayu Zhan,[2] Oliver Garrod,[1] Xuan Cui,[1] Robin A.A. Ince,[1] and Philippe G. Schyns[1,3,4,*]
[1]School of Psychology and Neuroscience, University of Glasgow, 62 Hillhead Street, Glasgow G12 8QB, UK
[2]School of Psychological and Cognitive Sciences, Peking University, 5 Yiheyuan Road, Beijing 100871, China
[3]X (formerly Twitter): @SchynsPhilippe
[4]Lead contact
*Correspondence: philippe.schyns@glasgow.ac.uk
https://doi.org/10.1016/j.cub.2023.10.042

SUMMARY

Prediction-for-perception theories suggest that the brain predicts incoming stimuli to facilitate their categorization.[1–17] However, it remains unknown what the information contents of these predictions are, which hinders mechanistic explanations. This is because typical approaches cast predictions as an underconstrained contrast between two categories[18–24]—e.g., faces versus cars, which could lead to predictions of features specific to faces or cars, or features from both categories. Here, to pinpoint the information contents of predictions and thus their mechanistic processing in the brain, we identified the features that enable two different categorical perceptions of the same stimuli. We then trained multivariate classifiers to discern, from dynamic MEG brain responses, the features tied to each perception. With an auditory cueing design, we reveal where, when, and how the brain reactivates visual category features (versus the typical category contrast) before the stimulus is shown. We demonstrate that the predictions of category features have a more direct influence (bias) on subsequent decision behavior in participants than the typical category contrast. Specifically, these predictions are more precisely localized in the brain (lateralized), are more specifically driven by the auditory cues, and their reactivation strength before a stimulus presentation exerts a greater bias on how the individual participant later categorizes this stimulus. By characterizing the specific information contents that the brain predicts and then processes, our findings provide new insights into the brain's mechanisms of prediction for perception.

RESULTS

Cued predictions bias perceptual decisions

We used a unique stimulus, Dali's *Slave Market with the Disappearing Bust of Voltaire*, which is known to elicit two distinct perceptions based on processing contents represented within different spatial frequency (SF) bands.[25–27] This image can be interpreted as either a Voltaire bust or as two nuns (Figure 1A, original). This differentiation arises from processing either low SF (LSF, at 8 cycles per image) or high SF (HSF, at 16 and 32 cycles per image),[25] as illustrated in Figure 1A, features (STAR Methods section stimuli; Figures S1A and S1B).

Before the experiment, we trained our participants (n = 10) to recognize and differentiate the specific features that enable each of these distinct perceptions. Additionally, we introduced auditory cues: a 250-ms pure tone at either 196 Hz (associated with Voltaire) or 1,760 Hz (associated with nuns). A neutral tone at 880 Hz was also introduced, which didn't predict any particular perception. During the main experiment, participants were presented with hybrid images. To create these images, we utilized Gabor sampling to adjust the visibilities of LSF Gabor features that induce Voltaire perception and, independently, HSF Gabor features that induce nuns perception. We manipulated these features by altering the proportions of visible information. The stimuli images were essentially a blend, containing varying proportions of the two features, added to a noise background. It is important to note that the proportions of Voltaire and nuns Gabor features were not fixed. They were independently and randomly sampled for each trial based on predetermined distributions, as shown in Figure 1A. The experiment was structured in two phases for each trial: a prediction stage followed by a categorization stage (Figure 1B). During prediction, one of the three auditory cues was randomly played to predict the features of the upcoming visual stimulus (Voltaire or nuns) or nothing. In the subsequent categorization stage, participants viewed the hybrid image and were tasked with reporting their strongest perception: Voltaire, nuns, or—in cases of uncertainty—they could respond with "don't know."

Figure 1C confirms that predictive cues biased perceptual decision behavior, as evidenced by the group-level psychometric curves. For each auditory cue (panel) and each potential decision response (colored curves), we analyzed how the presence of specific features (i.e., relative proportions of Voltaire versus nuns Gabor features in the stimulus, x axis) affected the probability of a particular response (y axis) (STAR Methods section cued-predictions bias perceptual decisions). The point of subjective equality (PSE) is a crucial metric here. It represents the level of stimulus evidence (x axis) where both the Voltaire and nuns responses (y axis)

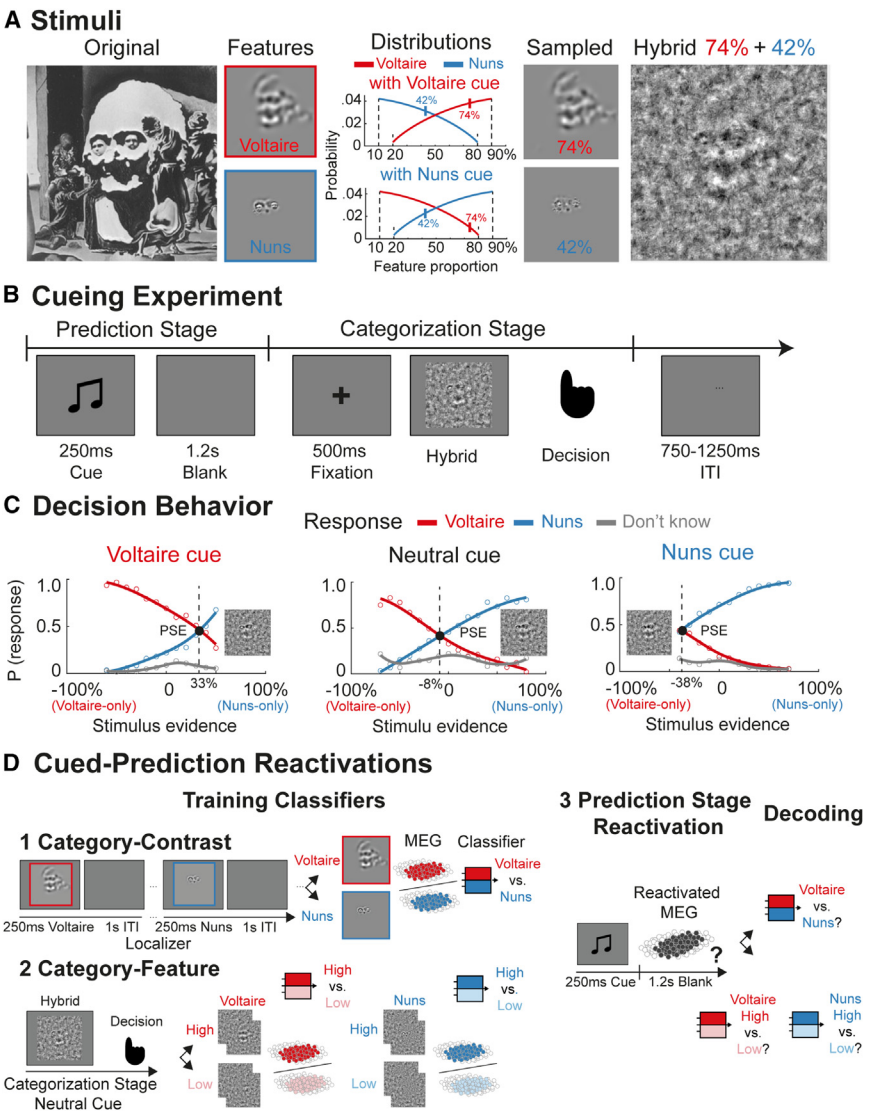

Figure 1. Experimental design

(A) Stimuli. From the original ambiguous stimulus, we applied filters to extract the Voltaire and nuns stimulus features,[25] respectively represented within LSF 8 cycles/image and HSF 16–32 cycles/image (see also STAR Methods section stimuli and Figure S1). On each trial, a hybrid stimulus comprised randomly and independently selected proportions of the Voltaire (red) + nuns (blue) features filtered with spatial Gabors. These proportions were independently and randomly sampled from the distributions, based on whether the cue was Voltaire versus nuns (under neutral cueing, proportions were random between 0% and 100%). The hybrid example illustrates proportions of 74% Voltaire + 42% nuns Gabor features (see vertical marks in the distribution) inserted in Gabor background noise (STAR Methods section stimuli).

(B) Cueing experiment. Prediction stage: a 250-ms pure tone cued either the Voltaire (196 Hz) or nuns (1,760 Hz) stimulus distribution, or was neutral (880 Hz), followed by a 1.2-s blank interval. Categorization stage: a 500-ms fixation was followed by a hybrid image that remained on the screen until response (nuns versus Voltaire versus don't know, 3-AFC), followed by a 750- to 1,250-ms inter-trial interval (ITI) with jitter.

(C) Decision behavior. Relationships between stimulus feature evidence and response probability, i.e., color-coded curves of p(Voltaire), p(nuns), and p(don't know), for each auditory cue (panel). The point of subjective equality (PSE, black dot) is the level of stimulus evidence of equally likely p(Voltaire) and p(nuns), as illustrated in the stimulus image. Left and right panels show that Voltaire and nuns cues shift the PSE in opposite directions compared with the neutral cue (central). See also Table S1.

(D) Cued-prediction reactivation. (1) Category-contrast classifier. Trained on a localizer run prior to the cueing experiment, category-contrast classifiers learn to discriminate the bottom-up patterns of sensor-level neural responses to each stimulus category (color-coded as Voltaire versus nuns and measured with MEG). (2) Category-feature classifier. Trained on categorization-stage sensor-level data under the neutral cue, category-feature classifiers learn to discriminate the high (>70%) versus low (<30%) proportions of Voltaire Gabor features and, separately, nuns Gabor features. (3) Prediction stage reactivation. Following the auditory cues that predict the Voltaire versus nuns Gabor feature distributions (A), the category-contrast and the category-feature classifiers quantify the single-trial reactivation strength of the sensor-level pattern of MEG activity every 4 ms of the prediction stage.

intersect and are equally probable. Under the neutral cue (Figure 1C, central), the PSE is −8%, meaning that 8% more Voltaire evidence is required for equiprobable responses, a small bias toward nuns responses. However, when we cue Voltaire (left), the PSE shifts because participants require 33% less Voltaire evidence. Similarly, when we cue nuns (middle) the PSE shifts to 38% less nuns evidence. These cue-induced shifts in PSE were replicated in each participant (Table S1), reinforcing the idea that predictive cueing strongly biases responses toward the cued behavior.

