JNeurosci Accepted Manuscript

JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles: Behavioral/Cognitive

A Gradient of Sharpening Effects by Perceptual Prior across the Human Cortical Hierarchy

https://doi.org/10.1523/JNEUROSCI.2023-20.2020

Cite as: J. Neurosci 2020; 10.1523/JNEUROSCI.2023-20.2020

Received: 30 July 2020
Revised: 6 November 2020
Accepted: 6 November 2020

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2020 the authors

JNeurosci Accepted Manuscript

A Gradient of Sharpening Effects by Perceptual Prior across the Human Cortical Hierarchy

Carlos González-García[1] and Biyu Jade He[2,3*]

[1] Department of Experimental Psychology, Ghent University, Ghent, Belgium B-9000.

[2] Neuroscience Institute, New York University Langone Medical Center, NY 10016.

[3] Departments of Neurology, Neuroscience and Physiology, and Radiology, New York University Langone Medical Center, NY 10016.

* Correspondence should be addressed to BJH at biyu.jade.he@gmail.com.

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgments: This research was supported by National Science Foundation (BCS-1926780, to BJH). CGG was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 835767.

JNeurosci Accepted Manuscript

ABSTRACT

Prior knowledge profoundly influences perceptual processing. Previous studies have revealed consistent suppression of predicted stimulus information in sensory areas, but how prior knowledge modulates processing higher up in the cortical hierarchy remains poorly understood. In addition, the mechanism leading to suppression of predicted sensory information remains unclear, and studies thus far have revealed a mixed pattern of results in support of either the 'sharpening' or 'dampening' model. Here, using 7T fMRI in humans (both sexes), we observed that prior knowledge acquired from fast, one-shot perceptual learning sharpens neural representation throughout the ventral visual stream, generating suppressed sensory responses. In contrast, the frontoparietal (FPN) and default-mode (DMN) networks exhibit similar sharpening of content-specific neural representation but in the context of unchanged and enhanced activity magnitudes, respectively—a pattern we refer to as 'selective enhancement'. Together, these results reveal a heretofore unknown macroscopic gradient of prior knowledge's sharpening effect on neural representations across the cortical hierarchy.

SIGNIFICANCE STATEMENT

A fundamental question in neuroscience is how prior knowledge shapes perceptual processing. Perception is constantly informed by internal priors in the brain acquired from past experiences, but the neural mechanisms underlying this process are poorly understood. To date, research on this question has focused on early visual regions, reporting a consistent downregulation when predicted stimuli are encountered. Here, using a dramatic one-shot perceptual learning paradigm, we observed that prior knowledge results in sharper neural representations across the cortical hierarchy of the human brain through a gradient of mechanisms. In visual regions, neural responses tuned away from internal predictions are suppressed. In frontoparietal regions, neural activity consistent with priors is selectively enhanced. These results deepen our understanding of how prior knowledge informs perception.

INTRODUCTION

Prior knowledge strongly shapes perception. For instance, we perceive a patch that is brighter at the top as concave, due to our lifetime experience with light coming from above (Ramachandran, 1988). Internal priors are thought to interact with bottom-up inputs in an iterative interplay that leads to perception (Friston, 2005; Yuille and Kersten, 2006; Albright, 2012). Although significant effort has been devoted to characterizing the neural correlates of such interactions in early sensory areas, how priors impact perceptually relevant neural processing in higher-order regions remains unclear. Earlier studies have shown that prior knowledge obtained from task cues elicits enhanced activity in posterior parietal

JNeurosci Accepted Manuscript

cortex, medial and lateral prefrontal cortex (PFC) as well as enhanced connectivity between PFC and visual areas, pointing to frontoparietal regions as potential sources of predictions guiding perception (Summerfield et al., 2006; Eger et al., 2007; Esterman and Yantis, 2010; Rahneet al., 2011) and top-down templates influencing information processing (Desimone and Duncan, 1995; Buschman and Kastner, 2015). However, how exactly neural representations in frontoparietal areas are altered by prior knowledge remains poorly understood. Recent studies showed that frontoparietal areas contain content-specific neural activity patterns that are strongly modulated by the availability of prior knowledge (González-García et al., 2018; Flounders et al., 2019), but the relative contribution of neural populations tuned towards or away from prior knowledge remains unknown. Answering this question would help explain how prior knowledge is encoded in neural populations and how it might be propagated across large-scale brain networks to inform perceptual processing.

At the same time, current models suggest that top-down predictions generated in higher-order cortical regions interact with sensory processing in lower-order regions to guide perception (Bar et al., 2006; Yuille and Kersten, 2006; Albright, 2012). A common observation is that perceptual expectations reduce activity in visual regions (Murray et al., 2002; de Lange et al., 2018), which supports a 'predictive coding' account of how such top-down generative models facilitate perception (Mumford, 1992; Friston, 2005). This downregulation, or 'expectation suppression', of sensory cortical activity for expected input can be explained by two alternative accounts (de Lange et al., 2018). Under the first account, expectation suppression arises as a consequence of filtering out the expected neural responses that are redundant with high-level predictions (Murray et al., 2004; Friston, 2005); thus, prior knowledge *dampens* activity in neural populations tuned towards the predicted stimulus (Fig. 1A; 'dampening' model). Alternatively, expectation suppression could result from *sharpening* of neural representation such that neural activity tuned away from (hence inconsistent with) the expected features are suppressed, resulting in a sharper neural representation (Fig. 1B; 'sharpening' model) (Lee and Mumford, 2003). We note that both models refer to changes in population neural responses while being agnostic about the underlying tuning curves of individual neurons.

Thus far, studies investigating the effect of prior knowledge on visual cortical representations have yielded mixed results, with some supporting the 'dampening' model (Meyer and Olson, 2011; Kumar et al., 2017; Schwiedrzik and Freiwald, 2017; Richter et al., 2018) and others supporting the 'sharpening' model (Kok et al., 2012; Bell et al., 2016). These previous studies typically employed clear images and induced prior knowledge using either top-down instructions or statistical learning. We reasoned that in real life, the impact of prior knowledge on perceptual processing is particularly pronounced when stimulus input is weak or ambiguous, such as detecting a camouflaged or hidden predator. In addition, in real life, the prior knowledge that guides perception typically derives from relevant past experiences. For instance, a previous sight of a predator in clear view will help guide future detection of the same

JNeurosci Accepted Manuscript

type of predators hidden or camouflaged. In such cases, animals and humans do not have the benefit of top-down instruction or the luxury of repeated associative learning.

Here, we combined high-field (7 Tesla) fMRI with a dramatic visual phenomenon to assess the impact of prior knowledge obtained from past experiences on the neural processing of ambiguous stimuli across the cortical hierarchy. Specifically, we employed Mooney images, black-and-white degraded images that are difficult to recognize at first. However, once participants are exposed to their unambiguous version (a process called “disambiguation”), recognition of the same Mooney images becomes effortless and can last for long periods of time (Ludmer et al., 2011; Albright, 2012). Thus, these images provide a robust, well-controlled experimental paradigm to investigate prior knowledge’s influence on perceptual processing in an ecologically relevant manner (e.g., anyone who has taken psychology 101 recognizes the Dalmatian dog picture instantly). Previous studies using Mooney images have revealed an increased similarity in the neural representation between post-disambiguation Mooney images and their unambiguous counterparts in regions along the ventral visual stream (Hsieh et al., 2010; Van Loon et al., 2016) as well as the FPN and DMN (González-García et al., 2018). However, these correlational approaches are agnostic regarding the specific tuning properties of neural populations before and after disambiguation, since positive correlations could reflect both dampening and sharpening of responses.

To anticipate the findings, we found that prior knowledge induces a sharper neural representation throughout the cortical hierarchy, with overall suppressive effects in the sensory end of the hierarchy, overall enhancing effects in the DMN end of the hierarchy, and little net effect on activation magnitudes in the intermediate FPN (Fig. 1B-D). These results reveal a macroscopic gradient of prior knowledge’s effect on neural representations, strengthen the evidence for a sharpening account of prior knowledge’s effect in sensory areas, and point to a new mechanistic model for explaining neural sources of prior knowledge: selective enhancement.

METHODS

Subjects

Twenty-three healthy volunteers participated in the study. All participants were right-handed and neurologically healthy, with normal or corrected-to-normal vision. The experiment was approved by the Institutional Review Board of the National Institute of Neurological Disorders and Stroke. All subjects provided written informed consent. Four subjects were excluded due to excessive movements in the scanner, leaving 19 subjects for the analyses reported herein (age range = 19 – 32; mean age = 24.6; 11 females). The data set analyzed herein was previously published in González-García et al., 2018.

JNeurosci Accepted Manuscript

125 **Visual stimuli**

126 Thirty-three Mooney and gray-scale images were generated from gray-scale photographs of real-world
127 man-made objects and animals selected from the Caltech
128 (http://www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html) and Pascal VOC
129 (http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html) databases. First, gray-scale images were
130 constructed by cropping gray-scale photographs with a single man-made object or animal in a natural
131 setting to 500 x 500 pixels and applying a box filter. Mooney images were subsequently generated by
132 thresholding the gray-scale image. Threshold level and filter size were initially set at the median
133 intensity of each image and 10 $\times$ 10 pixels, respectively. Each parameter was then titrated so that the
134 Mooney image was difficult to recognize without first seeing the corresponding gray-scale image. An
135 independent subject group (N = 6) was used to select the 33 images used in this study from an initial
136 set of 252 images (for details, see (Chang et al., 2016)). Images were projected onto a screen located
137 at the back of the scanner and subtended approximately 11.9 $\times$ 11.9 degrees of visual angle.

138 **Experimental design**

139 Each trial started with a red fixation cross presented in the center of the screen for 2 seconds, and then
140 a Mooney image or a gray-scale image presented for 4 seconds. The fixation cross was visible during
141 image presentation, and subjects were instructed to maintain fixation throughout. A 2-sec blank period
142 appeared next, followed by a brighter fixation cross (lasting 2 sec) that prompted participants to
143 respond (see Fig. 1A). Participants were instructed to respond to the question “Can you recognize and
144 name the object in the image?” with an fMRI-compatible button box using their right thumb. Trials were
145 grouped into blocks. Each block contained fifteen trials: three gray-scale images followed by six
146 Mooney-images and a shuffled repetition of the same six Mooney-images. Three of the Mooney images
147 had been presented in the previous run and corresponded to the gray-scale images in the same block
148 (these are post-disambiguation Mooney images). The other three Mooney-images were novel and did
149 not match the gray-scale images (pre-disambiguation); their corresponding gray-scale images would be
150 presented in the following run. Each fMRI run included three blocks of the same images; within each
151 block, image order was shuffled but maintained the same block structure (gray-scale followed by
152 Mooney images). After each run, a verbal test was conducted between fMRI runs. During the verbal
153 test, the six Mooney images from the previous run were presented one by one for 4 sec each, and
154 participants were instructed to verbally report the identity of the image. Each participant completed 12
155 runs. In order for all Mooney images to be presented pre- and post- disambiguation, the first and last
156 runs were “half runs”. The first run contained only three novel Mooney images (pre-disambiguation).
157 The last run consisted of 3 gray-scale images and their corresponding post-disambiguation Mooney

JNeurosci Accepted Manuscript

images. The total duration of the task was ~90 minutes. The order of Mooney images presentation was randomized across participants.

Data acquisition and preprocessing

Imaging was performed on a Siemens 7T MRI scanner equipped with a 32-channel head coil (Nova Medical, Wilmington, MA, USA). T1-weighted anatomical images were obtained using a magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) sequence (sagittal orientation, 1×1×1 mm resolution). Additionally, a proton-density (PD) sequence was used to obtain PD-weighted images also with 1×1×1 mm resolution, to help correct for field inhomogeneity in the MP-RAGE images. Functional images were obtained using a single-shot echo planar imaging (EPI) sequence (TR = 2000 ms, TE = 25 ms, flip angle = 50°, 52 oblique slices, slice thickness = 2 mm, spacing = 0 mm, in-plane resolution = 1.8×1.8 mm, FOV = 192 mm, acceleration factor / GRAPPA = 3). We note that because signal-to-noise ratio decreases with smaller voxel size, in order to maintain a high signal-to-noise ratio with whole-brain coverage, we chose a modest spatial resolution. Respiration and cardiac data were collected using a breathing belt and a pulse oximeter, respectively.

For anatomical data preprocessing, MP-RAGE and PD images were first skull-stripped. Then, the PD image was smoothed using a 2-mm full-width at half maximum (FWHM) kernel. Afterwards, the MP-RAGE image was divided by the smoothed PD image to correct for field inhomogeneity. Functional data preprocessing started with the removal of physiological (respiration- and cardiac-related) noise using the RETROICOR method (Glover et al., 2000). The next preprocessing steps were performed using the FSL package (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL). These included: ICA cleaning to remove components corresponding to physiological or movement-related noise, >0.007 Hz high-pass filtering to remove low-frequency drifts, rigid-body transformation to correct for head motion within and across runs, slice timing correction to compensate for systematic differences in the time of slice acquisition, and spatial smoothing with a 3-mm FWHM Gaussian kernel. Last, images were registered to the atlas in two steps. First, functional images were registered to the individual's anatomical image via global rescale (7 degrees of freedom) transformations. Affine (12 degrees of freedom) transformations were then used to register the resulting functional image to a 2×2×2 mm MNI atlas. Registration to MNI space was performed on voxel-wise GLM and searchlight MVPA results to obtain population-level whole-brain maps. RSA analyses were conducted in individual subject's functional data space and results were pooled across subjects.

Region-of-interest (ROI) definition

ROIs (shown in Fig. 1F-G) were defined as follows (for further details, see González-García et al., 2018). A separate retinotopic localizer and a lateral occipital complex (LOC) functional localizer were

JNeurosci Accepted Manuscript

performed for each subject to define bilateral early visual ROIs and LOC, respectively. The average numbers of voxels across subjects are: right V1 (577 voxels, vx), left V1 (550 vx), right V2 (637 vx), left V2 (564 vx), right V3 (540 vx), left V3 (477 vx), right V4 (256 vx) and left V4 (198 vx), left LOC (1195 vx) and right LOC (744 vx).

Left (4176 vx) and right (4572 vx) fusiform gyrus (FG) ROIs were extracted using the Harvard-Oxford Cortical Structural Atlas (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases). Default mode network (DMN) regions, including bilateral lateral parietal cortices (LatPar; left: 1361 vx; right: 1469 vx), medial prefrontal cortex (mPFC; 1135 vx), and posterior cingulate cortex (PCC; 2145 vx), were defined using a general linear model (GLM) of the disambiguation contrast (pre-disambiguation-not-recognized vs. post-disambiguation-recognized). Lastly, statistical map from searchlight decoding of the disambiguation effect (i.e., decoding recognition status: pre-not-recognized vs. post-recognized) was used to define the frontoparietal network (FPN) ROIs: bilateral frontal (left: 13383 vx; right: 14427 voxels) and parietal (left: 3949 vx; right: 3381 vx) cortices.

Additionally, we replicated results from FPN and DMN using ROIs defined from an independent resting-state data set (Power et al., 2011).

Statistical Analysis

Univariate analysis of ROI activity. A General Linear Model (GLM) was used to assess changes in activation magnitude across conditions. At the individual subject level, a model was constructed including regressors for each of the 33 images under the three experimental conditions (pre-, post-disambiguation, and gray-scale conditions), resulting in a total of 99 regressors. All regressors were convolved with a hemodynamic response function (HRF). After model estimation, regressors of “pre-disambiguation not-recognized” and “post-disambiguation recognized” images were selected. To do so, we used subjective recognition responses: based on the bimodal distribution of recognition responses (Fig. 2C), an image was defined as not-recognized if the subject responded “yes” in 2 or fewer out of 6 presentations of that image; it was defined as recognized if the subject responded “yes” in 4 or more out of 6 presentations. Importantly, only images fulfilling both criteria were selected for analyses; therefore, for each subject the same set of images was used in the pre- and post-disambiguation conditions (14 ± 3).

For each subject and ROI, we obtained the mean beta estimate for the pre-disambiguation and post-disambiguation conditions, averaged across voxels and images. First, beta estimated of each condition were compared to baseline using across-participant two-tailed one-sample t-tests. Then, they were submitted to across-participant paired t-tests to compare activity levels before and after disambiguation. In all cases, results were FDR-corrected for multiple comparisons across all ROIs.

JNeurosci Accepted Manuscript

Image preference analysis. To perform the image preference analysis (see schematic in Fig. 1E), we used the same beta estimates as in the GLM analysis, ensuring that pre- and post-disambiguation images were matched for each subject. For each voxel within a given ROI, images were ranked according to this voxel's activation level in the gray-scale condition.

For each voxel, this image preference ranking based on its activation level to gray-scale images was then applied to the same voxel's beta estimates from the pre- and post-disambiguation conditions. Thus, for each voxel, we obtained a vector containing its beta estimates for each Mooney image ranked by how strongly this voxel activated to its (prior-inducing) matching gray-scale image, separately for pre- and post- disambiguation conditions. Since the images included in the analysis depended on individual behavioral performance, the number of images varied across participants. To account for the different number of images, these vectors were resampled to 10 bins. We then averaged the beta estimates across voxels for each bin in the ranking, separately for pre- and post- conditions. To test whether the preference ranking of gray-scale images generalized to pre- and post-disambiguation conditions we fitted a linear regression to each vector of ranked beta estimates. This allowed us to obtain the slope coefficient for each condition. These coefficients were then subjected to an across-participant right-tailed one-sample Wilcoxon test against 0 to test for the presence of significant positive regression slopes, which would indicate a generalization of gray-scale image preference ranking to the Mooney image condition. Subsequently, we compared the slopes of pre- and post-disambiguation conditions using two-tail paired Wilcoxon tests. In this analysis, a larger slope coefficient in the pre- than post-disambiguation condition indicates a dampening mechanism (Fig. 1A), whereas the opposite scenario (a steeper slope for post- than pre-disambiguation data) indicates a sharpening mechanism (Fig. 1B-D). In all tests, results were FDR-corrected for multiple comparisons across all ROIs investigated.

RESULTS

Behavioral results

Nineteen subjects were shown 33 Mooney images containing animals or manmade objects. Each Mooney image was presented six times before its corresponding gray-scale image, and six times after. Following each Mooney image presentation, subjects reported whether they could recognize the image using a button press ("subjective recognition"). Each fMRI run included three distinct gray-scale images, their corresponding post-disambiguation Mooney images, and three new Mooney images shown pre-disambiguation (their corresponding grayscale images would be shown in the next run), with randomized order within each stage such that a gray-scale image was rarely followed immediately by its corresponding Mooney image (Fig. 2A; for details, see Methods, *Task paradigm*). To verify whether subjects correctly identified the Mooney images, at the end of each run, Mooney images presented

JNeurosci Accepted Manuscript

during that run were shown again and participants were asked to verbally report the identity of the image and were allowed to answer “unknown”. This resulted in a verbal test for each Mooney image once before disambiguation and once after disambiguation (“verbal identification”). Verbal responses were scored as correct or incorrect using a pre-determined list of acceptable responses for each image.

Disambiguation by viewing the gray-scale images had a substantial effect on participants’ subjective recognition responses, with a significantly higher rate of recognition for Mooney images presented post-disambiguation (86±1%; mean ± s.d. across subjects) compared to the same images presented before disambiguation (43±12%; $t_{1,18}$ = 18.6, p = 3.4e-13, Cohen’s d = 4.2; Fig. 2B). A similar pattern of results was observed using the verbal identification responses: Mooney images were correctly identified significantly more often after disambiguation (86±0.8%) than before (34±12%; $t_{1,18}$ = 25.3, p = 1.7e-15, Cohen’s d = 5.8).

Based on the bimodal distribution of subjective recognition rates across images (Fig. 2C), we established cut-offs to select “pre-(disambiguation) not-recognized” images as those recognized 2 or fewer times before disambiguation and “post-(disambiguation) recognized” images as those recognized 4 or more times after disambiguation. Accuracy of verbal identification responses in these two groups were 8.7±5.8%, and 93.6±4.5%, respectively. Using these cut-offs, for each participant we selected Mooney images that were both not-recognized in the pre-disambiguation stage and recognized in the post-disambiguation stage for further analyses.

We defined twenty regions of interest (ROIs, see Fig. 1F-G and Methods, *ROI definition*) that cover early visual areas, category-selective visual areas, the FPN and DMN, all of which show disambiguation-induced increase in content-specific information (González-García et al., 2018) for further analyses.

Priors suppress and sharpen representations in early and higher-order visual regions

To examine the impact of prior knowledge acquisition on neural representations in the ventral visual stream, we first tested for prior-induced suppression by comparing the overall activation magnitudes for pre- and post-disambiguated images. In all visual ROIs, from early visual regions (V1–V4; Fig. 3A, black asterisks) to category-selective visual regions (lateral occipital complex, LOC, and fusiform gyrus, FG; Fig. 4A, black asterisks), post-disambiguation images elicited weaker BOLD responses compared to pre-disambiguation images (all $p < 0.0005$; $q < 0.05$, FDR-corrected across all 20 ROIs; henceforth “FDR-corrected”, all Cohen’s $d > 0.95$). Pre-disambiguation images triggered significant activation in all visual ROIs (blue asterisks in Fig. 3A and 4A, all $p < 0.03$, $q < 0.05$, FDR-corrected, all $d > 0.67$), while post-disambiguation images triggered significant activation in V1, V2, V3 and the LOC (red asterisks in Fig. 3A and 4A, $p < 0.02$, $q < 0.05$, FDR-corrected, $d > 0.67$; for the rest of ROIs, corrected $p > 0.1$). Thus, the availability of prior knowledge resulted in suppressed neural responses in visual regions,

JNeurosci Accepted Manuscript

consistent with previous studies reporting an expectation suppression effect (Murray et al., 2002; Meyer and Olson, 2011; Summerfield and de Lange, 2014). However, these results remain agnostic regarding the specific mechanism (dampening or sharpening; Fig. 1A-B) underlying such suppression.

To adjudicate between these alternative models, we performed an image preference analysis (Richter et al., 2018) (see Fig. 1E for a schematic of the analysis, and Methods, *Image preference analysis* for details). As a first step, for each ROI and participant, we extracted BOLD activation magnitudes (beta values per voxel) corresponding to each Mooney image under the two experimental conditions (pre- and post-disambiguation), as well as during the presentation of their matching unambiguous, gray-scale counterparts. For each voxel within the ROI, images were ranked according to its activation level in the gray-scale condition from the 'least preferred' to the 'most preferred'. This step allowed us to characterize voxel tuning preference based on the unambiguous images inducing prior knowledge. We then asked whether the expectation suppression effect observed above was driven by voxels tuned towards or away from the predicted features, which would support the dampening and sharpening account, respectively.