Brain predictions: Category contrast versus category features

We sought to determine how auditory cues reactivate from the participant's memory specific visual representations, namely the Gabor features associated with Voltaire and nuns behavioral responses. This was based on an extensive dataset of 3,375 trials per participant, with all decoding and statistical inferences performed with each participant.[28,29] The core of our analyses is a comparison between two types of cued reactivations: (1) our new category features reactivations, where specific features of Voltaire and nuns are evoked, and (2) the more typical category contrast reactivations, where there is an underconstrained contrast being made between Voltaire and nuns, which could be feature specific to Voltaire or nuns, or features from both categories.

Our methods are illustrated in Figure 1D. Initially, we developed category-contrast classifiers (Figure 1D1). These were trained to discriminate between images containing 100% Voltaire features and those containing 100% nuns features. To

achieve this, we relied on the bottom-up MEG responses from a localizer run before the main experiment. In this localizer, the participants categorized an outlier in a sequence (e.g., in a series of 100% Voltaire images, spot the lone 100% nuns image). Using these category-contrast classifiers, we gauged the prediction strength during the prediction stage, essentially evaluating the per-trial cued reactivation of the Voltaire versus nuns category contrast based on the auditory cues, where prediction strength is the per-trial decision value of the classifier. However, these classifiers cannot separately quantify the reactivation of the feature content unique to Voltaire and to nuns. To overcome this, we developed category-feature classifiers. These learned the relationship between Voltaire-specific features and brain responses, and independently between nuns-specific features and brain responses. Training for these classifiers occurred using the categorization stage data, under neutral cueing (Figure 1D2). In the prediction stage, the per-trial decision values of these classifiers quantified how strongly the cues reactivated the predicted category features (Figure 1D3). Our final task was comparing how the per-trial prediction strengths of these two types of classifiers influence participant behavior. To preview our results, predictions rooted in category features are superior. They offer more precise localization, are more specifically driven by auditory cues, and their per-trial reactivations more strongly bias participant behavior. We now detail these results below.

Reactivation of predictions with category-contrast classifiers

Previous research indicates that top-down predictions of category contrast reverse their bottom-up flow.[18,19] We therefore applied category-contrast classifiers during the prediction stage of our study to set a comparison standard for our novel category-feature classifiers.

For every participant, binary category-contrast classifiers were trained and cross-validated on images of 100% Voltaire and 100% nuns, every 4 ms of the 0–400 ms post-stimulus response of the MEG localizer (Figure 1D1). We similarly trained classifiers in a separate MEG localizer of the auditory cues (196 Hz, Voltaire versus 1,760 Hz, nuns). Both these localizers were trained before the primary cueing experiment to avoid contamination (STAR Methods sections visual localizer and auditory localizer). Classifiers were further split into the "early" set, trained 75–150 ms post stimulus (capturing low-level visual features processing,[30,31] early P1 event-related potential [ERP]), and the "late" set, trained 150–280 ms post stimulus (capturing advanced categorization stages, N170 and N250 ERPs[26,32–37]) (STAR Methods section localizer cross-validation).

During the top-down prediction stage, we applied these early and late bottom-up category-contrast classifiers (illustrated in Figures 1B and 1D3). This generated per-trial classifier decision values every 4 ms, representing the prediction strength of category contrast. We calculated the progression of this prediction strength over time (i.e., reactivation performance), computing for each classifier mutual information (MI) between ground truth Voltaire versus nuns cue and decision value$_t$, (FWER corrected over 100 localizer training time points and 150 prediction stage testing time points, $p < 0.05$, one-tailed; STAR Methods section category-contrast reactivation). We then identified two classifiers: one with maximum prediction reactivation performance from the early set (75–150 ms) and one with maximum prediction reactivation from the late set (150–280 ms)—i.e., both selected from the matrix of localizer training time points × prediction testing time points (see Figure S1C for complete data). These classifiers gauged the strength of the auditory-cue-induced reactivation of the prediction of Voltaire versus nuns category contrast.

As depicted in Figure 2, dark purple curves reveal across participants that auditory cues reactivate late classifiers more strongly compared with the early ones (shown in light purple). Source localizations of these predictions were determined by computing MI(late classifier decision value; MEG source activity) on all 8,196 individual sources during the prediction stage. Initial ventral stream involvement from the temporal cortex (middle temporal gyrus [MTG], superior temporal gyrus [STG], parahippocampal cortex [PHC], and fusiform gyrus [FG]) moving down to the occipital cortex (lingual gyrus [LG], pericalcarine cortex [PCAL], cuneus [CUN], and lateral occipital cortex [LOC]), followed by the bilateral occipital cortex.[18,19] 9/10 participants displayed such consistent visual category-contrast predictions (Figure S1; FWER, $p < 0.05$, one-tailed, Bayesian population prevalence [BPP][28,38]) with maximum *a posteriori* probability (MAP) estimate of the population prevalence of the effect of 9/10 replications = 0.9 (95% highest posterior density interval [HPDI] [0.61 0.99]).

We controlled the distinct propagation of the auditory cue contrast by decoding the prediction stage with classifiers trained on the auditory localizer data (orange). Initially discriminated in the temporal cortex (75–150 ms post-cue), the auditory cue contrast does not propagate beyond MTG and STG. It is important to note that while the auditory cue contrast and the reactivated visual category contrast occur at similar times, they spread differently within the brain (see regions' histograms and STAR Methods section source representation of category contrast). We replicated this result in all participants (BPP = 1.0 [0.75 1.0]; MAP [95% HPDI]; see Figure S1 for individual results). Moreover, when we accounted for and removed the effect of the auditory cue contrast, the spread pattern of the visual reactivation remained similar—i.e., computed as conditional MI (CMI), that is CMI(ground truth Voltaire versus nuns cue; visual decision value|auditory decision value; Figure S2A). This analysis clearly shows that the reactivation patterns observed for visual category contrast are not conflated with the contrasts present in the auditory cues.

Reactivation of predictions with category-feature classifiers

We now turn to study the Gabor features that the participant's brain reactivates separately for Voltaire and for nuns. We divided the stimuli into those with high (>70%) versus low (<30%) Gabor features (Figure 1A). Then, we trained our new bottom-up category-feature classifiers (Figure 1D2) to discriminate high versus low Gabor features. We did this from multivariate MEG sensor responses every 4 ms of the categorization stage, under the neutral cue (STAR Methods section category-feature classifier cross-validation). With these we could study how the brain reactivates the Gabor features of Voltaire and those of nuns at prediction, separately under Voltaire and under nuns cueing—i.e., separately computing MI(Voltaire versus neutral cue; Voltaire-feature-classifier decision value) and MI(nuns versus neutral

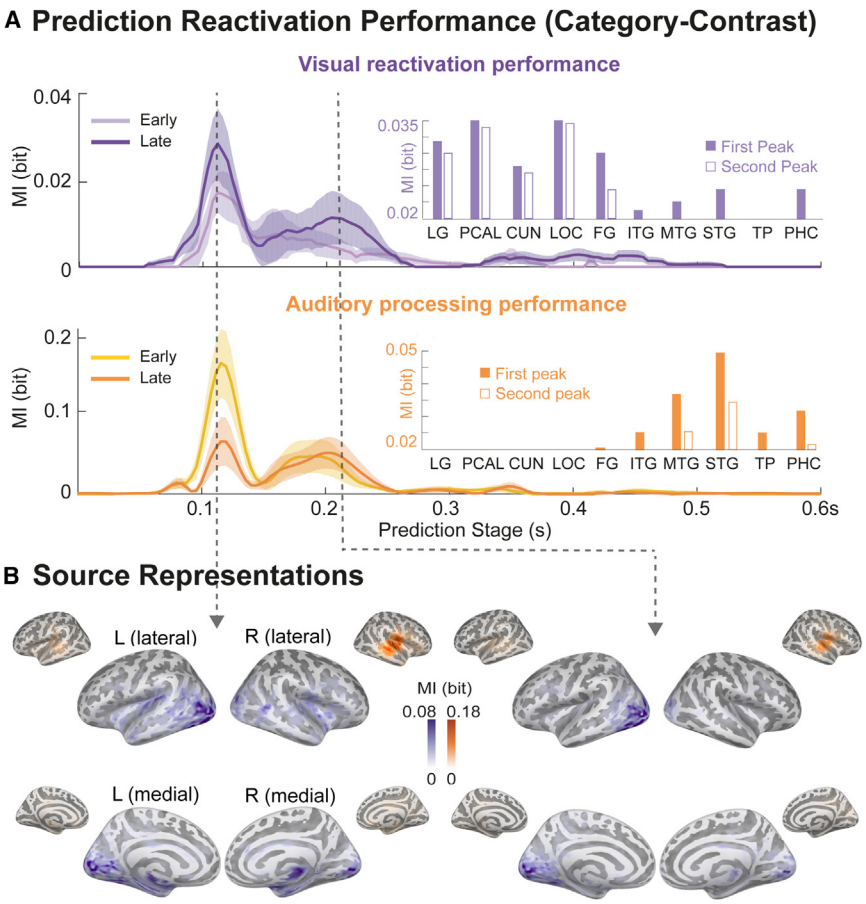

Figure 2. Voltaire versus nuns category-contrast reactivation

(A) Prediction reactivation performance. Purple curves show the dynamic prediction reactivation performance of the Voltaire versus nuns category contrast at each time point of the prediction stage, for the early classifier (with the highest performance over 75–150 ms post-cue, light purple) and late classifier (with highest performance over 150–280 ms post-cue, dark purple) averaged across participants—shaded regions denote ±SEM. To control for the propagation of the auditory cue contrast, we trained auditory classifiers to discriminate the two cues on the auditory localizer data and similarly tested their decoding performance on the same prediction stage responses (orange). The curves show that the late bottom-up classifier (dark purple) better models the two peaks of Voltaire versus nuns visual prediction reactivation (dashed arrows) whereas the early classifier (light orange) better models the two peaks of auditory decoding.