Thus, for each voxel, we obtained a vector containing the beta estimates for each Mooney image, ranked from the least to the most preferred image based on its matching grayscale image, separately for pre- and post-disambiguation condition. To account for the different numbers of images across participants, we resampled these vectors to 10 bins. After averaging across voxels within each ROI, we fitted a linear regression to test whether the image-preference ranking based on gray-scale images can explain activation magnitudes to Mooney images in the pre- and post-disambiguation stage. A positive slope here would indicate that the neural code organizing voxel-tuning preference of gray-scale images generalized to pre- or post-disambiguation Mooney images. This was the case for both pre- and post-conditions in all visual ROIs (V1–V4, LOC, FG, all $p < 0.03$; $q < 0.05$, FDR-corrected, all effect sizes > 0.60; except the right V4 in the pre-disambiguation stage ($p = 0.28$); see Figs. 3B and 4B, blue and red asterisks). This result suggests that the neural representation of a Mooney image in visual areas has significant overlap with that of its matching grayscale image, regardless of whether the Mooney image is presented before or after disambiguation—and recognized or not. The significant finding for pre-disambiguation (and not-recognized) Mooney images is likely due to shared physical features between a Mooney image and its original non-degraded grayscale image, such as edges and shapes.

Next, to ascertain the neural mechanism driving the observed prior-induced suppression, we compared the steepness of pre- and post-disambiguation slopes. A steeper slope in the pre-disambiguation condition would indicate a dampening mechanism, whereby the amount of suppression scales positively with image preference (thus, neural activity coding the predicted features is primarily suppressed) (Fig. 1A). In contrast, a steeper slope in the post-disambiguation condition would indicate a sharpening mechanism, in which neural activity tuned away from (hence inconsistent with) the

JNeurosci Accepted Manuscript

predicted features is primarily suppressed (Fig. 1B). In all visual ROIs (V1–V4, LOC, FG), we found significantly larger slope coefficients in the post-disambiguation condition (all $p < 0.003$; $q < 0.05$, FDR-corrected, all effect sizes > 0.83; Figs. 3B and 4B, black asterisks), suggesting that disambiguation sharpens neural representations across the ventral visual stream (Fig. 1B). This pattern of results contrasts with a previous observation of dampening throughout the human ventral visual stream by using statistical learning to induce prior knowledge (Richter et al., 2018).

Prior-induced selective enhancement and sharpening in the FPN

The results above reveal a clear pattern of prior-induced effects in the sensory representation of Mooney images. We then tested whether a similar modulation took place in regions higher up in the cortical hierarchy: the FPN and DMN, two networks that recent studies have shown to contain content-specific activity in prior-guided visual perception (González-García et al., 2018; Flounders et al., 2019).

Results in the FPN revealed a distinct picture from the ventral visual stream. First, none of the ROIs showed significant activation in the pre-disambiguation period (all $p > 0.13$) Moreover, only the right parietal cortex showed a significant post-disambiguation activation ($p = 0.006$; $q < 0.05$, FDR-corrected, $d = 0.76$) and a significant difference in the overall BOLD magnitude between pre- and post-disambiguation Mooney images ($p = 0.026$; $q < 0.05$, FDR-corrected, $d = 0.58$), whereas no activation magnitude difference was found in the remaining FPN regions (all $p > 0.31$; Fig. 5A). The increased activation in the right parietal region after disambiguation is in contrast to prior-induced activity suppression in visual regions.

Despite a largely absent effect in the overall magnitude, image preference analysis revealed a robust pattern of results in all of the FPN regions and a potential hemispheric asymmetry (Fig. 5B). Left parietal and left frontal regions of the FPN showed a similar pattern to visual areas, with significantly positive slopes for both pre- and post-disambiguation Mooney images (all $p < 0.031$; $q < 0.05$, FDR-corrected, all effect sizes > 0.49), and significantly larger slope coefficients for post- than pre-disambiguation images (all $p < 0.001$, all effect sizes > 0.89). By contrast, in the right parietal and right frontal regions of the FPN, only post-disambiguation slopes were significant (all $p < 7.16e\text{-}5$, all effect sizes > 0.98), not pre-disambiguation slopes (all $p > 0.16$), and again with significant differences between them ($p = 0.003$, effect size = 0.77, and $p = 0.0015$, effect size = 0.83, respectively). However, no significant differences were found between pre-disambiguation slopes in left and right hemisphere ROIs (all $p > 0.12$). The same analysis repeated on FPN ROIs defined from an independent resting-state data set (Power et al., 2011) reproduced all of the results.

These results suggest selective enhancement of neural populations coding for features consistent with prior knowledge in the FPN, resulting in sharper neural representations despite largely no change in the overall activation magnitudes (Fig. 1C). As will be seen later, these results also suggest a pattern of

JNeurosci Accepted Manuscript

results in the FPN that is intermediate between the ventral visual stream and the DMN, consistent with the notion that the FPN is functionally situated between sensory regions and the DMN (Margulies et al., 2016; González-García et al., 2018).

Prior-induced enhancement and sharpening in the DMN

Within the DMN, we found significant deactivation in the pre-disambiguation period (all $p < 0.0003$; $q < 0.05$, FDR-corrected, all $d > 1.2$; blue asterisks in Fig. 6A). Interestingly, this deactivation completely disappears in the post-disambiguation period (all $p > 0.22$; Fig. 6A), resulting in higher overall magnitudes in all DMN regions in the post- compared to pre- period (all $p < 1.72e\text{-}6$; $q < 0.05$, FDR-corrected, all $d > 1.75$; black asterisk in Fig. 6A). This overall activity magnitude difference in the DMN as a result of Mooney image disambiguation is consistent with prior observations using a whole-brain approach (Dolan et al., 1997; Gorlin et al., 2012; González-García et al., 2018).

Analysis of voxel-level image preference in these regions revealed, first, no positive slope in the pre-disambiguation condition (all $p > 0.13$; Fig. 6B, blue lines), suggesting that there is no shared representation for a pre-disambiguation Mooney image and its matching grayscale image. Interestingly, slope coefficients for post-disambiguation images were significantly positive in all DMN regions (all $p < 2.15e\text{-}4$; $q < 0.05$, FDR-corrected, all effect sizes > 0.92; Fig. 6B, red asterisks) and were significantly larger than pre-disambiguation slopes (all $p < 0.013$; $q < 0.05$, FDR-corrected, all effect sizes > 0.65; Fig. 6B, black asterisks). Using DMN ROIs defined from an independent resting-state data set (Power et al., 2011) reproduced all the results.

Together, these results suggest that activity in the DMN is uninfluenced by physical image features that are shared between a Mooney image and its matching grayscale image, and that the availability of perceptually relevant prior knowledge globally enhances neural responses throughout the DMN but more so for those congruent with prior knowledge. Therefore, prior knowledge results in both enhanced neural activity and a sharper neural representation in the DMN (Fig. 1D).

Disambiguation impacts image preference rankings above and beyond the effect of repetition

Given that our paradigm relies on displaying Mooney images before and after disambiguation, an alternative explanation of our voxel-level image preference results is that they do not reflect disambiguation but rather repetition-suppression effects. In order to control for the effect of repetition, we repeated the image preference analysis using beta estimates obtained from two different sets of images. First, the same images used in the previous analyses (not-recognized pre-disambiguation, and recognized post-disambiguation), or 'Disambiguation set'. Second, images that were recognized both *before* and *after* disambiguation ('Repetition set'). For both image sets, the same subset of images was used for Pre- and Post-disambiguation conditions, and the numbers of images included were similar

JNeurosci Accepted Manuscript

between image sets ('Disambiguation set': 14.6 ± 3.5; 'Repetition set': 13.5 ± 4.7). The comparison between these two sets of images allowed us to test whether the acquisition of prior knowledge had an effect on the sharpening of neural responses above and beyond the effect of repetition.

We carried out a repeated-measures ANOVA with three factors [Image set ('Disambiguation' vs. 'Repetition'), Condition (Pre- vs. Post-disambiguation), and Network (Visual, LOC, FG, FPN, and DMN)] and the slope coefficients from the image preference analysis as the dependent variable. This analysis yielded a significant Image set × Condition interaction ($F_{1,14}$ = 7.25, p = 0.018, η^2_p = 0.34), revealing that sharpening (difference between Post and Pre slopes) was stronger in the 'Disambiguation set' compared to the 'Repetition set' (Figure 7). This interaction was not modulated by the Network factor ($F_{22.24,1.59}$ = 0.67, p = 0.49, η^2_p = 0.05). This control analysis thus confirms that although repetition has an effect on the sharpening of neural responses, the impact of disambiguation goes above and beyond this effect.

DISCUSSION

In summary, we observed a macroscopic gradient of prior knowledge's influence on perceptual processing across the cortical hierarchy. Successful disambiguation suppressed neural responses throughout the ventral visual stream by sharpening neural representations (Fig. 1B). In the FPN, neural responses consistent with prior knowledge were enhanced, while those conflicting with prior knowledge were reduced, resulting in a sharper neural representation with little change in the overall activation magnitudes (Fig. 1C). In the DMN, when relevant prior knowledge is acquired, stimulus-triggered deactivation is largely abolished, and activity enhancement is stronger for neural responses that are consistent with prior knowledge (Fig. 1D). Thus, the availability of perceptually-relevant prior knowledge induces sharper neural representations across the human cortex—from ventral visual stream to the FPN and DMN—albeit with different mechanisms: through suppression of irrelevant neural responses in visual areas and selective enhancement of relevant neural responses in higher-order frontoparietal cortices.

Previous studies have shown a consistent downregulation of activity in regions along the ventral visual stream for expected stimulus input (Murray et al., 2002; Summerfield and de Lange, 2014; de Lange et al., 2018) (but see (De Gardelle et al., 2013)). Yet, it remained unknown whether such downregulation of visual activity also occurs when priors derived from past viewing experiences guide perception of impoverished sensory input—a scenario of great relevance to natural vision, because stimuli in the natural world are often weak or ambiguous (Olshausen and Field, 2005) and past experiences play an important role in guiding future perception (Dolan et al., 1997; Hsieh et al., 2010; Gorlin et al., 2012; Van Loon et al., 2016). We addressed this question using the dramatic Mooney image effect. Our

JNeurosci Accepted Manuscript

results reveal a systematic downregulation of activity in early and ventral visual regions following prior acquisition, suggesting a similar 'expectation suppression' effect.

Two mutually exclusive mechanisms, sharpening and dampening, have been proposed to account for 'expectation suppression' and previous studies have produced a mixed pattern of results supporting both accounts (see Introduction). Here, we observed a pattern of findings consistent with the sharpening account throughout the ventral visual stream (V1–V4, LOC, FG). In all visual areas, the magnitude of activity suppression scales with image preference such that voxels whose responses are *inconsistent* with prior knowledge are most strongly suppressed. This result contrasts with previous studies reporting a dampening pattern of responses in the ventral visual stream for expected stimuli (Meyer and Olson, 2011; Kumar et al., 2017; Schwiedrzik and Freiwald, 2017; Richter et al., 2018). These previous studies investigated neural responses elicited by high-contrast, clear images, unlike the degraded Mooney images used here. Interestingly, a recent monkey study reporting 'expectation sharpening'—similar to our findings here—also adopted degraded stimuli (Bell et al., 2016). Thus, together with prior literature, our results suggest that sharpening is more likely to occur when stimulus input is weak or ambiguous—when filtering out irrelevant features and enhancing relevant ones is especially important for recognition. This interpretation of our results provides a potential way to reconcile previous disparate findings and a computational hypothesis about how priors facilitate recognition via modulating sensory responses.

Across visual areas, FPN and DMN we found that the neural representation of a post-disambiguation Mooney image has significant overlap with that of its matching grayscale image (as evidenced by significant post-disambiguation slope coefficients with images sorted according to grayscale image responses). However, only visual areas and the left-hemisphere regions of the FPN showed a significant overlap between pre-disambiguation Mooney images and their matching grayscale images. This shared neural code between pre-disambiguation Mooney images and their matching grayscale images likely reflects shared physical features between a Mooney image and its matching grayscale image such as common edges and shapes.

What might be the functional role of the FPN during prior-guided visual perception? Our study reveals a striking pattern of findings in the FPN: despite a largely absent effect in the overall activation magnitudes across FPN regions, voxels tuned towards prior knowledge are selectively enhanced, while those whose tuning preferences are inconsistent with prior knowledge are selectively inhibited. These results raise the intriguing possibility that the FPN stores perceptually relevant prior knowledge that is recruited by bottom-up input triggered by a post-disambiguation Mooney image, which in turn sends a top-down 'template' (Mumford, 1992; Summerfield et al., 2006) that induces suppression and sharpening of neural representations in the ventral visual stream. This hypothesis is consistent with a previous observation that frontal regions of the FPN rise higher up in the cortical hierarchy after

JNeurosci Accepted Manuscript

Mooney image disambiguation (González-García et al., 2018), because storage of prior knowledge is a kind of implicit memory that dissociates processing from the immediate sensory input (see more discussion below). In addition, a recent study integrating magnetoencephalography and fMRI data revealed that these regions encoded changes related to the recognition status of Mooney images before transitioning into content-specific neural processing (Flounders et al., 2019), suggesting that activation of the FPN might be the switch that brings prior knowledge into online perceptual processing. Altogether, our results suggest that the FPN might function as a source of prior-induced top-down modulations over sensory cortices by storing content-specific perceptual templates. This interpretation aligns with proposals in previous studies that used task cues to induce top-down expectation (Eger et al., 2007; Esterman and Yantis, 2010; Rahnev et al., 2011), and with the idea that the FPN sends top-down templates to guide information processing in lower-level regions (Desimone and Duncan, 1995; Buschman and Kastner, 2015). Importantly, the current results not only extend this idea to prior knowledge derived from past viewing experiences but also provide a concrete neural mechanism—selective enhancement.

In the DMN, we first replicated previous findings of increased activity magnitudes after disambiguation (Dolan et al., 1997; Gorlin et al., 2012), which results from a strong deactivation driven by pre-disambiguation Mooney images and near-baseline activity in response to the same Mooney images presented post-disambiguation. Previous studies have revealed an important role of the DMN in tasks that require making associations between relevant information, such as planning about the future or retrieving information from memories (Bar, 2007, 2009; Schacter et al., 2007). Thus, our findings suggest that DMN might represent prior-induced abstract features, such as relevant conceptual knowledge (Fairhall and Caramazza, 2013), semantic information (Binder et al., 2009) and/or internally driven memory processes (Sestieri et al., 2011; Shapira-Lichter et al., 2013; Konishi et al., 2015). According to this view, after disambiguation, the DMN is involved in generating associations to attribute meaning to the ambiguous input. Interestingly, a recent study showed that DMN deactivations encode the spatial location of visually presented stimuli (Szinte and Knapen, 2019); thus, an alternative, not mutually exclusive, possibility is that the DMN encodes stimulus-relevant prior knowledge in the sensory space.

Is the source of top-down prior knowledge guiding visual processing located in the FPN or the DMN? Our results suggest that both networks may be involved. Given that the FPN (especially its left-hemisphere components) exhibits a pattern of effects more similar to visual areas, it may encode perceptually-relevant prior knowledge more directly in the sensory space (e.g., this part of the image forms a head). In turn, the FPN may receive top-down knowledge, in more abstract terms, from the DMN, such as the conceptual knowledge of an animal. Importantly, such a large-scale hierarchy view would accommodate previous findings that have variably proposed the FPN (Eger et al., 2007;

JNeurosci Accepted Manuscript

Esterman and Yantis, 2010; Rahnev et al., 2011) or the DMN (Dolan et al., 1997; Summerfield et al., 2006) as the source of top-down priors.

Importantly, the present findings cannot be accounted for by a change in recognition status between pre- and post-disambiguation Mooney images. Successful recognition alone is widely documented to induce increased activation in visual areas and the frontoparietal network in multiple previous studies (Grill-Spector et al., 2000; Dehaene et al., 2001). By contrast, despite that post-disambiguation images are recognized while pre-disambiguation images are not, we found that post-disambiguation images are associated with reduced or unchanged activity in visual areas and the FPN, respectively. In addition, the effect probed in this study cannot be attributed to episodic memory. The Mooney image disambiguation effect relies on one-shot perceptual learning, where explicit memories with contextual associations (e.g. where and when I saw the grayscale picture of a crab) are not required. Indeed, previous work has showed that this effect still exists even when grayscale pictures are masked and not consciously recognized (Chang et al., 2016). Moreover, there is little hippocampal involvement in the Mooney image disambiguation effect (Ludmer et al., 2011; González-García et al., 2018). Therefore, this effect likely belongs to cortex-dependent perceptual learning instead of hippocampal-dependent episodic memory.

Altogether, our results reveal a large-scale cortical hierarchy of prior knowledge's influence on visual perceptual processing. Visual cortical activity is suppressed, most strongly for voxels whose responses are inconsistent with prior knowledge (Fig. 1B); by contrast, DMN activity is enhanced and most strongly for voxels whose responses align with prior knowledge (Fig. 1D). FPN exhibits a pattern of findings intermediate of visual areas and the DMN, with both selective enhancement and selective suppression (Fig. 1C). These results reinforce previous findings showing a large-scale cortical gradient with the FPN situated between sensory cortices and the DMN (Margulies et al., 2016; González-García et al., 2018). This hierarchical organization probably entails specific directionality in the interaction between different brain networks that awaits future investigation. A detailed characterization of such a cortical hierarchy will also have important clinical implications, for instance, to better understand the physiopathological processes involved in illnesses in which top-down priors might overwhelm bottom-up inputs to generate an over-reliance on prior knowledge in perception (Teufel et al., 2015; Zarkali et al., 2019).

What are the neural mechanisms underlying the sharpening effects we observed at the fMRI voxel level? Inferring tuning properties of individual neurons based on fMRI data is not straightforward, as different neural mechanisms could underlie similar fMRI responses (Alink et al., 2018); accordingly, the current study remains agnostic regarding the tuning curves of individual neurons. Our study builds on previous conceptions of sharpening at the population level, where prior knowledge produces more selective responses (de Lange et al., 2018). From this perspective, sharpening at the population level is

JNeurosci Accepted Manuscript

compatible with local neural scaling, whereby neurons tuned away from the predicted features are suppressed without a change in the width of the tuning curve (Alink et al., 2018). Whether such local scaling underlies prior-induced sharpening of population responses across the cortical hierarchy or if tuning properties of individual neurons are modulated [as, e.g., in the case of feature-based attention (Martinez-Trujillo and Treue, 2004; David et al., 2008)] deserves future investigation.

Another question for future investigation is whether our findings in the FPN and visual areas have shared mechanisms with object-based attention. Parts of the FPN, such as the intraparietal sulcus, frontal eye field, and inferior frontal junction, are involved in the top-down control of object-based attention (Baldauf and Desimone, 2014; Liu, 2016). According to the biased-competition model of selective visual attention (Desimone and Duncan, 1995; Duncan, 1998), neural representation of the attended object's features is enhanced, while that of unattended objects' features is reduced. This can result in sharper population responses as we observed in visual regions. Because in our task, object-based attention should follow object segmentation and recognition, not precede it, future studies employing techniques with high temporal resolution to record activities from these regions (such as intracranial electrocorticography) might be able to disentangle these different but inter-related processes. In addition, although the Mooney image paradigm taps mainly into prior knowledge, it does not explicitly control for object-based attention. In this regard, future development of experimental paradigms that allow a full orthogonalization of prior knowledge and attention in the context of one-shot perceptual learning would allow a more precise characterization of each process.

In conclusion, our results fill in a gap in our knowledge regarding how fast, one-shot perceptual learning influences future perceptual processing, strengthen the evidence for the sharpening account of prior knowledge's effect on visual cortical processing, and point to the FPN and DMN as potential sources of perceptually relevant prior knowledge.

REFERENCES

Albright TD (2012) On the Perception of Probable Things: Neural Substrates of Associative Memory, Imagery, and Perception. Neuron 74:227–245.

Alink A, Abdulrahman H, Henson RN (2018) Forward models demonstrate that repetition suppression is best modelled by local neural scaling. Nat Commun 9:1–10.

Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. Science 344:424-427.

Bar M (2007) The proactive brain: using analogies and associations to generate predictions. Trends Cogn Sci 11:280–289.

Bar M (2009) The proactive brain: memory for predictions. Philos Trans R Soc B Biol Sci 364:1235–

JNeurosci Accepted Manuscript

1243.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hamalainen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci 103:449–454.

Bell AH, Summerfield C, Morin EL, Malecek NJ, Ungerleider LG (2016) Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. Curr Biol 26:2280–2290.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cereb Cortex 19:2767–2796.

Buschman TJ, Kastner S (2015) From Behavior to Neural Dynamics: An Integrated Theory of Attention. Neuron 88:127-144.

Chang R, Baria AT, Flounders MW, He BJ (2016) Unconsciously elicited perceptual prior. Neurosci Conscious 2016.

David SV, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron 59:509-521.

De Gardelle V, Waszczuk M, Egner T, Summerfield C (2013) Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. Cereb Cortex 23:2235–2244.

de Lange FP, Heilbron M, Kok P (2018) How Do Expectations Shape Perception? Trends Cogn Sci 22:764–779.

Dehaene S, Naccache L, Cohen L, Bihan D Le, Mangin JF, Poline JB, Rivière D (2001) Cerebral mechanisms of word masking and unconscious repetition priming. Nat Neurosci 4:752–758.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193-222.

Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RS, Friston KJ (1997) How the brain learns to see objects and faces in an impoverished context. Nature 389:596–599.

Duncan J (1998) Converging levels of analysis in the cognitive neuroscience of visual attention. Philos Trans R Soc Lond B Biol Sci 353:1307-1317.

Eger E, Henson R, Driver J, Dolan R (2007) Mechanisms of Top-Down Facilitation in Perception of Visual Objects Studied by fMRI. Cereb Cortex 17:2123–2133.

Esterman M, Yantis S (2010) Perceptual Expectation Evokes Category-Selective Cortical Activity. Cereb Cortex 20:1245–1253.

JNeurosci Accepted Manuscript

Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. J Neurosci 33:10552–10558.

Flounders MW, González-García C, Hardstone R, He BJ (2019) Neural dynamics of visual ambiguity resolution by perceptual prior. Elife 8.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc B Biol Sci 360:815–836.

Glover GH, Li T-Q, Ress D (2000) Image-Based Method for Retrospective Correction of Physiological Motion Effects in fMRI: RETROICOR. Magn Reson Med 167:162–167.

González-García C, Flounders MW, Chang R, Baria AT, He BJ (2018) Content-specific activity in frontoparietal and default-mode networks during prior-guided visual perception. Elife 7.

Gorlin S, Meng M, Sharma J, Sugihara H, Sur M, Sinha P (2012) Imaging prior information in the brain. Proc Natl Acad Sci 109:7935–7940.

Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. Nat Neurosci 3:837–893.

Hsieh PJ, Vul E, Kanwisher N (2010) Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. J Neurophysiol 103:1501–1507.

Kok P, Jehee JFM, de Lange FP (2012) Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron 75:265–270.

Konishi M, McLaren DG, Engen H, Smallwood J (2015) Shaped by the Past: The Default Mode Network Supports Cognition that Is Independent of Immediate Perceptual Input Hayasaka S, ed. PLoS One 10:e0132209.

Kumar S, Kaposvari P, Vogels R (2017) Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons. J Cogn Neurosci 29:1445–1454.

Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A 20:1434.

Liu T (2016) Neural representation of object-specific attentional priority. Neuroimage 129:15-24.

Ludmer R, Dudai Y, Rubin N (2011) Uncovering Camouflage: Amygdala Activation Predicts Long-Term Memory of Induced Perceptual Insight. Neuron 69:1002–1014.

Margulies DS, Ghosh SS, Goulas A, Falkiewicz M, Huntenburg JM, Langs G, Bezgin G, Eickhoff SB, Castellanos FX, Petrides M, Jefferies E, Smallwood J (2016) Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc Natl Acad Sci 113:12574–

JNeurosci Accepted Manuscript

12579.

Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr Biol 14:744-751.

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci U S A 108:19401–19406.

Morey RD (2008) Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutor Quant Methods Psychol.

Mumford D (1992) On the computational architecture of the neocortex - II The role of cortico-cortical loops. Biol Cybern 66:241–251.

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci U S A 99:15164–15169.

Murray SO, Schrater P, Kersten D (2004) Perceptual grouping and the interactions between visual cortical areas. Neural Networks 17:695–705.

Olshausen BA, Field DJ (2005) How close are we to understanding V1? Neural Comput 17:1665–1699.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE (2011) Functional Network Organization of the Human Brain. Neuron 72:665–678.

Rahnev D, Lau H, de Lange FP (2011) Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. J Neurosci 31:10741–10748.

Ramachandran VS (1988) Perception of shape from shading. Nature 331:163–166.

Richter D, Ekman M, de Lange FP (2018) Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. J Neurosci 38:7452–7461.

Schacter, Addis DR, Buckner RL (2007) Remembering the past to imagine the future: The prospective brain. Nat Rev Neurosci 8:657–661.

Schwiedrzik CM, Freiwald WA (2017) High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. Neuron 96:89-97.e4.

Sestieri C, Corbetta M, Romani GL, Shulman GL (2011) Episodic memory retrieval, parietal cortex, and the default mode network: Functional and topographic analyses. J Neurosci 31:4407–4420.

Shapira-Lichter I, Oren N, Jacob Y, Gruberger M, Hendler T (2013) Portraying the unique contribution of the default mode network to internally driven mnemonic processes. Proc Natl Acad Sci U S A

JNeurosci Accepted Manuscript

110:4950–4955.

Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci.

Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006) Predictive codes for forthcoming perception in the frontal cortex. Science (80-) 314:1311–1314.

Szinte M, Knapen T (2019) Visual Organization of the Default Network. Cereb Cortex 30:3518–3527.

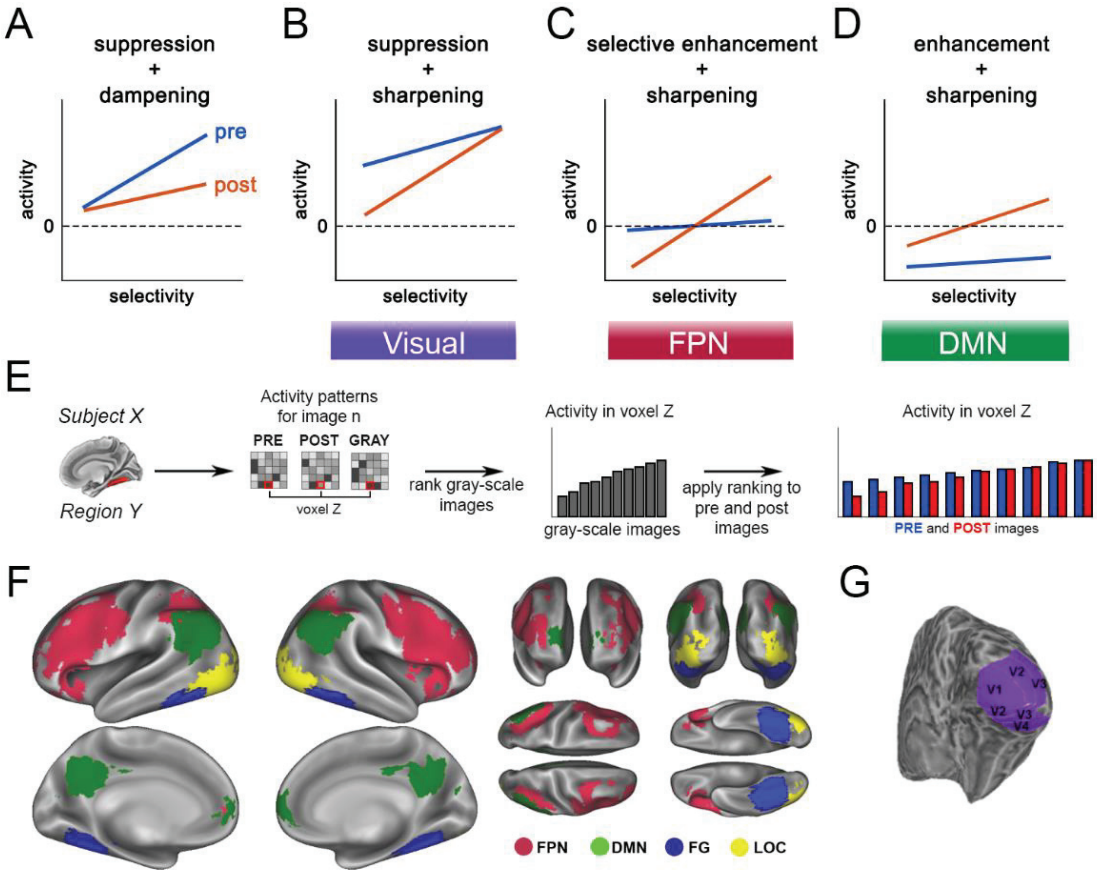
Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, Goodyer IM, Fletcher PC (2015) Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci 112:13401–13406.

Van Loon AM, Fahrenfort JJ, Van Der Velde B, Lirk PB, Vulink NCC, Hollmann MW, Steven Scholte H, Lamme VAF (2016) NMDA Receptor Antagonist Ketamine Distorts Object Recognition by Reducing Feedback to Early Visual Cortex. Cereb Cortex 26:1986–1996.

Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? Trends Cogn Sci 10:301–308.

Zarkali A, Adams RA, Psarras S, Leyland L-A, Rees G, Weil RS (2019) Increased weighting on prior knowledge in Lewy body-associated visual hallucinations. Brain Commun 1.

JNeurosci Accepted Manuscript



677

678 **Figure 1.** (**A-D**) Alternative explanatory mechanisms of prior-induced changes in neural activity. (**E**)
679 Schematic for the image preference analysis. For each voxel within an ROI, images are ranked based
680 on the activation magnitudes in the gray-scale condition. This image ranking is then applied to the
681 same voxel's activity during pre- and post-disambiguation conditions. For details, see Methods, *Image*
682 *preference analysis*. **(F)** Locations of FPN and DMN ROIs (red and green), and category-selective
683 visual areas (FG and LOC, blue and yellow). (**G**) Location of early visual areas in an example subject.

JNeurosci Accepted Manuscript

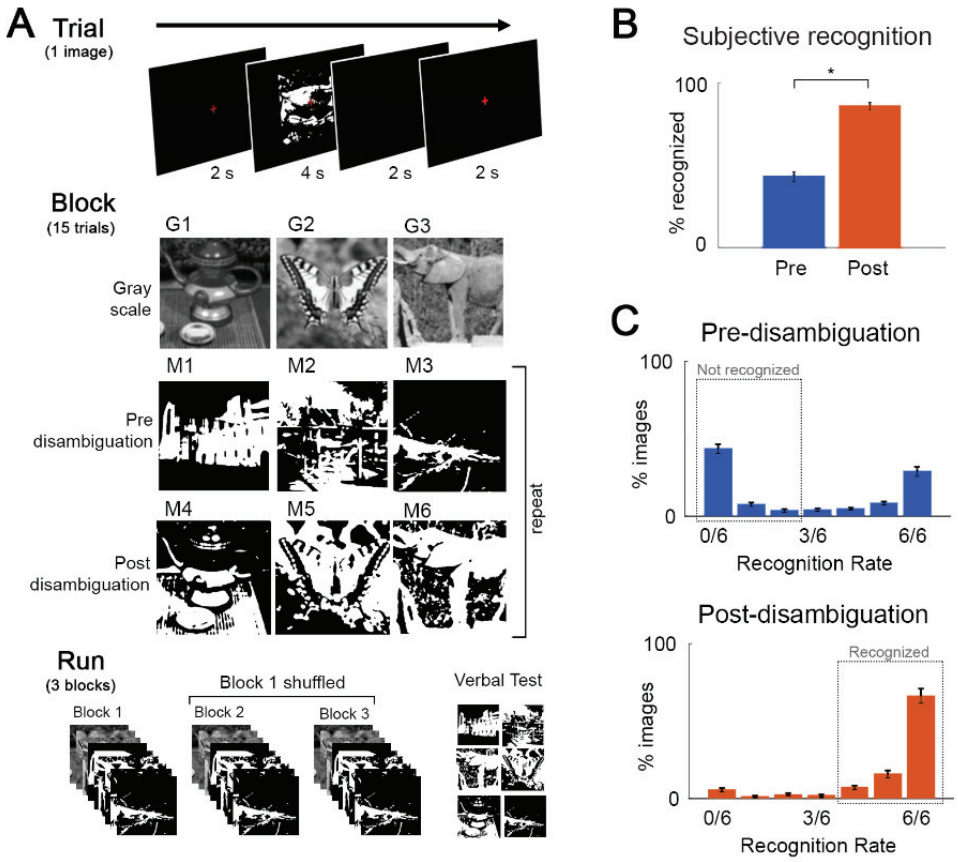


Figure 2. Paradigm and behavioral results. (**A**) Task design, and flow of events at trial, block and fMRI run level. Participants viewed gray-scale and Mooney images and were instructed to respond to the question 'Can you recognize and name the object in the image?'. 33 unique images were used and each was presented six times before and six times after disambiguation. (**B**) Percentage of 'recognized' answers across all Mooney image presentations before and after disambiguation, which differed significantly (p=3.4e-13). (**C**) Distribution of recognition rates across 33 Mooney images pre- and post-disambiguation. Dashed boxes depict the cut-offs used to classify an image as recognized or not-recognized. Error bars denote s.e.m. across subjects.