(B) Source representations. Cortical surface maps reveal the MEG sources that contribute to the two reactivation peaks of the late category-contrast classifier (dark purple) and early auditory classifier (light orange), computed as MI(classifier decision value; MEG source activity).

Bar plots in (A) show their mean representation strength (MI) across all sources in each ROI at the two peak time points (early, filled bars; late, unfilled bars). LG, lingual gyrus; PCAL, pericalcarine cortex; CUN, cuneus; LOC, lateral occipital cortex; FG, fusiform gyrus; ITG, inferior temporal gyrus; MTG, middle temporal gyrus; STG, superior temporal gyrus; TP, temporal lobe; PHC, parahippocampal cortex.

See also Figures S1 and S2.

cue; nuns-feature-classifier decision value). Furthermore, we selected the best-performing classifiers at the peak time point of prediction decoding (FWER corrected over 100 categorization-stage training time points and 150 prediction stage testing time points, $p < 0.05$, one-tailed; STAR Methods section category-feature reactivation). We replicated this significant decoding (FWER corrected over 100 categorization-stage training time points and 150 prediction-stage testing time points, $p < 0.05$, one-tailed) in 10/10 participants for Voltaire (BPP = 1.0 [0.75 1.0]; MAP [95% HPDI]) and 9/10 participants for nuns (BPP = 0.9 [0.61 0.99]; MAP [95% HPDI]; Figure S1).

We then localized the brain sources giving rise to these category-feature predictions by computing MI(Voltaire-feature-classifier decision value; MEG source activity) and MI(nuns-feature-classifier decision value; MEG source activity) on 8,196 sources at peak reactivation time (see complete data in Figure S1D; STAR Methods section category-feature reactivation). What we found (as shown in Figure 3A) is that category-feature predictions comprise bilateral PHC, but also that they lateralize to left LG, LOC, and FG for coarser-scale Voltaire, but bilateral to the early visual cortex and right LOC for finer-scale nuns. Thus, category-feature classifiers lateralized the reactivations of the predicted features, a result that aligns with lateralized bottom-up representations of scale information shown in other work[39,40] and in our localizer task (Figure S2).

Additionally, Figure S3 highlights an interesting pattern: when prediction is lateralized to one hemifield to the left or right HSF nun's face, individual participants subsequently represent the same-side HSF nun's face for categorization behavior. This suggests a relationship between prediction and categorization behavior that the next sections develop.

Specificity of cued reactivations

To investigate how selectively auditory cues reactivate visual predictions, we contrasted reactivations when cues are Voltaire versus neutral and when they are nuns versus neutral. For both contrasts, we compared the performance of the category-contrast and category-feature classifiers (Figure 3B; STAR Methods section specificity of cued reactivations). For each feature, we found that selective reactivation for the cue contrast is highest with the category-feature classifier, indicating its higher cue specificity. That is, the Voltaire cue contrast reactivates the Voltaire-feature classifier more strongly than the nuns-feature classifier and category-contrast classifier. We replicated this cue specificity in 8/10 participants for Voltaire-cued reactivations (BPP = 0.8 [0.49 0.96]; MAP [95% HPDI]) and 7/10 participants for nuns-cued reactivations (BPP = 0.7 [0.53 0.82]; MAP [95% HPDI]). Thus, with greater specificity of cued reactivations, the category-feature classifiers (versus category-contrast classifier) provide a finer conceptual resolution of

CellPress
OPEN ACCESS

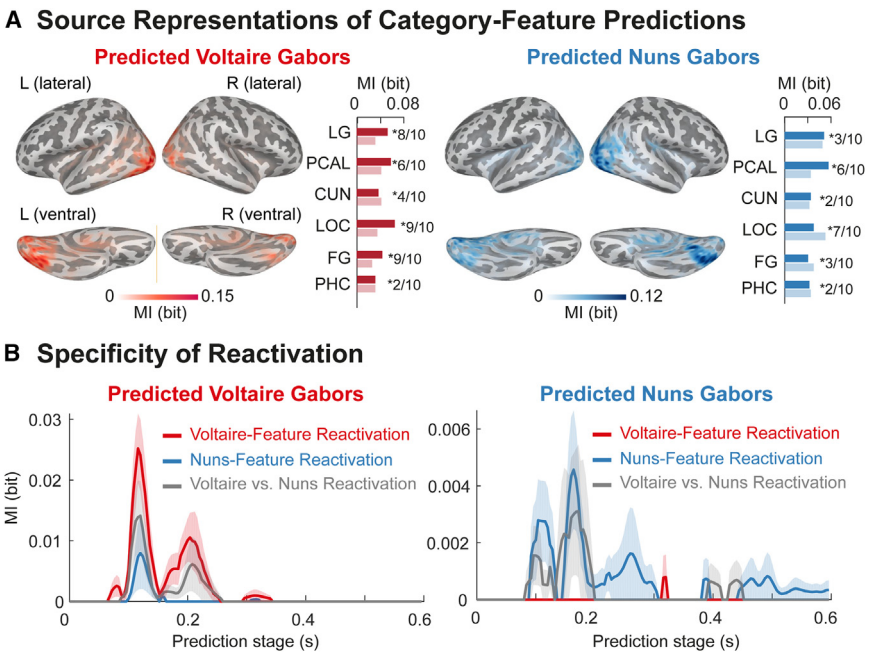

Figure 3. Predictions of Voltaire and nuns Gabor features

(A) Source representations. Cortical surface plots localize the MEG sources that contribute to category-feature predictions, at the maximal time point of cued reactivation of the pattern that represents the discrimination of high versus low Gabor features for Voltaire (red) and for nuns (blue). See also Figure S2. Bar plots indicate mean (across participants) reactivated category-feature prediction strength (MI) at peak time, in left (darker) and right (lighter) hemispheres, showing that (1) bilateral PHC represent Voltaire and nuns Gabor predictions, (2) Voltaire Gabor predictions are left lateralized (in LG, LOC, and FG), and nuns Gabor predictions are right lateralized (in LOC). For each ROI, we compared the source representations between the left and right hemispheres, separately for Voltaire and nuns feature predictions. Asterisk (*) provides the prevalence of participants with significant lateralization ($p < 0.05$, independent t tests with Bonferroni correction). See also Figures S1 and S2.

(B) Specificity of reactivation. Color-coded curves show the relative performance of the best category feature (red for high versus low Voltaire Gabors; blue for nuns) and best category-contrast (Voltaire versus nuns) classifiers during the prediction stage. See also Figure S3.

the visual contents predicted when studying the mechanisms of prediction for categorization.

Reactivation of category features at prediction bias behavior at categorization

Finally, we wanted to compare how cued reactivations of category contrast versus category features during prediction change decision behavior during categorization. Specifically, we compared how the top 30% (strong) versus bottom 30% (weak) prediction strengths (i.e., the classifier decision values) trials change the psychometric relationship between stimulus evidence and decision probabilities (STAR Methods section reactivation biases behavior). Figures 4A and 4B present these results.

Figure 4A shows that the strength (strong versus weak) of category-feature reactivations at prediction biases decision behavior—i.e., shift the PSE by 11% (25%–36%) on Voltaire-cued trials and −16% (−24% to −40%) on nuns-cued trials. We replicated these PSE shifts (FWER corrected over 100 categorization-stage training time points and 150 prediction-stage testing time points, $p < 0.05$, two-tailed) in 8/10 participants (BPP = 0.8 [0.49 0.96]; MAP [95% HPDI]) (see Table S1 for the individual results). To compare, Figure 4B shows that the strength of category-contrast reactivations (nuns versus Voltaire, trained on 100% localizer stimuli), shifts the PSE by only −2% (nuns predictions, only 2/10 participants with significant effect) and 9% (Voltaire, only 3/10 participants with significant effect) (see Table S1 for the individual results).

Notably, the trial-by-trial prediction strength of category features biases the participant's decisions under both auditory cues. Further, the magnitude of this bias originating from the brain (before the stimulus is shown) is comparable to the full effect of the cue itself on behavior following stimulus presentation (as shown in Figure 1C). Specifically, the PSE changes due to brain reactivations of category features are 27% (Voltaire) and 52% (nuns) of the changes in PSE that are due to the auditory cues themselves on behavior (cf. Figure 1C). That is, strong (versus weak) category-feature reactivations under Voltaire cueing change behavioral PSE by 11%, whereas Voltaire versus neutral cueing changes behavioral PSE by 41%; for the nuns cue, numbers are −16% (strong versus weak nuns-cued reactivations) versus −30% (nuns versus neutral cue).

Finally, we found that participants tend to rely more strongly on predictions when the evidence is ambiguous.[9] Voltaire behavioral responses when brain predictions are strong versus weak are significantly higher when stimuli are more ambiguous ([0%,40%], i.e., closer to PSE versus [−0%,−40%], $t(9) = 3.648$, $p = 0.005$, paired-sample t tests, Bonferroni correction).

In sum, the effect of category-feature reactivations on behavior is stronger and across a wider range of stimulus evidence compared with the category-contrast classifier. Thus, focusing on category-specific feature predictions can disentangle the decoded category contrast to improve our understanding of the mechanisms of prediction, from cue reactivations of category-specific features to the effect that these predictions exert on decision behavior.