JNeurosci Accepted Manuscript

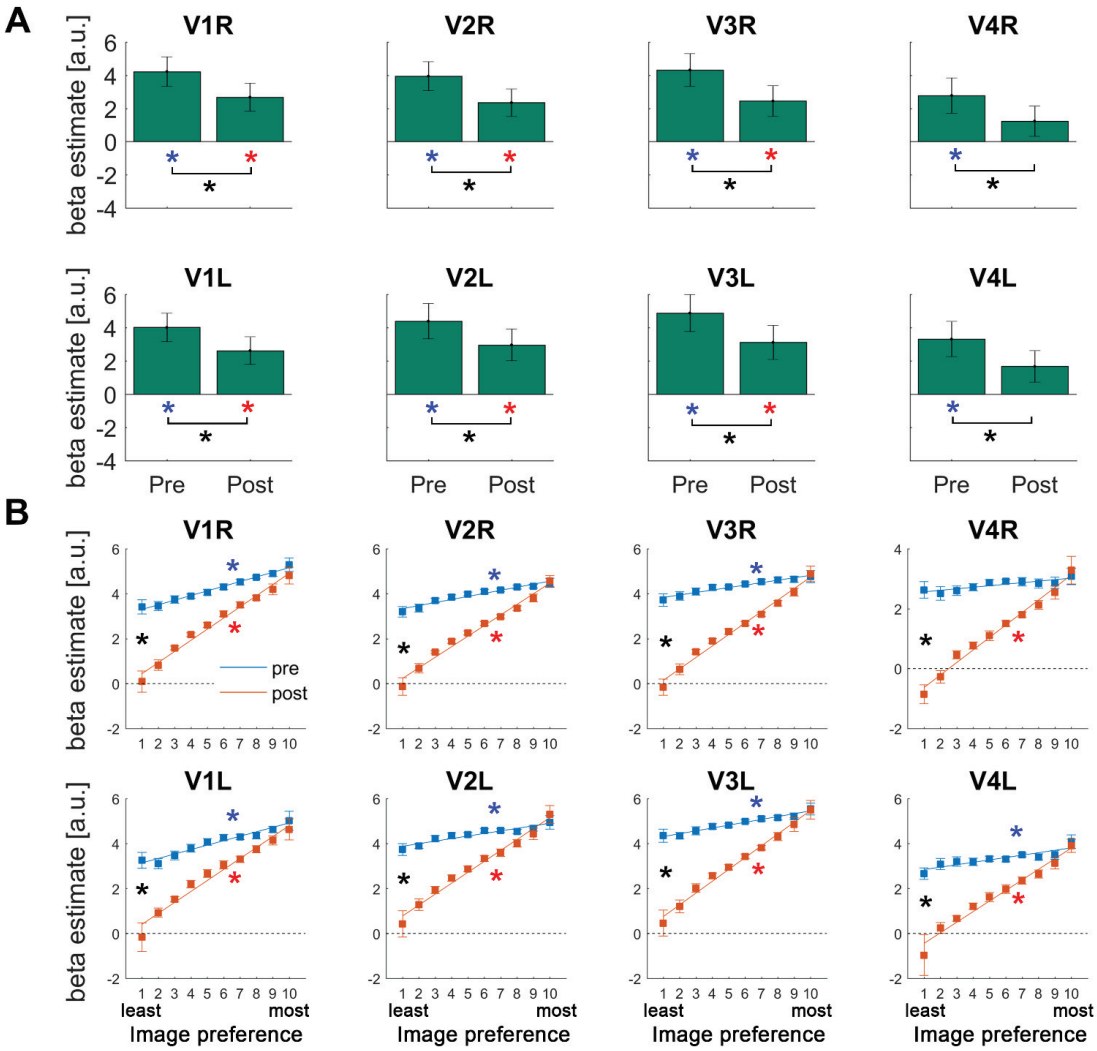


Figure 3. **(A)** Activation magnitudes in early visual regions. Each bar represents the across-subject mean beta value in response to Mooney images presented before or after disambiguation. Blue (/red) asterisks denote significant pre- (/post-) disambiguation activation above baseline, and black asterisks denote significant differences between conditions ($p < 0.05$, paired t-test, FDR-corrected). Error bars denote s.e.m. across subjects. (**B**) Results of the image preference analysis in early visual regions. For each voxel, Mooney images were sorted based on that voxel's activation magnitudes in response to their matching grayscale images, such that images with a high ranking contain predicted features that maximally activate this voxel, whereas images with a low ranking contain predicted features that do not activate this particular voxel. The activation magnitudes (in response to pre- or post-disambiguation Mooney images) were averaged across voxels for each position in this image preference ranking. A

JNeurosci Accepted Manuscript

linear regression was then fit to each subject's data for each ROI, and the regression coefficients (the slope) was subjected to a right-tailed Wilcoxon sign-rank test against 0 for each condition (pre- and post-disambiguation) and to a two-tailed Wilcoxon sign-rank test between conditions. Lines reflect the across-subject mean regression line fit to the pre- (blue) and post- (red) disambiguation images, respectively. Error bars represent s.e.m. corresponding to the paired test (Morey, 2008). Blue and red asterisks denote significant positive slopes for the pre- and post-disambiguation data, respectively. Black asterisks denote a significant difference in slope coefficients between pre- and post-disambiguation conditions. All presented results in Figs. 3-6 are FDR-corrected across all 20 ROIs.

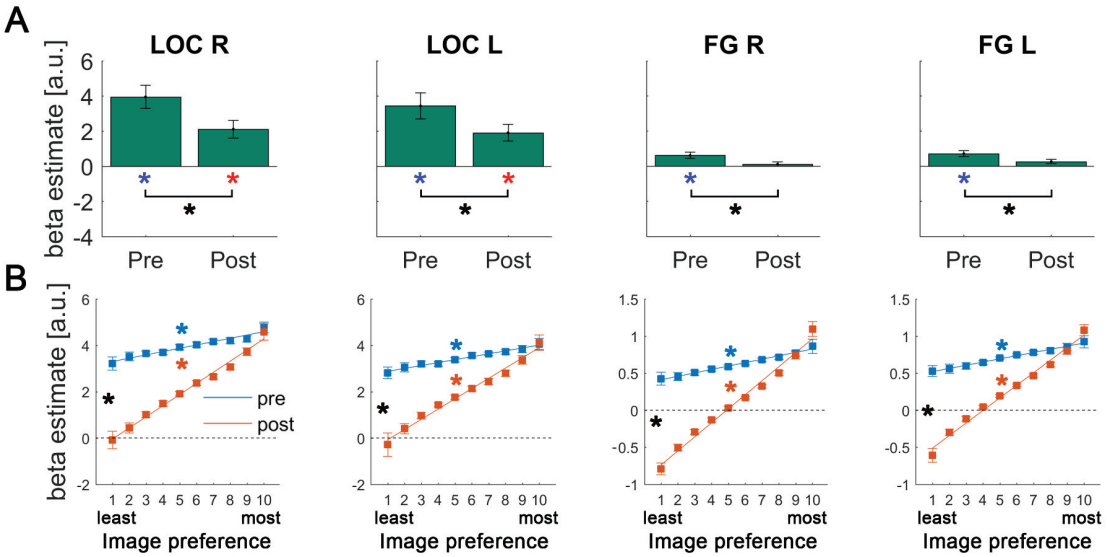
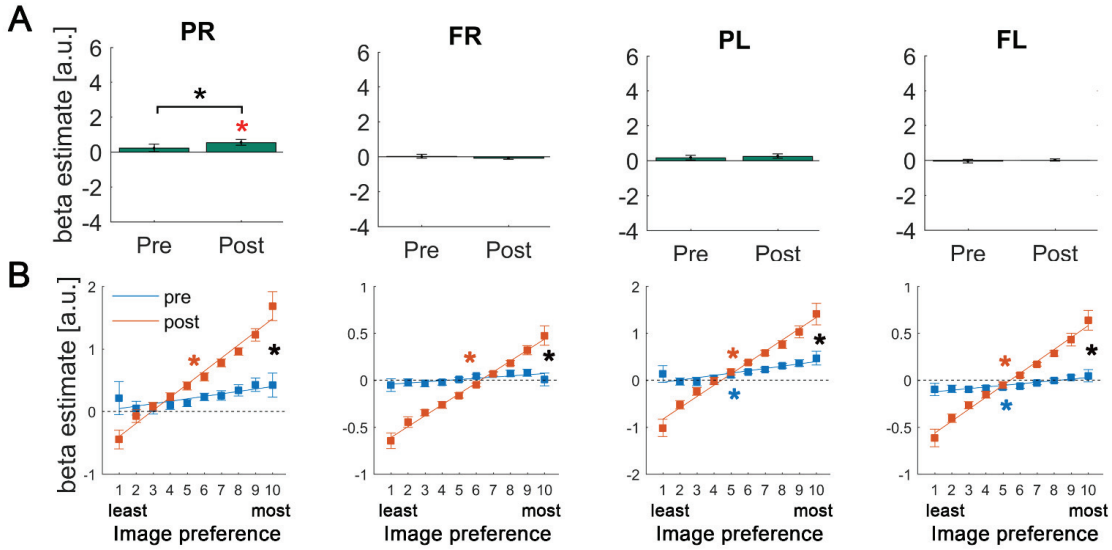


Figure 4. Results of the univariate GLM (**A**) and image preference analysis (**B**) in the LOC and FG. Format is the same as in Figure 3.

JNeurosci Accepted Manuscript

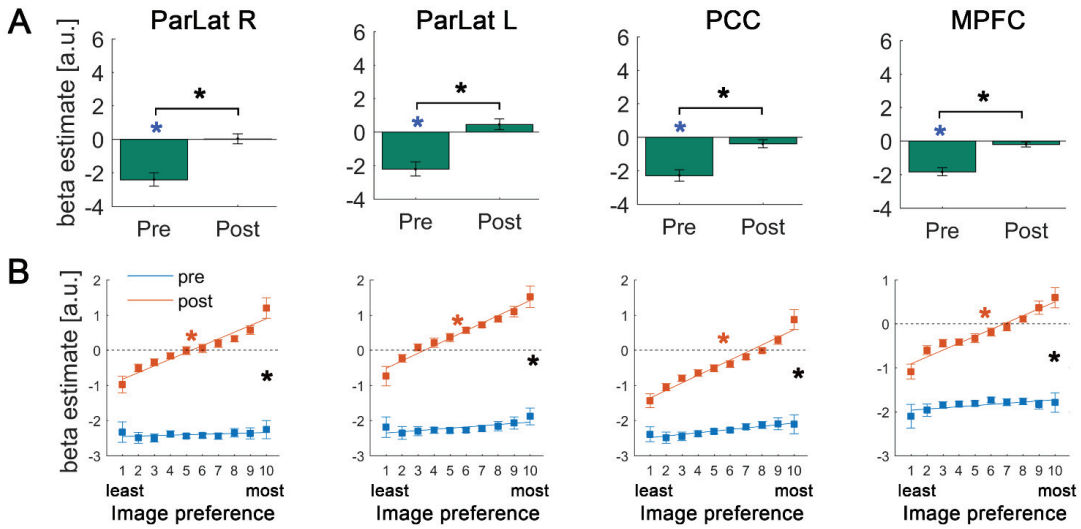


717

718 **Figure 5.** Results of the univariate GLM (**A**) and image preference analysis (**B**) in the FPN regions.

719 Format is the same as in Figure 4.

720



721

722 **Figure 6.** Results of the univariate GLM (**A**) and image preference analysis (**B**) in the DMN regions.

723 Format is the same as in Figure 4.

JNeurosci Accepted Manuscript

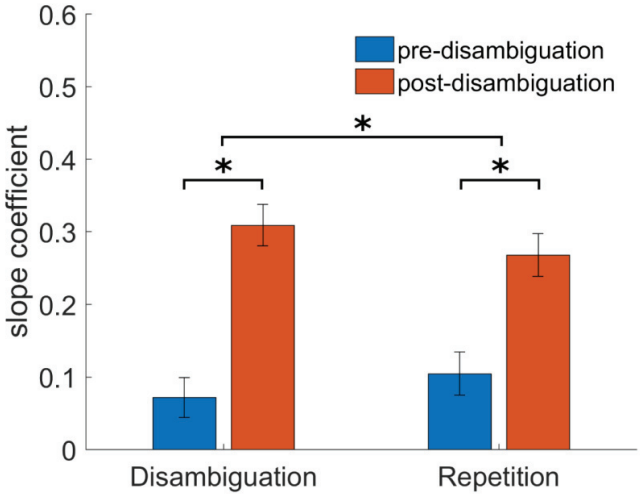
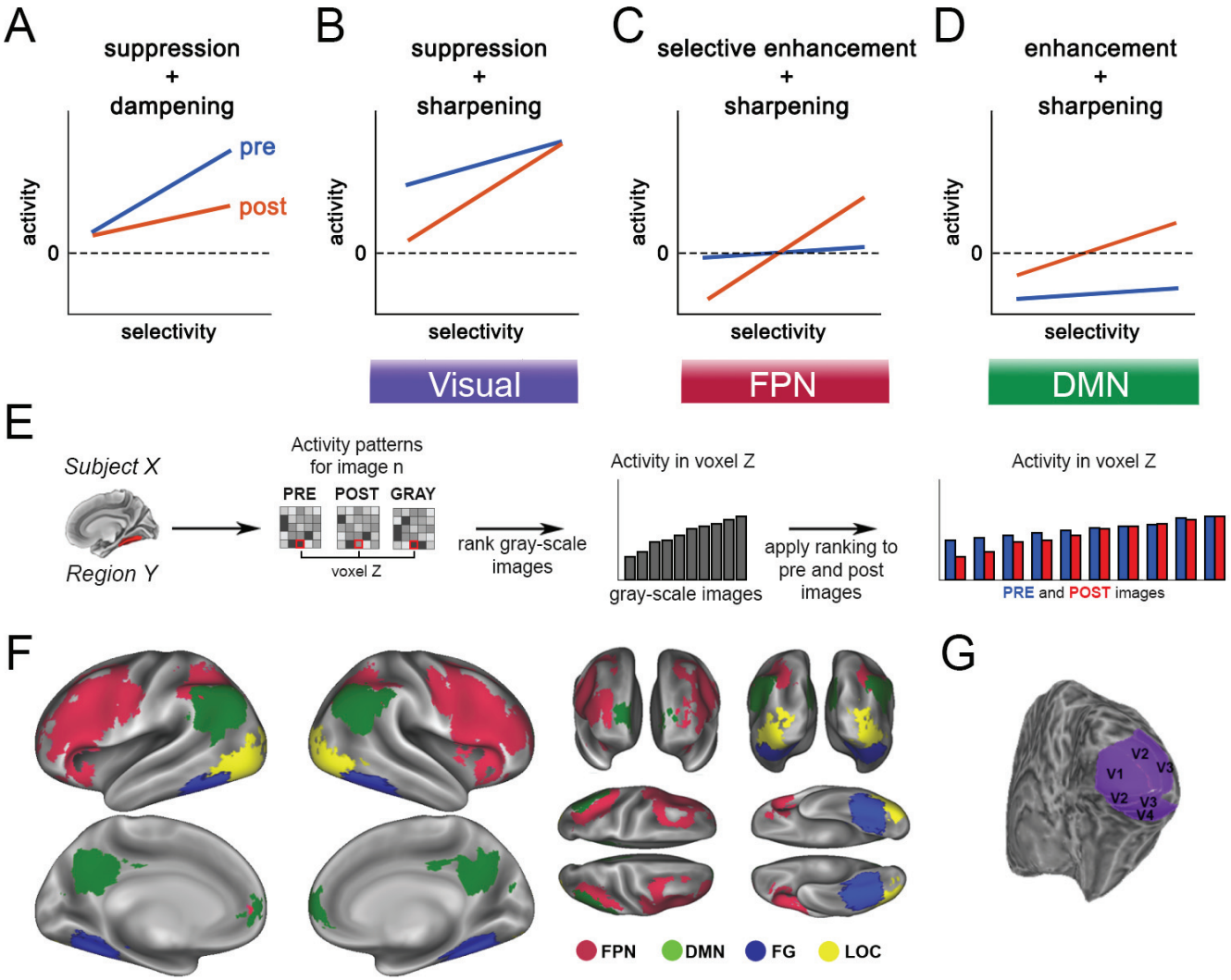
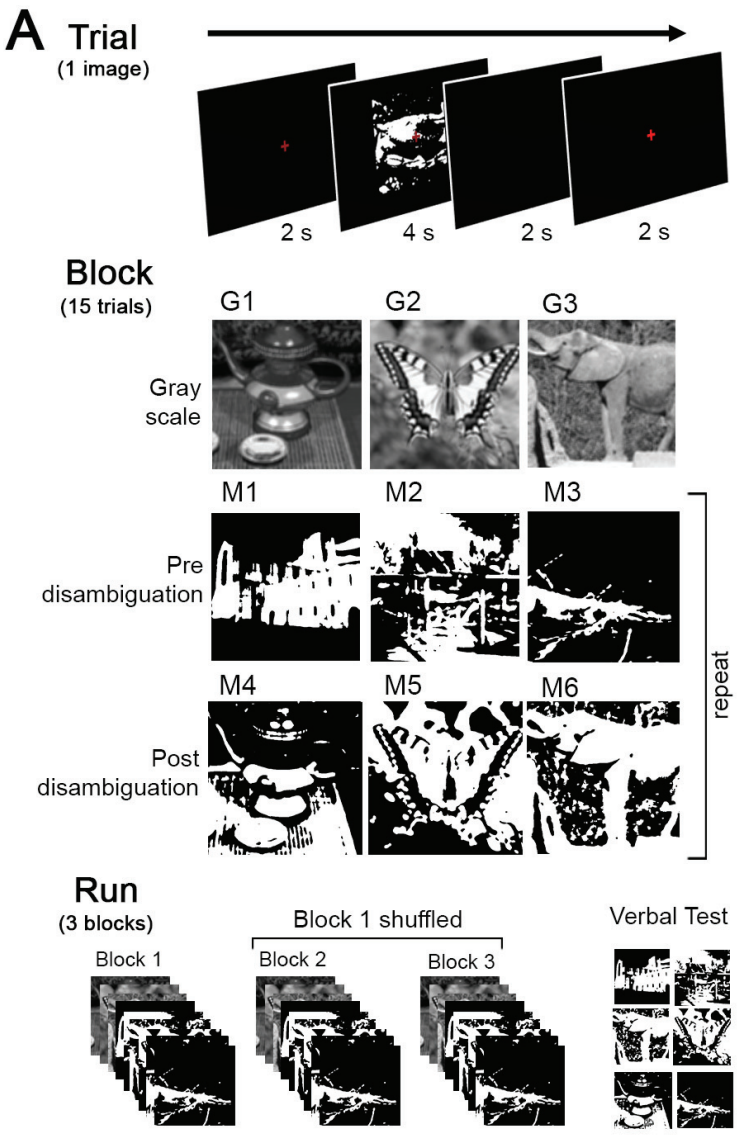


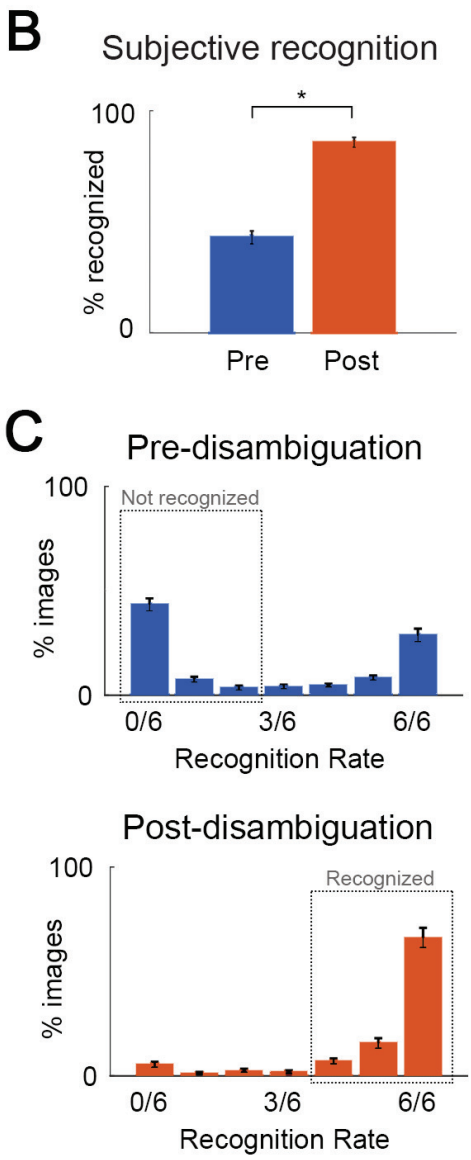
Figure 7. Control analysis for the effect of recognition in our image preference results. Bars depict the across-network mean slope coefficient pre- and post-disambiguation for two different sets of images that underwent an identical sequence of repetition, but different sequence recognition: The Disambiguation set includes images that were not-recognized before and recognized after disambiguation, while the Repetition set includes images that were recognized both before and after disambiguation. The upper star denotes a significant interaction of Image Set × Condition (Pre vs. Post) ($p < 0.02$), with significant differences between pre- and post-disambiguation in both sets of images (lower stars), but stronger sharpening in the Disambiguation set. Error bars depict s.e.m.