DISCUSSION

To better understand prediction for perception, we must access the dynamic perceptual contents that the brain predicts about a category—i.e., the features that represent this category in memory. We argued that typical decoders trained to discriminate the bottom-up contrast between two categories (or

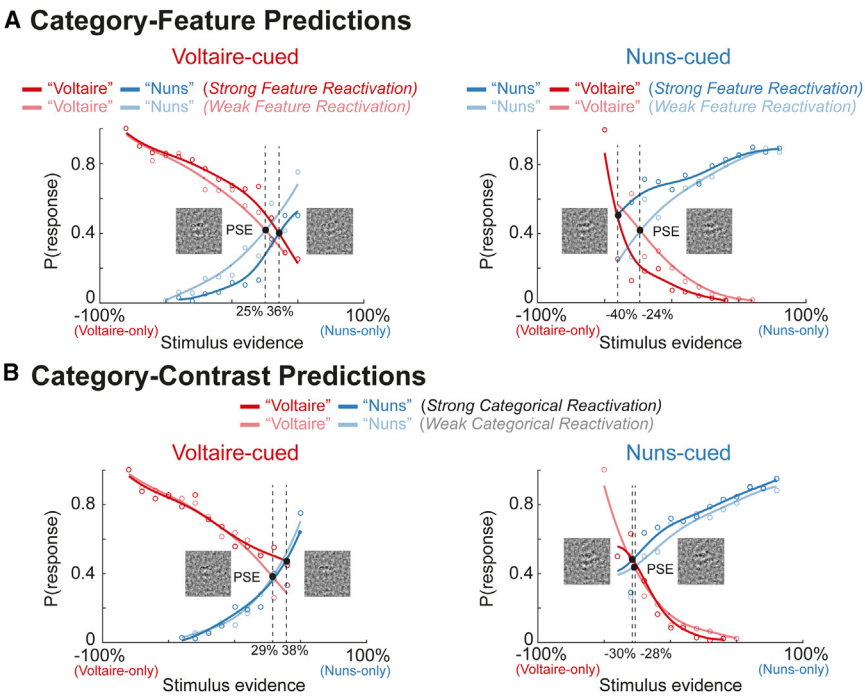

Figure 4. Strength of predictions biases decision behavior

(A) Category-feature predictions. For Voltaire- and nuns-cued trials (panels), we independently computed the relationships between stimulus feature evidence (i.e., proportions of Voltaire versus nuns Gabor features, x axis) and response probabilities (i.e., p(Voltaire) [red curves], p(nuns) [blue curves], y axis) for different strengths of predicted category (i.e., Voltaire-cued [left], nuns-cued [right], and strong [opaque curves] versus weak [transparent curves] prediction reactivation strengths of the category-feature decoders). The black dot is the PSE of equal p(Voltaire) and p(nuns), as illustrated with the stimulus image. The curves reveal response bias (vertical offset of light versus dark curves) across a wide range of Gabor feature evidence. The differences in the PSE between the strong and weak reactivation conditions quantifies this bias.

(B) Category-contrast predictions. Similar analyses with the category-contrast decoder of Voltaire versus nuns only weakly biases perceptual decisions, and only at high levels of Gabor feature evidence.

See also Figure S3 and Table S1.

stimuli) do not access the top-down reactivations of these category-specific features. This is because the way in which we set up a category contrast, such as the difference between a target category (e.g., car) and a chosen contrast category (like buses or faces), is inevitably relative to the contrast category (and often arbitrary) and not necessarily relative to the target category itself.

To address this, we considered classifiers that decode the specific visual features underlying each category—i.e., Voltaire and nuns Gabor features—and then used these to quantify the reactivation of the representations of these specific visual features at prediction. We showed that top-down reactivations of the category features (versus category contrast) are more precisely localized (i.e., lateralized) and more specifically driven by the cues, with per-trial predicted reactivation strength that more strongly biases participant's perceptual categorization behavior across a wider range of stimulus evidence.

Lateralized source mapping of category-feature predictions

Previous research has shown that a category contrast can be predicted across the visual hierarchy, from the prefrontal cortex to sensory areas.[20,41] However, we demonstrated why it is critical to understand the specific category features that are predicted. Although category-contrast classifiers showed bilateral reactivations of the predictions, we showed that classifiers of the specific category features lateralize the predictions of Voltaire features in the left occipital cortex and of nuns features in right LOC. Importantly, this top-down lateralization of predicted features is consistent with their bottom-up representations in the localizers and with the SF-related lateralization in other studies that show that LSF versus HSF processing is lateralized in left versus right hemispheres, respectively.[39,40]

Generalizing from these insights, it is critical to characterize the predicted features because prediction mechanisms will likely align them with the brain's representations of the input,[6,42] which could change according to the considered region of the occipito-ventral hierarchy. Future studies that control stimulus features could therefore parametrically change the scale or the orientation features of the same 3D objects to investigate whether cueing specifically these features (in addition to the object category) reactivates category features that are scale- and orientation-invariant in higher-level rFG[43–46] but scale- and orientation-dependent in the early visual cortex.[47–50]

Relationship between neural representations of prediction and perceptual behavior

Advances in neuroimaging have further demonstrated that these predictions can modulate the neural responses to inputs by modulating premotor cortex activity[51,52] and by suppressing responses of sensory cortices to predicted stimuli.[10,53–56] Such suppression is tuned to the expected stimulus, indirectly showing feature specificity[57,58] and facilitation of categorization RT.[59] However, these studies did not quantify how strongly the cue reactivated the predicted contents in the brain nor did they demonstrate a direct relationship between these reactivated contents and decision behavior as we did here. One study has modeled the single-trial relationship between MVPA performance and categorization RT.[60] Here, we established the direct relationship between the single-trial prediction strength of specific visual contents and subsequent perceptual decision behavior. Our category-feature classifiers directly quantified the effect of reactivating perception-related category features on decision behavior bias, whereby stronger feature reactivations led to stronger decision bias across a wider range of feature evidence, all compared with typical category-contrast

reactivations. This relationship between reactivation strength and bias was computed for each cue separately, showing that the neural reactivations provide additional information about response bias—i.e., over and above the strong effect of the behavioral cue.

Our results show that predictions following cue onset impact subsequent behavior. However, the reactivation of predicted features is not sustained. Future work should specifically address how a brief reactivation influences stimulus representation[61] and subsequent behavior. For example, we could aim to identify the network mechanisms involving frontal, parietal, and occipital cortex regions[62] that could maintain, trial-by-trial, the readiness for stimulus features following their cued-predictions, all before stimulus onset.[20]

Perceptual bias, biased response to the cue, or spatial attention?

One could oppose the idea that the reported effects of prediction on behavior (a bias on stimulus perception) might, in fact, reflect a simpler response bias to the auditory cue (e.g., responding Voltaire when hearing the Voltaire cue). This concern might arise because we did not explicitly model the decision-making process while separating the stimulus evidence. However, our data provide insights into this issue. Specifically, the behavioral results show that participants 3 and 9 (cf. Figure S1) did not engage with the perceptual task as instructed. Rather than reporting how they perceived the stimuli, the behavior of these two participants matched the auditory cue (i.e., a response bias) and we could not decode cued reactivations of the predicted category features. In contrast, the strength of these cued reactivations directly and conclusively biased decision behavior in the remaining 8/10 participants, who therefore demonstrate a perceptual bias rather than a response bias to the auditory cue.

One could also argue that our effects of prediction are partially conflated with those of attention. This is a thorny issue because features will have spatial locations in 2D images, and here will be represented across SF bands with Gabor filters. Hence, the predicted pixels associated with each perception will necessarily have attentional correlates. However, we explicitly trained participants to learn the relationship between 2D image pixels across SF bands and different perceptual responses. Therefore, successful learning prior to the cueing experiment necessarily implies specifically discriminating the features associated with each perception (STAR Methods section cue-feature training). Still, teasing apart the specific effects of visual representations across high-dimensional SF Gabors versus spatial attention (to the global features or their detailed Gabor features, and how this differs from "representation") in the context of visual predictions should be the object of detailed investigations,[63–65] including in the laminar layers of the early visual cortex.[66–68]

To conclude, we traced prediction for perception using a realistic and well-known ambiguous stimulus and its two perceptual categorizations. We showed that predictions of category-relevant features disentangle the typical category contrast to improve a mechanistic understanding of prediction processes in the brain, from the top-down cue reactivation of category-feature representations to the trial-by-trial effect of these reactivations on behavior. This approach generalizes to a wider range of cognitive neuroscience, such as more naturalistic face (e.g., facial identity or expression),[69–71] object (e.g., cars or buildings),[72] or scene (e.g., indoor rooms, outdoor environments)[51,73] predictions and categorizations. This point brings up the complex question of what the memorized stimulus features are and how we can embed them into a generative model of the stimulus, so we can psychophysically model their dynamic predictions in the brain,[70,74–79] as we started developing here. However, to minimally apply our approach to a cueing design, all we need is to characterize the features that subtend a recognition task and then demonstrate that their cued reactivation facilitates task-behavior as we did here. Such feature characterization is applicable to any stimulus category, feature sampling scheme, and visual categorization task.[75,76]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
- METHOD DETAILS
  - Stimuli
  - Procedure
  - MEG Data Acquisition and Pre-processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Cued-Predictions bias perceptual decisions
  - Category-contrast decoding of predictions
  - Category-feature decoding of predictions
- REACTIVATION OF LEFT VS. RIGHT NUNS FACE FEATURES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.10.042.

ACKNOWLEDGMENTS

We thank Christoph Daube and Lukas Snoek for their comments on earlier versions of this manuscript. This work was funded by the Wellcome Trust (Senior Investigator Award, UK; 107802) and the Multidisciplinary University Research Initiative/Engineering and Physical Sciences Research Council (USA, UK; 172046-01), awarded to P.G.S., and the Wellcome Trust (214120/Z/18/Z), awarded to R.A.A.I. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.Y., J.Z., and P.G.S.; methodology, O.G., R.A.A.I., and P.G.S.; investigation, Y.Y. and X.C.; analysis, Y.Y. and R.A.A.I.; writing, Y.Y., J.Z., R.A.A.I., and P.G.S.; funding acquisition, R.A.A.I. and P.G.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 22, 2023
Revised: October 11, 2023
Accepted: October 23, 2023
Published: December 7, 2023

REFERENCES

1. Smith, F.W., and Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. Proc. Natl. Acad. Sci. USA *107*, 20099–20103.
2. Uran, C., Peter, A., Lazar, A., Barnes, W., Klon-Lipok, J., Shapcott, K.A., Roese, R., Fries, P., Singer, W., and Vinck, M. (2022). Predictive coding of natural images by V1 firing rates and rhythmic synchronization. Neuron *110*, 1240–1257.e8.
3. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. *36*, 181–204.
4. Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. *11*, 127–138.
5. Gilbert, C.D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. Neuron *54*, 677–696.
6. Yuille, A., and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? Trends Cogn. Sci. *10*, 301–308.
7. Glenberg, A.M. (1997). What memory is for. Behav. Brain Sci. *20*, 1–19.
8. Ye, Z., Shi, L., Li, A., Chen, C., and Xue, G. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. eLife *9*, e57023.
9. De Lange, F.P., Heilbron, M., and Kok, P. (2018). How do expectations shape perception? Trends Cogn. Sci. *22*, 764–779.
10. Kok, P., Jehee, J.F.M., and de Lange, F.P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. Neuron *75*, 265–270.
11. Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Dale, A.M., Hämäläinen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., et al. (2006). Top-down facilitation of visual recognition. Proc. Natl. Acad. Sci. USA *103*, 449–454.
12. Stein, T., and Peelen, M.V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. J. Exp. Psychol. Gen. *144*, 1089–1104.
13. Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.M., Kennedy, H., and Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. Neuron *89*, 384–397.
14. Benedek, M., Bergner, S., Könen, T., Fink, A., and Neubauer, A.C. (2011). EEG alpha synchronization is related to top-down processing in convergent and divergent thinking. Neuropsychologia *49*, 3505–3511.
15. Lobier, M., Palva, J.M., and Palva, S. (2018). High-alpha band synchronization across frontal, parietal and visual cortex mediates behavioral and neuronal effects of visuospatial attention. NeuroImage *165*, 222–237.
16. Brandman, T., Avancini, C., Leticevscaia, O., and Peelen, M.V. (2020). Auditory and semantic cues facilitate decoding of visual object category in MEG. Cereb. Cortex *30*, 597–606.
17. Treder, M.S., Charest, I., Michelmann, S., Martín-Buro, M.C., Roux, F., Carceller-Benito, F., Ugalde-Canitrot, A., Rollings, D.T., Sawlani, V., Chelvarajah, R., et al. (2021). The hippocampus as the switchboard between perception and memory. Proc. Natl. Acad. Sci. USA *118*, e2114171118.
18. Linde-Domingo, J., Treder, M.S., Kerrén, C., and Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. Nat. Commun. *10*, 179.
19. Dijkstra, N., Ambrogioni, L., Vidaurre, D., and van Gerven, M. (2020). Neural dynamics of perceptual inference and its reversal during imagery. eLife *9*, e53588.
20. Kok, P., Mostert, P., and De Lange, F.P. (2017). Prior expectations induce prestimulus sensory templates. Proc. Natl. Acad. Sci. USA *114*, 10473–10478.
21. Lee, S.H., Kravitz, D.J., and Baker, C.I. (2012). Disentangling visual imagery and perception of real-world objects. NeuroImage *59*, 4064–4073.
22. Hindy, N.C., Ng, F.Y., and Turk-Browne, N.B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat. Neurosci. *19*, 665–667.
23. Dijkstra, N., Bosch, S.E., and van Gerven, M.A.J. (2019). Shared neural mechanisms of visual perception and imagery. Trends Cogn. Sci. *23*, 423–434.
24. Kerrén, C., Linde-Domingo, J., Hanslmayr, S., and Wimber, M. (2018). An optimal oscillatory phase for pattern reactivation during memory retrieval. Curr. Biol. *28*, 3383–3392.e6.
25. Bonnar, L., Gosselin, F., and Schyns, P.G. (2002). Understanding Dali's Slave Market with the Disappearing Bust of Voltaire: a case study in the scale information driving perception. Perception *31*, 683–691.
26. Zhan, J., Ince, R.A.A., Van Rijsbergen, N., and Schyns, P.G. (2019). Dynamic construction of reduced representations in the brain for perceptual decision behavior. Curr. Biol. *29*, 319–326.e4.
27. Smith, M.L., Gosselin, F., and Schyns, P.G. (2006). Perceptual moments of conscious visual experience inferred from oscillatory brain activity. Proc. Natl. Acad. Sci. USA *103*, 5626–5631.
28. Ince, R.A.A., Kay, J.W., and Schyns, P.G. (2022). Within-participant statistics for cognitive science. Trends Cogn. Sci. *26*, 626–630.
29. Ince, R.A., Paton, A.T., Kay, J.W., and Schyns, P.G. (2021). Bayesian inference of population prevalence. eLife *10*, e62461.
30. Shigeto, H., Tobimatsu, S., Yamamoto, T., Kobayashi, T., and Kato, M. (1998). Visual evoked cortical magnetic responses to checkerboard pattern reversal stimulation: a study on the neural generators of N75, P100 and N145. J. Neurol. Sci. *156*, 186–194.
31. Clark, V.P., Fan, S., and Hillyard, S.A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. Hum. Brain Mapp. *2*, 170–187.
32. Cichy, R.M., Pantazis, D., and Oliva, A. (2014). Resolving human object recognition in space and time. Nat. Neurosci. *17*, 455–462.
33. Hillyard, S.A., and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. Proc. Natl. Acad. Sci. USA *95*, 781–787.
34. Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological studies of face perception in humans. J. Cogn. Neurosci. *8*, 551–565.
35. Jaworska, K., Yan, Y., Van Rijsbergen, N.J., Ince, R.A.A., and Schyns, P.G. (2022). Different computations over the same inputs produce selective behavior in algorithmic brain networks. eLife *11*, e73651.
36. Ince, R.A.A., Jaworska, K., Gross, J., Panzeri, S., Van Rijsbergen, N.J., Rousselet, G.A., and Schyns, P.G. (2016). The deceptively simple N170 reflects network information processing mechanisms involving visual feature coding and transfer across hemispheres. Cereb. Cortex *26*, 4123–4135.
37. Schyns, P.G., Petro, L.S., and Smith, M.L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. Curr. Biol. *17*, 1580–1585.
38. Ince, R.A., Paton, A.T., Kay, J.W., and Schyns, P.G. (2021). Bayesian inference of population prevalence. eLife *10*, e62461.
39. Eger, E., Schyns, P.G., and Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. NeuroImage *22*, 232–242.
40. Rotshtein, P., Vuilleumier, P., Winston, J., Driver, J., and Dolan, R. (2007). Distinct and convergent visual processing of high and low spatial frequency information in faces. Cereb. Cortex *17*, 2713–2724.

41. Demarchi, G., Sanchez, G., and Weisz, N. (2019). Automatic and feature-specific prediction-related neural activity in the human auditory system. Nat. Commun. *10*, 3440.
42. Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as Bayesian inference. Annu. Rev. Psychol. *55*, 271–304.
43. Chang, L., and Tsao, D.Y. (2017). The code for facial identity in the primate brain. Cell *169*, 1013–1028.e14.
44. Huth, A.G., Nishimoto, S., Vu, A.T., and Gallant, J.L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron *76*, 1210–1224.
45. Kornblith, S., Cheng, X., Ohayon, S., and Tsao, D.Y. (2013). A network for scene processing in the macaque temporal lobe. Neuron *79*, 766–781.
46. Pietrini, P., Furey, M.L., Ricciardi, E., Gobbini, M.I., Wu, W.H., Cohen, L., Guazzelli, M., and Haxby, J.V. (2004). Beyond sensory images: object-based representation in the human ventral pathway. Proc. Natl. Acad. Sci. USA *101*, 5658–5663.
47. Niemeier, M., Goltz, H.C., Kuchinad, A., Tweed, D.B., and Vilis, T. (2005). A contralateral preference in the lateral occipital area: sensory and attentional mechanisms. Cereb. Cortex *15*, 325–331.
48. Fang, F., Boyaci, H., Kersten, D., and Murray, S.O. (2008). Attention-dependent representation of a size illusion in human V1. Curr. Biol. *18*, 1707–1712.
49. Jones, J.P., and Palmer, L.A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. *58*, 1233–1258.
50. Hubel, D.H., and Wiesel, T.N. (1998). Early exploration of the visual cortex. Neuron *20*, 401–412.
51. Yan, Y., Zhan, J., Ince, R.A., and Schyns, P.G. (2022). Network predictions sharpen the representation of visual features for categorization. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.01.498431.
52. Haegens, S., Händel, B.F., and Jensen, O. (2011). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. J. Neurosci. *31*, 5197–5204.
53. Todorovic, A., and de Lange, F.P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. J. Neurosci. *32*, 13389–13395.
54. Summerfield, C., and de Lange, F.P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. Nat. Rev. Neurosci. *15*, 745–756.
55. Garrido, M.I., Rowe, E.G., Halász, V., and Mattingley, J.B. (2018). Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimuli. Cereb. Cortex *28*, 1771–1782.
56. Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc. Natl. Acad. Sci. USA *108*, 20754–20759.
57. Kumar, S., Kaposvari, P., and Vogels, R. (2017). Encoding of predictable and unpredictable stimuli by inferior temporal cortical neurons. J. Cogn. Neurosci. *29*, 1445–1454.
58. Richter, D., Heilbron, M., and de Lange, F.P. (2022). Dampened sensory representations for expected input across the ventral visual stream. Oxford Open Neurosci. *1*, kvac013.
59. Richter, D., and de Lange, F.P. (2019). Statistical learning attenuates visual activity only for attended stimuli. eLife *8*, e47869.
60. Carlson, T.A., Ritchie, J.B., Kriegeskorte, N., Durvasula, S., and Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. J. Cogn. Neurosci. *26*, 132–142.
61. Yan, Y., Zhan, J., Ince, R.A.A., and Schyns, P.G. (2023). Network communications flexibly predict visual contents that enhance representations for faster visual categorization. J. Neurosci. *43*, 5391–5405.
62. Gruber, O., and Goschke, T. (2004). Executive control emerging from dynamic interactions between brain systems mediating language, working memory and attentional processes. Acta Psychol. *115*, 105–121.
63. Liu, T. (2019). Feature-based attention: effects and control. Curr. Opin. Psychol. *29*, 187–192.
64. Schoenfeld, M.A., Hopf, J.M., Martinez, A., Mai, H.M., Sattler, C., Gasde, A., Heinze, H.J., and Hillyard, S.A. (2007). Spatio-temporal analysis of feature-based attention. Cereb. Cortex *17*, 2468–2477.
65. Summerfield, C., and Egner, T. (2016). Feature-based attention and feature-based expectation. Trends Cogn. Sci. *20*, 401–404.
66. Lawrence, S.J.D., Formisano, E., Muckli, L., and de Lange, F.P. (2019). Laminar fMRI: applications for cognitive neuroscience. Neuroimage *197*, 785–791.
67. Huber, L., Finn, E.S., Chai, Y., Goebel, R., Stirnberg, R., Stöcker, T., Marrett, S., Uludag, K., Kim, S.G., Han, S., et al. (2021). Layer-dependent functional connectivity methods. Prog. Neurobiol. *207*, 101835.
68. Stephan, K.E., Petzschner, F.H., Kasper, L., Bayer, J., Wellstein, K.V., Stefanics, G., Pruessmann, K.P., and Heinzle, J. (2019). Laminar fMRI and computational theories of brain function. Neuroimage *197*, 699–706.
69. Schyns, P.G., Gosselin, F., and Smith, M.L. (2009). Information processing algorithms in the brain. Trends Cogn. Sci. *13*, 20–26.
70. Schyns, P.G., Zhan, J., Jack, R.E., and Ince, R.A.A. (2020). Revealing the information contents of memory within the stimulus information representation framework. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190705.
71. Zhan, J., Garrod, O.G.B., and O.V.R.. (2018). Someone like you? Modelling face memory reveals task-generalizable representations. Preprint at PsyArXiv. https://doi.org/10.31234/osf.io/wsvu8.
72. Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., and Gore, J.C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat. Neurosci. *2*, 568–573.
73. Malcolm, G.L., Nuthmann, A., and Schyns, P.G. (2014). Beyond gist: strategic and incremental information accumulation for scene categorization. Psychol. Sci. *25*, 1087–1097.
74. Archambault, A., O'Donnell, C., and Schyns, P.G. (1999). Blind to object changes: when learning the same object at different levels of categorization modifies its perception. Psychol. Sci. *10*, 249–255.
75. Schyns, P.G., Snoek, L., and Daube, C. (2022). Degrees of algorithmic equivalence between the brain and its DNN models. Trends Cogn. Sci. *26*, 1090–1102.
76. Jack, R.E., and Schyns, P.G. (2017). Toward a social psychophysics of face communication. Annu. Rev. Psychol. *68*, 269–297.
77. Kay, K., Bonnen, K., Denison, R.N., Arcaro, M.J., and Barack, D.L. (2023). Tasks and their role in visual neuroscience. Neuron *111*, 1697–1713.
78. Gosselin, F., and Schyns, P.G. (2010). Bubbles: a new technique to reveal the use of information in recognition tasks. J. Vision *1*, 333.
79. de-Wit, L., Alexander, D., Ekroll, V., and Wagemans, J. (2016). Is neuroimaging measuring information in the brain? Psychon. Bull. Rev. *23*, 1415–1428.
80. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al. (2013). MEG and EEG data analysis with MNE-Python. Front. Neurosci. *7*, 267.
81. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hämäläinen, M.S. (2014). MNE software for processing MEG and EEG data. Neuroimage *86*, 446–460.
82. Taulu, S., and Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. *51*, 1759–1768.
83. Taulu, S., and Kajola, M. (2005). Presentation of electromagnetic multichannel data: the signal space separation method. J. Appl. Phys. *97*, 124905.
84. King, J.-R., and Gramfort, A. (2018). Encoding and decoding neuronal dynamics: methodological framework to uncover the algorithms of cognition. Preprint at HAL Open Science. https://hal.science/hal-01848442.
85. Ince, R.A., Giordano, B.L., Kayser, C., Rousselet, G.A., Gross, J., and Schyns, P.G. (2017). A statistical framework for neuroimaging data

analysis based on mutual information estimated via a gaussian copula. Hum. Brain Mapp. *38*, 1541–1573.

86. Quian Quiroga, R., and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. Nat. Rev. Neurosci. *10*, 173–185.
87. Smith, S.M., and Nichols, T.E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage *44*, 83–98.
88. Nichols, T.E., and Holmes, A.P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. *15*, 1–25.
89. King, J.R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. *18*, 203–210.
90. Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage *87*, 96–110.
91. Franzen, L., Delis, I., De Sousa, G., Kayser, C., and Philiastides, M.G. (2020). Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making. Nat. Commun. *11*, 5440.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data	This paper	https://doi.org/10.17632/72ggfgw9f7.1
Stimuli	This paper	https://doi.org/10.17632/72ggfgw9f7.1
Software and algorithms		
MATLAB R2015b	Mathworks	RRID:SCR_001622
MNE	https://mne.tools	RRID:SCR_005972
Psychtoolbox-3	http://psychtoolbox.org	RRID:SCR_002881
Custom Code	This paper	https://doi.org/10.17632/72ggfgw9f7.1

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Philippe G. Schyns (philippe.schyns@glasgow.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Original and analyzed data reported in this study have been deposited to Mendeley Data: https://doi.org/10.17632/72ggfgw9f7.1. The code for analyses has been deposited to Mendeley Data: https://doi.org/10.17632/72ggfgw9f7.1. The code for stimuli reconstruction, experiment and visualization is available by request to the lead contact.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

Ten participants (18-35 years old, mean=25.1, SD=3.4, 3 males and 7 females) took part in the experiment and provided informed consent. All had normal or corrected-to-normal vision and reported no history of any psychological, psychiatric, or neurological condition that might affect visual or auditory perception. The University of Glasgow College of Science and Engineering Ethics Committee approved the experiment (Application Number: 300190094).

METHOD DETAILS

Stimuli

At the Prediction Stage (see Figure 1B), a different pure tone cued Voltaire, Nuns, or had no predictive value (described in cue-feature training below). The Categorization Stage showed a different hybrid image on each trial. We detail these stimuli below.

Prediction Stage: Auditory cues

Pure tones were played for 250 ms, with auditory frequencies of 196Hz (cueing Voltaire), 1760Hz (cueing Nuns) or 880Hz (no prediction).

Categorization Stage: Hybrid stimuli

We cropped a grey-level copy of Dali's *Slave Market with the Disappearing Bust of Voltaire* to retain the ambiguous part of the image that simultaneously shows the bust of Voltaire and the two Nuns across different image scales (256 × 256 pixels, Figure 1A, Original). To extract scale information (i.e. Spatial Frequencies, SF), we decomposed the ambiguous image in 5 bands at 128 (22.4), 64 (11.2), 32 (5.6), 16 (2.8), 8 (1.4) cycles per image (c/deg of visual angle), with brightness set at 0.55.

To quantify the image pixels that underlie selective perceptions of Voltaire and Nuns, we computed the Mutual Information (MI) between the pixel visibility each SF band and perceptual decisions, using data from a previous study.[25] Specifically, we computed MI(Voltaire vs. Don't Know; pixel visibility) and MI(Nuns vs. Don't Know; pixel visibility), FWER-corrected over all pixels, $p<0.05$, one-tailed. We found significant pixels at 8 cycles/image for Voltaire and at 32 and 16 cycles/image for Nuns.

To synthesize Hybrid stimuli, we Gabor filtered these significant Voltaire and Nuns pixels to represent them with Gabor coefficients at 6 orientations (0, 30, 60, 90, 120, 150 deg.). Then, we randomly sampled X% of the Voltaire Gabor coefficients and independently Y % of the Nuns Gabor coefficients (2% resolution), while randomly and independently choosing the X and Y proportions according to the distributions shown in in Figure 1A. Finally, we added the X% Voltaire to the Y% Nuns coefficients and shuffled the Gabor coefficients across the background image. We preserved the original contrasts of the Voltaire, Nuns and noisy background Gabors and set their brightness to the same 0.6 value. We presented the stimuli at 5.72° × 5.72° of visual angle on a projector screen at a viewing distance of 115cm.

Procedure

Auditory localizer

Prior to the main experiment, we ran a MEG localizer session to model the bottom-up processing of each auditory cue. Each trial started with an 250ms pure tone, followed by a 750ms ITI blank. In each 12-trial block, 10 presented the same primary tone; the remaining two tones were catch trials. We instructed participants to press a key whenever they heard a catch tone. Each participant completed 48 such blocks of trials, repeated 16 times for each primary tone, for a total of 576 trials (160 trials per tone).

Visual localizer

Prior to the main experiment, we ran another MEG localizer to model the bottom-up processing of 100% Nuns and 100% Voltaire visual features. Each trial started with a 250ms image (with 100% Nuns features, 100% Voltaire features on a mid-grey background) followed by a 1s ITI blank screen. In each 11-trial block, 10 presented the same primary features (e.g. of Voltaire); the remaining image was a catch trial (e.g. of the Nuns). We instructed participants to press a key whenever they saw a catch image. Each participant completed 48 such blocks (i.e., 24 blocks per primary image), for a total of 528 trials (240 trials per primary image).

Cue-feature training

Following the completion of the localizer tasks, we trained each participant to learn the association between the auditory cues for Voltaire and Nuns and the 100% Voltaire and Nuns features. Each trial started with an auditory cue, followed by a 1s blank screen and an image (100% Nuns or Voltaire) that they categorized as "Voltaire" or "Nuns" (2AFC), followed by feedback (correct vs. incorrect). All participants achieved over 95% accuracy in 75 trials, while implicitly learning the coupling between cues and perceptions. In a second training phase, participants heard an auditory cue and chose the corresponding image (amongst 100% Nuns vs. 100% Voltaire vs. blank image, 3AFC task), followed by feedback (correct vs. incorrect). All participants achieved >90% accuracy in 36 trials.

Cueing experiment (Figure 1B)

Prediction Stage. Each trial started with a 250ms pure tone (Voltaire, 196Hz; Nuns, 1,760Hz; neutral, 880Hz, each presented on 1/3 of all trials), followed by a 1s blank screen.

Categorization Stage. Started with a 500ms fixation followed by a Hybrid image on the center of the screen that remained until response. Participants responded "Voltaire" vs. "Nuns" vs. "Don't know" as quickly as they possibly could (3-AFC). A 750ms to 1.25s inter-trial interval (ITI) with jitter followed the response. We counterbalanced the use of the three keys (i.e., "Voltaire," "Nuns," and "Don't know") across participants, which helped to minimize any effect from specific fingers.

Each participant completed 45 blocks of 75 such prediction-then-categorization trials, in 4-5 sessions run over 4-5 days, for a total of 3,375 trials per participant.

MEG Data Acquisition and Pre-processing

We measured each participant's MEG activity with a 306-channel Elekta Neuromag MEG scanner (MEGIN) at a 1,000Hz sampling rate. We performed the analyses according to recommended guidelines using the MNE-python software[80,81] and in-house Python/MATLAB code.

We rejected noisy channels with Maxwell filtering and visual inspection, and blocks of trials with a head movement > 0.6cm (tracked by cHPI measurement). For each remaining block, we applied signal-space separation (SSS)[82,83] to the raw data to reduce environmental noise and compensate for head movement. We band-pass filtered the data between 1-150Hz (Hamming FIR filter), notch-filtered them at 50, 100 and 150Hz and rejected muscle artifacts with automatic detection. We epoched the output data into [-200ms to 2.2s] trial windows around cue onset (visual stimulus onset is at 1.75s), and rejected jump artifacts with automatic detection. We concatenated the epoched data of all blocks per session (i.e., ~10-16 blocks/day), decomposed the output dataset with ICA, identified and removed the independent components corresponding to artifacts (eye movements, heartbeat—i.e., 2-5 components/participant).

We resampled the output data at 250 Hz, low-pass filtered them at 25Hz (5^{th} Hamming FIR filter) and performed the minimum-norm estimate (MNE) analysis with an empty-room recording. We reconstructed the time series of MEG sources on a 5mm grid of boundary element model (BEM) surface (computed with Freesurfer and MNE software per participant). We applied this reconstruction to each session of trials.

These computations produced for each participant a matrix of single-trial MEG response time series–of dimensions 8,196 MEG sources x 250Hz sampling rate. We applied the same pre-processing pipeline to the MEG localizer, using the epoched data [0 to 1000ms] following auditory and visual stimulus onset.

QUANTIFICATION AND STATISTICAL ANALYSIS

Cued-Predictions bias perceptual decisions

To reveal the influence of prediction on decision behavior, we computed the relationship between cues, stimulus feature evidence and perceptual decision probabilities. Pooling all trials across all participants, we quantified the stimulus feature evidence presented on each trial as the difference between its percentage of Voltaire and Nuns Gabor features. We then binned the trials by levels of feature evidence (from -100% to 100%, with 10% steps, Figure 1C, X axis) and calculated "Voltaire," "Nuns" and "Don't Know" response probabilities (Figure 1C, Y axis), for each cue type (i.e. Voltaire, Neutral and Nuns panels of Figure 1C). We regressed the feature evidence and the decision probabilities with local linear Gaussian kernels. We computed the Point of Subjective Equality (PSE) as the level of stimulus evidence (i.e. of Voltaire% - Nuns% feature evidence) for which Voltaire and Nuns decisions are equiprobable (Figure 1C).

Category-contrast decoding of predictions

Localizer cross-validation

To model the bottom-up representations of the auditory cues and of the visual Gabor features, we trained different linear classifiers[84] (Linear Discriminant Analysis, LDA) using the MEG localizer sensor data. That is, separately for the auditory and visual localizers, we randomly segmented each participant's MEG trials into 5 folds (without repetitions) and performed a 5-fold cross-validation.

In each validation iteration, we proceeded in 2 steps

Step 1: Training. We trained linear classifiers[84] (with MNE decoding module[80]) to discriminate the Voltaire vs. Nuns auditory cue and the 100% Voltaire vs. Nuns Gabor features, every 4ms between 0 and 400ms post stimulus, using as training set the MEG sensor data from 4 folds.

Step 2: Validation. We tested these trained classifiers every 4ms with the left-out fold. At each time point, we computed classifier decision value as the inner product of the learned linear weights with the held-out fold sensor data.

Following all 5 iterations, we proceeded to Step 3

Step 3: Cross-validation performance. To quantify decoding performance at each time point on a common scale, we concatenated trials across folds and calculated every 4ms the MI[85] between this classifier decision value and the true stimulus label (i.e. auditory cues for Voltaire vs. Nuns in the auditory classifiers; 100% Voltaire vs. Nuns Gabor features in the visual classifiers). MI quantifies the discrimination information about the stimulus that is available from the classifier weights, without forcing a discrete classification.[86]

We repeated Steps 1 to 3 three times and averaged the resulting three MI matrices (of 100 training time points ×100 testing time points) to quantify the cross-validated decoding performance. To establish statistical significance, we repeated this procedure 1,000 times with shuffled stimulus labels. We applied Threshold-Free Cluster Enhancement[87] (TFCE, E=0.5, H=0.5), thresholding for significance with the 95th percentile of the distribution of 1,000 maximum values, each taken across all training × testing time points in the temporal generalization matrix of each shuffle after TFCE (FWER corrected over 100 localizer training time points and 100 localizer testing time points,[88] p<0.05, one-tailed). The resulting matrices of significant MI comprise the time points with significant cross-validation performance. We repeated this cross-validation independently for each participant.

Category-contrast reactivation

We used category-contrast classifiers to compute temporal generalization cross-decoding[89] on the bottom-up processing of the auditory cues contrast and the top-down cued-reactivations of the Voltaire vs. Nuns visual contrast as follows (explained with visual classifiers):

Step 1: Training. We trained Voltaire vs. Nuns classifiers on the MEG localizer data at time points of significant cross-validation (see localizer cross-validation) between 0 and 400ms post stimulus onset.

Step 2: Testing. At the Prediction Stage, every 4 ms between 0 and 600ms post auditory cue, we computed the singular classifier decision value from single-trial MEG sensor response. This produced a 2D (training × testing time) matrix of decision values from the category-contrast classifiers, where each value indicates the reactivation strength of the category contrast at this time point.

Step 3: Reactivation performance quantification. For each training x testing time combination, we computed across trials the MI between decision values and ground truth stimulus category (i.e. Voltaire vs. Nuns Gabor features). Permutation testing (1,000 repetitions) established statistical significance (corrected for multiple comparisons with TFCE, FWER corrected over 100 localizer training time × 150 Prediction Stage testing time points, one-tailed, $p < 0.05$).

Step 4: Early vs. Late classifiers. We split the localizer classifiers into the Early (trained 0-150ms post-stimulus) and Late sets (trained 150-280ms post-stimulus). To measure performance, we chose the Early classifier and the Late classifier with maximum prediction reactivation performance. As above inference was FWER corrected over all classifiers considered.

We repeated Steps 1 to 4 to generate the performance curves of each participant. Figure 2A averages them across participants, separately for Early and Late category-contrast classifiers, for the auditory cues (indicating cue processing) and the visual category contrast (indicating prediction reactivation).

Additionally, for each participant and tested time point t of the Prediction stage, we removed the effect of the auditory cue-contrast–i.e. computing conditional MI, CMI(ground truth of Voltaire vs. Nuns cue; visual decision value$_t$ | auditory decision value$_t$) using the Late visual classifier and the auditory classifier trained at localizer time t. Figure S2A averages the CMI curves across participants.

Source representation of category-contrast

To localize the MEG source activity underlying auditory and visual category-contrast performance, we must be careful with direct interpretation of weight vectors in sensor space.[90] To address this we used a correlation forward model, where we computed MI between the classifier decision value and source activity, to determine the contribution of each source to the classifier performance.[91] We proceeded as follows:

Step 1: Time selection. We selected the two time points of the Prediction Stage when classifier performance peaks–i.e. one in Early prediction (before 150ms post cue); one in Late prediction (after 150ms post-cue).

Step 2: Source representation reconstruction. At Early and Late time points, for all 8,196 sources, we computed MI between single-trial category-contrast classifier decision values and single-trial source activity.

We repeated this two-step analysis in each participant to reconstruct their source representations of the auditory and visual category-contrasts at Early and Late Prediction time points. Figure 2B shows their group average; Figure S1 shows individual participant results.

Category-feature decoding of predictions

Category-feature classifier cross-validation

To separately model the dynamic bottom-up representations of Nuns features and Voltaire features, we trained classifiers[84] (using the MNE decoding module) at the Categorization Stage under neutral cueing. Specifically, every 4ms post-stimulus, we trained binary category-feature Voltaire classifiers on sets of trials with >70% Voltaire (i.e., the top 30% of the trial distribution) vs. <30% Voltaire (i.e., the bottom 30%), and category-feature Nuns classifiers on sets of trials with >70% Nuns vs. <30%. We segmented the participant's trials into 5 folds based on stratified sampling and performed a 5-fold cross-validation.

In each iteration, and separately for Voltaire and Nuns classifiers, we proceeded as follows:

Step 1: Training. We trained linear classifiers to discriminate >70% Voltaire vs.<30% Voltaire, every 4ms between 0 and 400ms post stimulus at the Categorization Stage, under neutral cueing, using MEG sensor data from 4 folds as the training set. We repeated training for the >70% Nuns vs. < 30% Nuns classifiers.

Step 2: Validation. We tested the trained classifiers every 4ms on the left-out fold, computing at each time point the decision value as the inner product between classifier weights and held-out fold trials.

Following all 5 iterations, we proceeded to Step 3

Step 3: Cross-validation performance. To quantify decoding performance every 4ms on a common scale, we concatenated trials across folds and calculated MI between this classifier decision value and the true stimulus label (>70% Voltaire/Nuns vs. <30% Voltaire/Nuns).

To quantify cross-validation, we repeated Steps 1 to 3 three times and averaged the resulting three MI matrices (of 100 training × 100 testing time points). We established statistical significance with permutation testing (1,000 repetitions), corrected for multiple comparisons with TFCE–FWER over 100 Categorization Stage training time points × 100 Categorization Stage testing time points, one-tailed, $p < 0.05$. The resulting matrices of significant MI comprised the time points with significant cross-validation performance. We repeated this cross-validation independently for each participant.

Category-feature reactivation

We used the category-feature classifiers to separately cross-decode reactivations of Voltaire and Nuns at the Prediction Stage. We traced their source representations as follows:

Step 1: Training. We trained the >70% vs. <30% category-feature Voltaire and category-feature Nuns classifiers at time points when they significantly cross-validate (see category-feature classifier cross-validation), using neutral-cued trials with >70% and <30% Voltaire and Nuns MEG responses.

Step 2: Testing. Every 4ms between 0 and 600ms of the Prediction Stage, we tested category-feature classifiers performance on the MEG sensor data. Each trial provides a 2D (training × testing time) matrix of decision values that quantify the reactivation strength of the category-feature prediction.

Step 3: Reactivation performance quantification. Separately for Voltaire and Nuns reactivations, and for Voltaire-cued, Nuns-cued and neutral-cued trials, we computed for each training x testing time cell of the matrix the MI between the category-feature classifier values and the ground truth cue labels (i.e. Voltaire, Nuns and neutral). We established statistical significance with permutation testing (1,000 repetitions) and corrected for multiple comparisons with TFCE–FWER over 100 Categorization Stage training time points × 150 Prediction Stage testing time points, one-tailed, $p < 0.05$. We then selected the Voltaire classifier and the Nuns classifier with maximum reactivation performance for the Voltaire Gabor features and separately for the Nuns Gabor features.

Step 5: Source representations.[90,91] To visualize the predictions of Voltaire- and Nuns-specific features on MEG sources at their peak reactivation during the Prediction Stage, we computed MI between single-trial category-specific Gabor feature classifier values and MEG source activity. We applied this analysis on all 8,196 sources covering the whole brain.

Step 6: Statistical test of lateralization. To test the significance of the lateralization in Voltaire- and Nuns- feature predictions, for each ROI (including LG, PCAL, CUN, LOC, FG, PHC), we performed independent-sample t-tests comparing the Step 5 source representations between the left and right hemispheres. We applied Bonferroni correction for multiple comparisons. We reported the prevalence of participants displaying significant lateralization patterns.

We repeated Steps 1 to 5 for each participant to produce MEG source representations of Voltaire and Nuns category-feature reactivations (see Figure 3A for their group averages, Figure S1 for per-participant results).

Specificity of cued-reactivations

To investigate how selectively the cue (for Voltaire and Nuns) reactivates predictions, we compared per participant the reactivation performance of the category-feature and the category-contrast classifiers as follows:

Step 1: Classifier selection. For each participant, across all training (0-400ms post visual stimulus) and testing time points (0-600ms post auditory cue), we selected the three classifiers with maximal reactivation performance: the Voltaire and the Nuns category-feature classifiers (see category-feature reactivation) and the category-contrast Voltaire vs. Nuns classifier (see category-contrast reactivation).

Step 2: Reactivation performance quantification. We compared these three classifiers on their classification of Prediction Stage MEG sensor data, every 4ms between 0–600ms post-cue (Figure 3B).

Reactivation biases behavior

To compare how the trial-by-trial category-feature decoding and category-contrast reactivation at Prediction change response probabilities at Categorization, we examined how their strong vs. weak reactivations change the psychometric relationship between stimulus evidence and response probabilities. For each participant, and separately for Nuns-cued and Voltaire-cued trials, we proceeded as follows:

Step 1: Time selection. Every 4ms of the Prediction Stage, we computed the difference of reactivation strength (i.e. classifier decision values) when the participant then categorizes the stimulus as "Voltaire" and as "Nuns" (e.g. on Voltaire-cued trials). We established statistical significance with permutation testing (1,000 repetitions), corrected for multiple comparisons over training × testing time, two-tailed, $p<0.05$. We extracted the single-trial reactivation strength when this significant difference is maximal.

Step 2: Reactivation split. We binned the Gabor feature evidence from -100% to 100% (10% step). In each bin, we selected trials with top vs. bottom 30% reactivation strength and compute the probabilities of "Voltaire," "Nuns," and "Don't Know" responses.

We repeated Steps 1 to 2 in each participant and computed the group median of decision probabilities for each bin of feature evidence. We then regressed the feature evidence and the group-median decision probabilities (local linear Gaussian kernel), separately for the top and bottom 30% of the trial distribution. Figures 4A and 4B compare these psychometric relationships between category-feature and category-contrast reactivations.

Localizer Linear Representation

To model the bottom-up representation of Voltaire and Nuns features in the localizer, every 4 ms between 0 and 400 ms post-stimulus, we computed independent multivariate linear regressions separately for Nuns (N) and Voltaire (V).

$$\mathbf{y} = \beta_0 + \beta_1 N$$

$$\mathbf{y} = \beta_0 + \beta_1 V$$

We fitted each model with least-squares, resulting in beta coefficients for the intercept and slope. We computed the fit in the MEG activity of the source with a multivariate R^2 that quantifies multivariate variance as the determinant of the covariance matrix:

$$R^2 = 1 - \frac{\left|(\boldsymbol{y} - \widehat{\boldsymbol{y}})^T(\boldsymbol{y} - \widehat{\boldsymbol{y}})\right|}{\left|(\boldsymbol{y} - \overline{\boldsymbol{y}})^T(\boldsymbol{y} - \overline{\boldsymbol{y}})\right|}$$

where $\boldsymbol{y}, \overline{\boldsymbol{y}}, \widehat{\boldsymbol{y}}$ are respectively source activity, its mean and the model prediction. This linear modelling produced a time course of R^2 values per source every 4ms. For each participant, we localized the Voltaire and Nuns source representation at the time point with highest R2 between 150-280ms post-stimulus. Figure S2 shows the average source representations across participants.

REACTIVATION OF LEFT VS. RIGHT NUNS FACE FEATURES

We used the left-nun and right-nun classifiers to separately cross-decode reactivations of the left and right Nuns at the Prediction Stage and their representations at the Categorization Stage. We traced their source representations as follows:

Step 1: Training. Every 4ms between 0 and 500ms, across all participants Categorization stage neutral-cued MEG responses, we trained high vs. low visibility feature classifiers separately for left and right nun face features. We used two images for high visibility and two images for low visibility of each feature (left feature, 46% and 76% for high visibility vs. 0% and 2% for low visibility, magenta in Figure S3; right feature, 42% and 56% vs. 0% and 2%, green).

Then, for each participant, we proceeded as follows

Step 2: Prediction Testing. On Nuns and neutral cued trials, every 4ms between 0 and 600ms of the Prediction Stage, we tested left and right nuns classifiers performance on MEG sensor data. Each trial provides a 2D (training × testing time) matrix of decision values that separately quantifies the reactivation strength of the left nun and right nun predictions. For each training x testing time cell of the matrix, we computed prediction performance as the MI between the left and right nuns classifier values and the cue (i.e. Nuns vs. neutral).

Step 3: Categorization Testing. Every 4ms between 0 and 500ms of the Categorization Stage with Nuns cue, we tested left and right nuns classifiers performance on MEG sensor data. For each training x testing time cell of the matrix, we computed categorization performance as the MI between nuns proportions of the stimuli and the left and right nuns classifier values.

Step 4: Comparison. To compare the prediction and categorization of left and right nun features, we took for each the maximum performance over the Prediction and Categorization stages across all training x testing time points.

Step 5: Source representations. To visualize predictions of left and right nun features with the highest performance at Prediction, we computed MI between single-trial classifier values and MEG source activity. We applied this analysis on all 8,196 sources covering the whole brain.

Step 6: Bias for behavioral response. For each stimulus pixel, on neutral cued trials we computed MI between single-trial pixel values and participant's responses (Nuns vs. other responses), establishing statistical significance with permutation testing–i.e. 100 repetitions, corrected for multiple comparisons over 256 × 256 pixels, one-tailed, $p<0.05$. Across the pixels representing the left-nun and separately the right nun, we computed the median of significant MI and compute the difference between these two medians. This difference indicates the bias for the left- or right-nun pixels for behavioral response.

We repeated Steps 2 to 6 for each participant to produce MEG source representations of left or right Nuns category-feature reactivations (see Figure S3 for per-participant results).