JNeurosci Accepted Manuscript

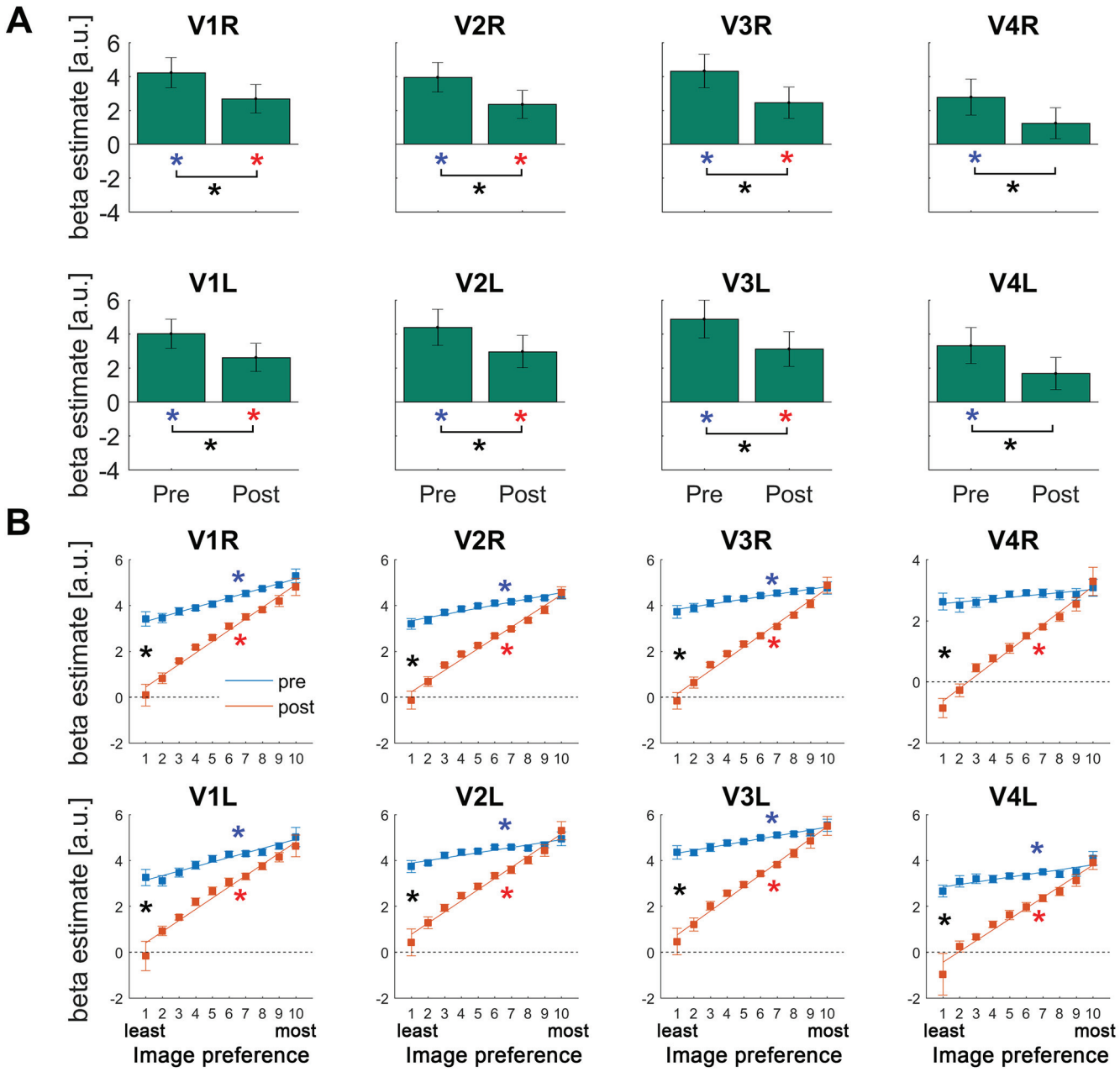


JNeurosci Accepted Manuscript

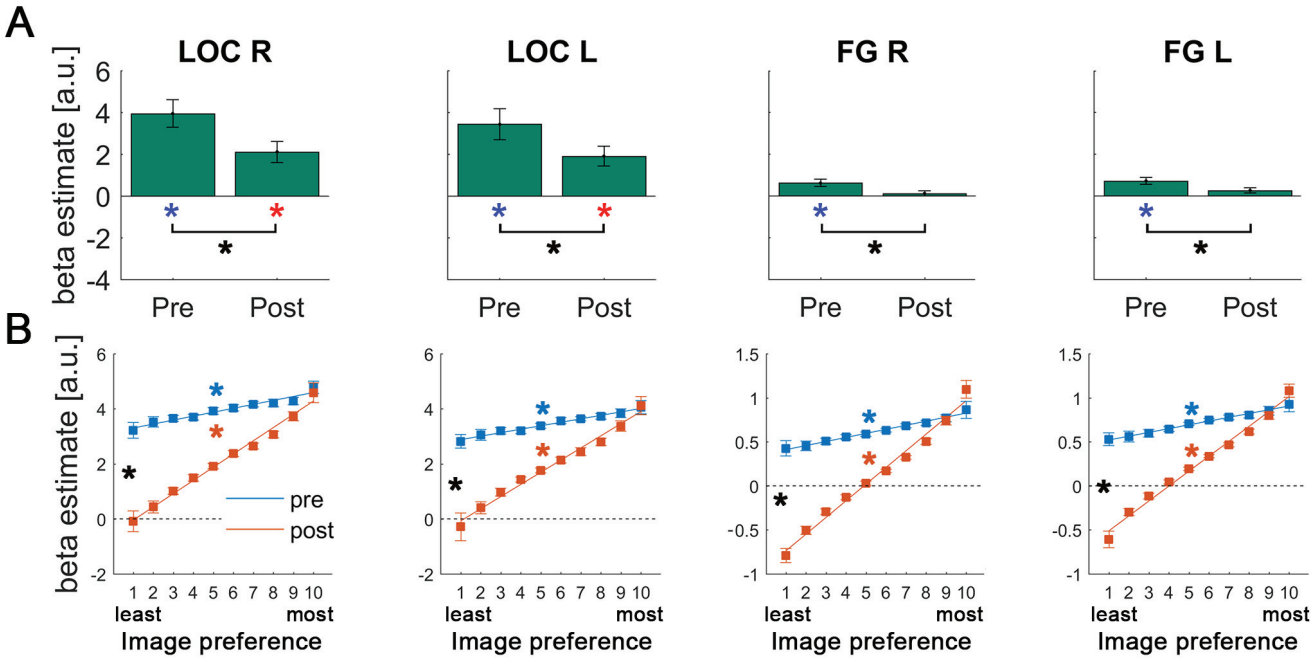




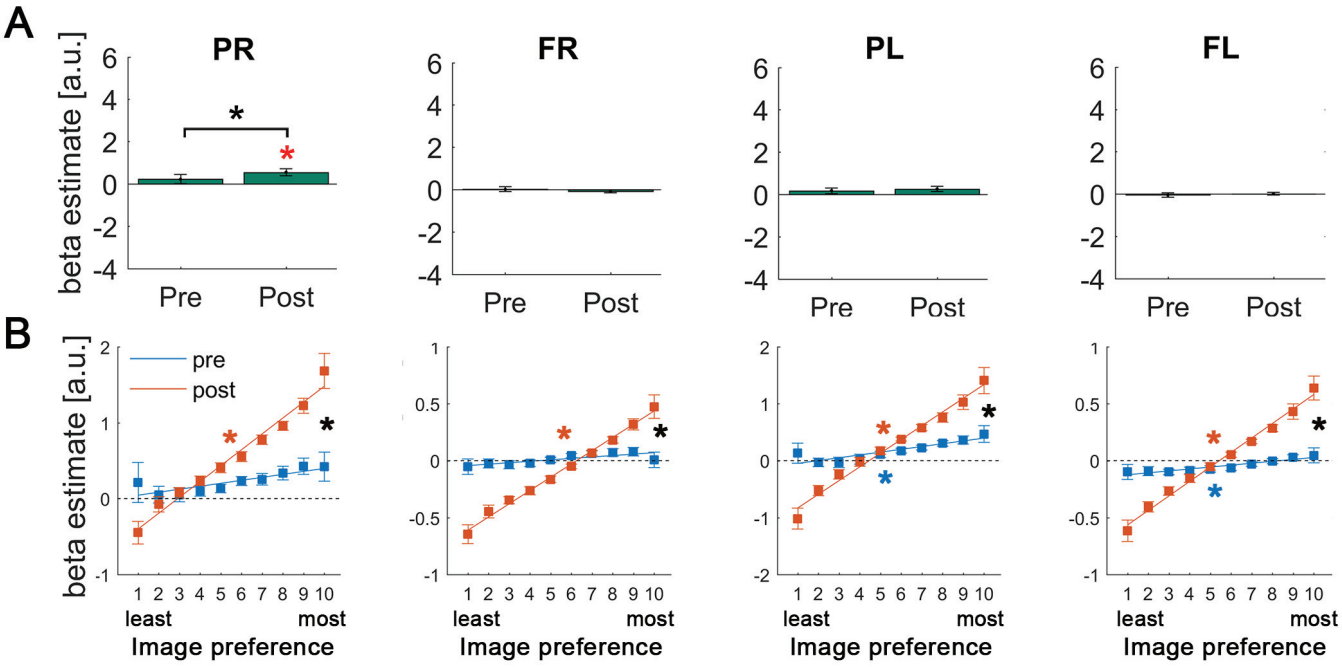
JNeurosci Accepted Manuscript



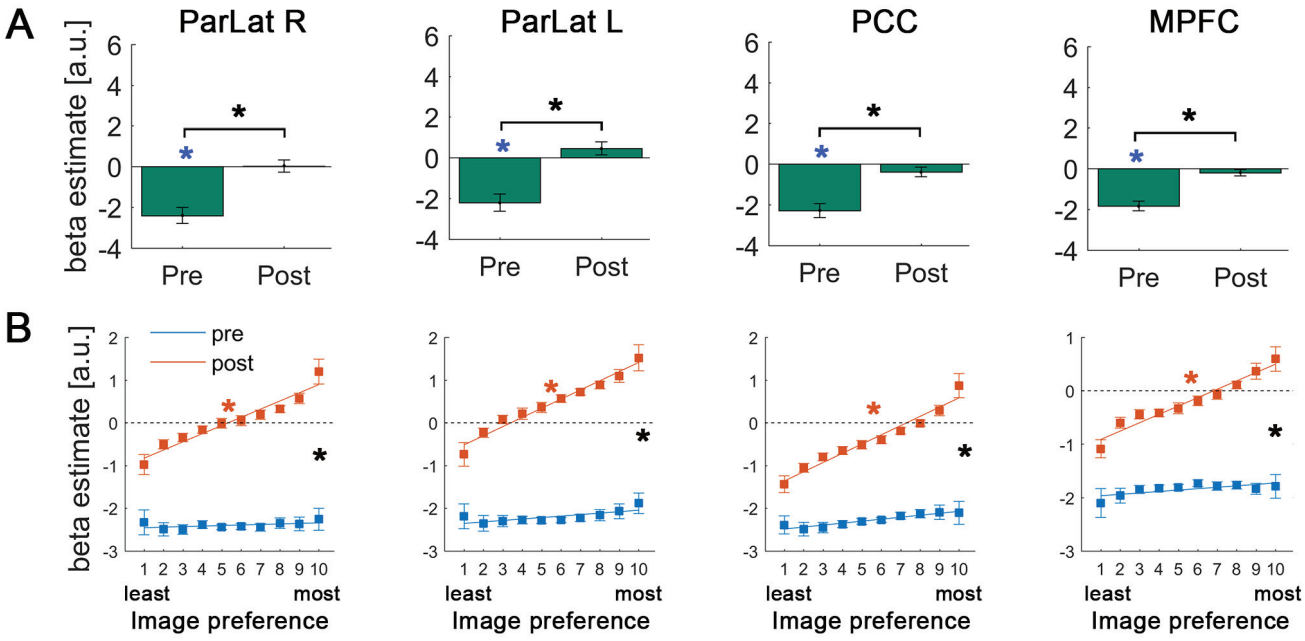
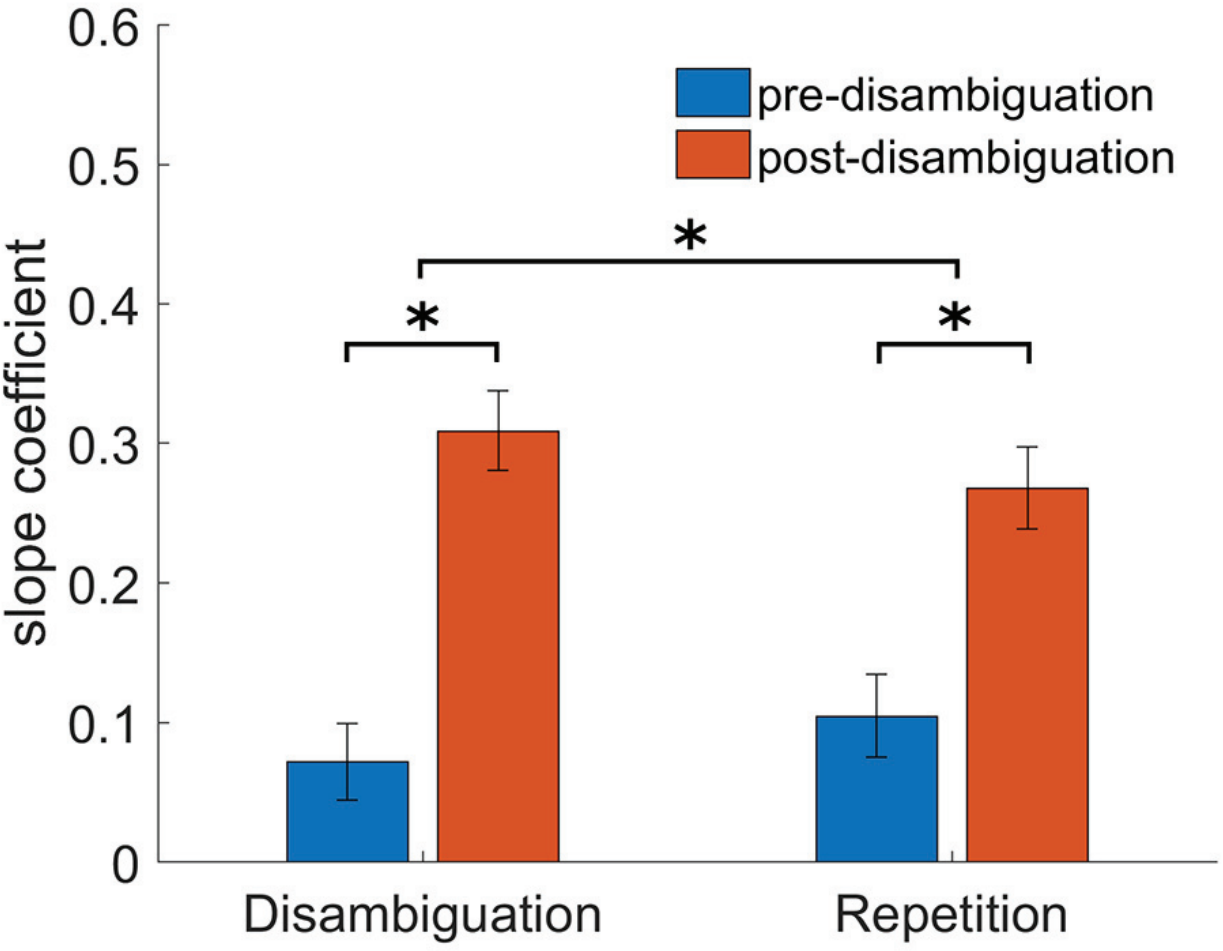
JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript