Brain mechanisms for simple perception and bistable perception

Megan Wang[1], Daniel Arteaga[1,2], and Biyu J. He[3]

National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892

Edited by Marcus E. Raichle, Washington University in St. Louis, St. Louis, MO, and approved July 15, 2013 (received for review December 16, 2012)

When faced with ambiguous sensory inputs, subjective perception alternates between the different interpretations in a stochastic manner. Such multistable perception phenomena have intrigued scientists and laymen alike for over a century. Despite rigorous investigations, the underlying mechanisms of multistable perception remain elusive. Recent studies using multivariate pattern analysis revealed that activity patterns in posterior visual areas correlate with fluctuating percepts. However, increasing evidence suggests that vision—and perception at large—is an active inferential process involving hierarchical brain systems. We applied searchlight multivariate pattern analysis to functional magnetic resonance imaging signals across the human brain to decode perceptual content during bistable perception and simple unambiguous perception. Although perceptually reflective activity patterns during simple perception localized predominantly to posterior visual regions, bistable perception involved additionally many higher-order frontoparietal and temporal regions. Moreover, compared with simple perception, both top-down and bottom-up influences were dramatically enhanced during bistable perception. We further studied the intermittent presentation of ambiguous images—a condition that is known to elicit perceptual memory. Compared with continuous presentation, intermittent presentation recruited even more higher-order regions and was accompanied by further strengthened top-down influences but relatively weakened bottom-up influences. Taken together, these results strongly support an active top-down inferential process in perception.

visual perception | fMRI | MVPA | Granger causality | ambiguous images

The problem of vision entails the constant interpretation of inherently ambiguous local components of a complex scene. In contrast to reduced visual stimuli routinely used in laboratory research such as Gabor patches and isolated faces, natural scenes contain many ambiguities caused by clutter, occlusion, shading, and the inherent complexity of natural objects (1, 2). Similarly, simple daily tasks, such as interpreting the handwriting of another individual, require a level of cognitive capability surmounting that of modern-day computers. The ease with which we are able to rapidly perform such tasks attests to the remarkable capacity of the human visual system, or alternatively, to the vast knowledge and templates stored in the human brain aiding in visual perception (3).

Ambiguous images such as the Necker cube and Rubin face-vase illusion provide a well-controlled experimental approach to studying the brain's processing when it is faced with ambiguities in sensory inputs. When multiple interpretations of the same sensory inputs are possible, subjective perception alternates between the different interpretations in a stochastic manner (for reviews, see refs. 2 and 4–6). In the case of ambiguous images containing two possible interpretations, this phenomenon is referred to as "bistable perception."

Neuroscientific studies of bistable perception over the past several decades have significantly advanced our understanding of this phenomenon. The literature has largely converged on several findings. Firstly, frontal and parietal brain regions seem to be involved in perceptual switching, as demonstrated by neuroimaging (7–11) (but see ref. 12) as well as transcranial magnetic stimulation (TMS) (13–15) and lesion (16–18) studies. Secondly, functional MRI (fMRI) activity patterns in visual regions, including the primary visual cortex (V1) and lateral geniculate nucleus (LGN), correlate with the content of fluctuating percepts, as reflected in both activity fluctuations of an entire brain region (19–24) and the fine spatial patterns of activity within a region (25–27). Thirdly, a progressively larger fraction of neurons show percept-modulated firing rate changes as one moves up the visual hierarchy, from ∼20% of neurons in V1 to ∼90% in the inferior temporal (IT) cortex (4, 28).

To date, studies decoding the content of fluctuating percepts in bistable perception have focused on the visual cortex. In light of the recent emerging framework that vision is not a bottom-up process with sensory inputs passively mapped across different levels of the brain, but rather an active inferential process with top-down processes actively guiding and shaping visual perception (1, 3, 29–34), it would be of great value to know how percept-related activity is distributed across the brain. (In using terms "top-down" and "bottom-up," we are under the assumption that cognitive and neural processes cannot be dissociated.) Supporting this idea, a recent primate study using a binocular flash suppression paradigm showed that a majority of visually responsive neurons in lateral prefrontal cortex correlate with perceptual experience in their firing rates (35). Nonetheless, the distribution of such percept-reflective activity patterns across the brain remains unclear.

Ambiguous images also lend themselves well to the study of perceptual memory (36, 37). Intermittent removal of ambiguous

Significance

When viewing an image with multiple interpretations such as the Necker cube, subjective perception alternates stochastically between the different interpretations. This phenomenon provides a well-controlled experimental approach to studying how the brain responds to ambiguities in sensory inputs—a ubiquitous problem in dealing with natural environment. We found that, compared with simple perception devoid of ambiguities, bistable perception requires additional higher-order brain regions and dramatically enhanced top-down and bottom-up influences in the brain. Intermittent viewing of ambiguous images elicits even stronger top-down brain activity. These results help elucidate the mechanisms of visual perception by demonstrating an active top-down inferential process.

Author contributions: B.J.H. designed research; M.W., D.A., and B.J.H. performed research; M.W., D.A., and B.J.H. contributed new reagents/analytic tools; M.W. and D.A. analyzed data; and M.W. and B.J.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Xnat Central database (https://central.xnat.org/).

[1]M.W. and D.A. contributed equally to this work.

[2]Present address: Vanderbilt University School of Medicine, Nashville, TN 37232.

[3]To whom correspondence should be addressed. E-mail: biyu.he@nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221945110/-/DCSupplemental.

images from view for >200 ms at a time slows down the rate of perceptual switching, even to the extent of bringing it to a standstill (38–40). This phenomenon suggests the existence of a perceptual memory trace during the intervening blank periods, such that the most prevalent percept in the recent past is likely to be reinstated when the image reappears. At present, the respective contributions of early visual, extrastriate visual and higher-order association regions to perceptual memory remain actively debated (40–44).

In this study, we asked the following questions: (*i*) Which brain regions, in their activity patterns, carry information about fluctuating perceptual content during bistable perception? (*ii*) Are similar brain regions involved in the simple perception of unambiguous visual images as compared with bistable perception as well as intermittent bistable perception? (*iii*) How do these brain regions interact with each other in the different perceptual conditions? To address the first two questions, we performed searchlight multivariate pattern analysis (MVPA) across the entire human brain to decode (*i*) perception of different unambiguous images, (*ii*) the content of fluctuating percepts elicited by ambiguous images, and (*iii*) the content of perception and perceptual memory during intermittent presentations of ambiguous images. To answer the last question, we contrasted directed influences between widespread brain regions across these three conditions using Granger causality (GC) analysis.

Results

Behavioral Results. To find brain mechanisms that are generalizable across different bistable images, we studied the well-known Necker cube and Rubin face-vase illusion (Fig. 1*A*, *Lower*) under identical task design and analyses. Thirteen healthy subjects participated in the experiment. Each subject completed the task under three conditions (Fig. 1*B*). First, in the unambiguous (UnAmb) condition, subjects continuously viewed Necker Cube and Rubin face-vase images that had been modified to accentuate one or the other percept (Fig. 1*A*, *Upper*). They pressed one of two buttons indicating their percept at each image presentation, which lasted 16 s. Subjects' percepts stably matched the intended disambiguation on 95.7 ± 4.9% (mean ± SD across subjects) of all Necker-cube trials, and 82.7 ± 18.8% of all Rubin face-vase trials. Second, in the ambiguous (Amb) condition, the original ambiguous images (Fig. 1*A*, *Lower*) were presented for 60 s at a time, and subjects indicated every spontaneous perceptual switch that they experienced throughout the duration of image presentation. Third, in the discontinuous (Disc) condition, each ambiguous image (Fig. 1*A*, *Lower*) was presented for 2 s followed by a 6-s blank period, and this sequence was repeated for nine times in each block. Subjects indicated their dominant percept in response to every image presentation. The mapping between percepts and buttons was identical across the three conditions. The distributions of percept durations in the Amb and Disc conditions for all subjects are shown in Fig. 1*C*. In the

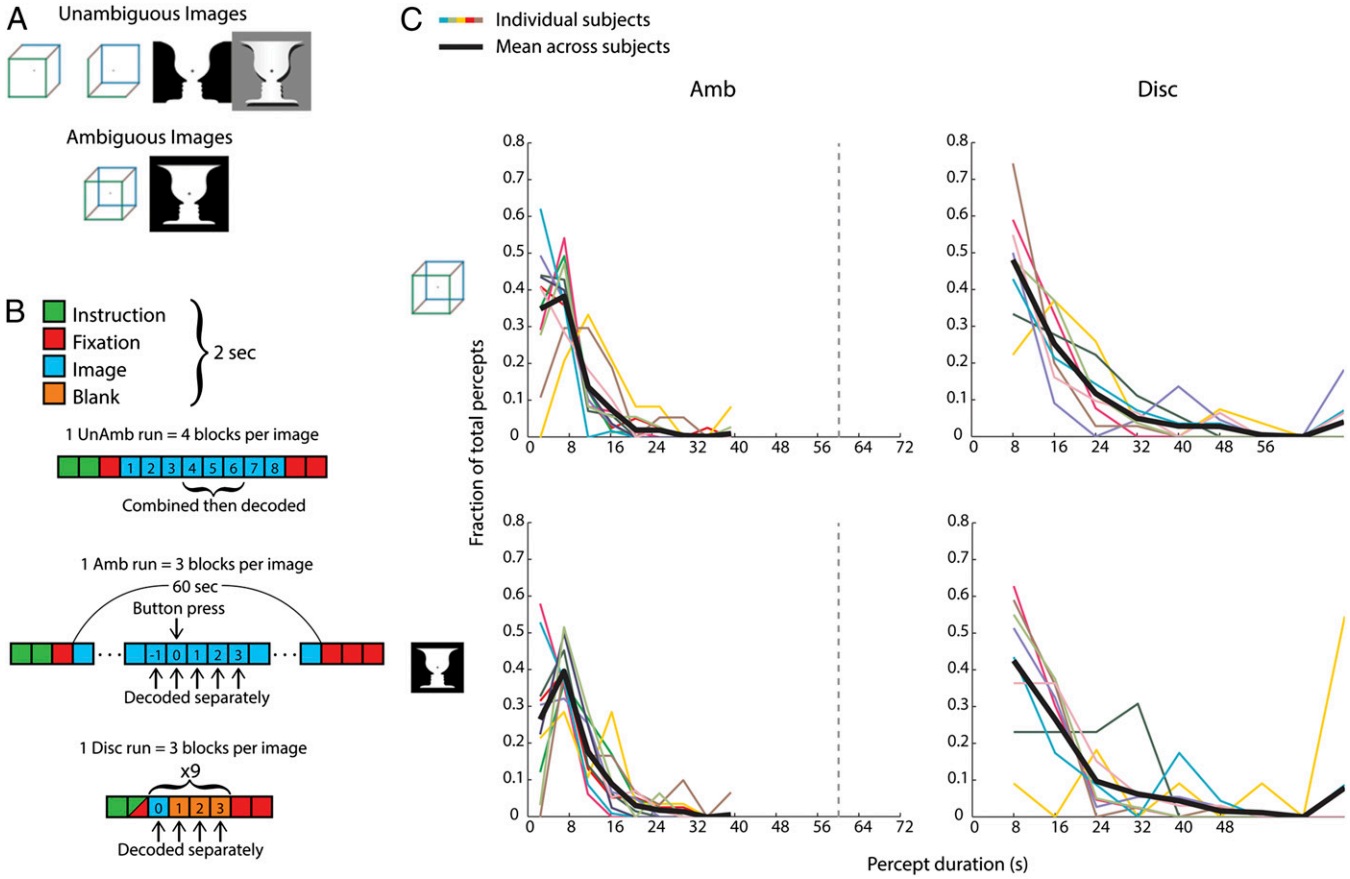

Fig. 1. Experimental paradigm and behavioral results. (*A*) Altered Necker cube and Rubin face-vase images (*Upper*) presented in the unambiguous (UnAmb) condition and the original images (*Lower*) presented in the ambiguous (Amb) and discontinuous (Disc) conditions. (*B*) Experimental design for each condition. Each UnAmb run contained 16 blocks (4 blocks per image), and each Amb and Disc run contained 6 blocks (3 blocks per image). The fMRI frames used for decoding are indicated in the graph. (*C*) Frequency histograms showing the distribution of percept durations in each subject (n = 11 for Amb; n = 8 for Disc). Percept durations were sorted into nine bins for both Amb and Disc data (combined across both percepts of each ambiguous image). Thick black lines indicate the mean across subjects.

Disc condition, the probability that percepts during consecutive image presentations were the same is significantly higher than their being different {P[report (N) = report ($N-1$)] > P[report (N) ≠ report ($N-1$)], $P = 0.026$, Wilcoxon signed-rank test}, indicating the presence of perceptual memory (38, 39).

Brain Activity Patterns Underlying Simple Perception and Bistable Perception. First, we investigated brain activity patterns underlying simple perception of unambiguous images. Searchlight-based MVPA (45) was applied across the whole brain on the peak of the fMRI response in UnAmb trials (6~12 s after image onset) to decode the percepts. Group analysis was conducted by a one-sample t test on decoding accuracy against chance level (0.5) at every voxel across subjects, and the results were thresholded at $P < 0.05$ after correcting for multiple comparisons (for details, see *SI Methods*). Brain regions containing percept-reflective activity in the UnAmb condition localized mainly to posterior visual areas (Fig. 2 and Fig. S1, *Top*).

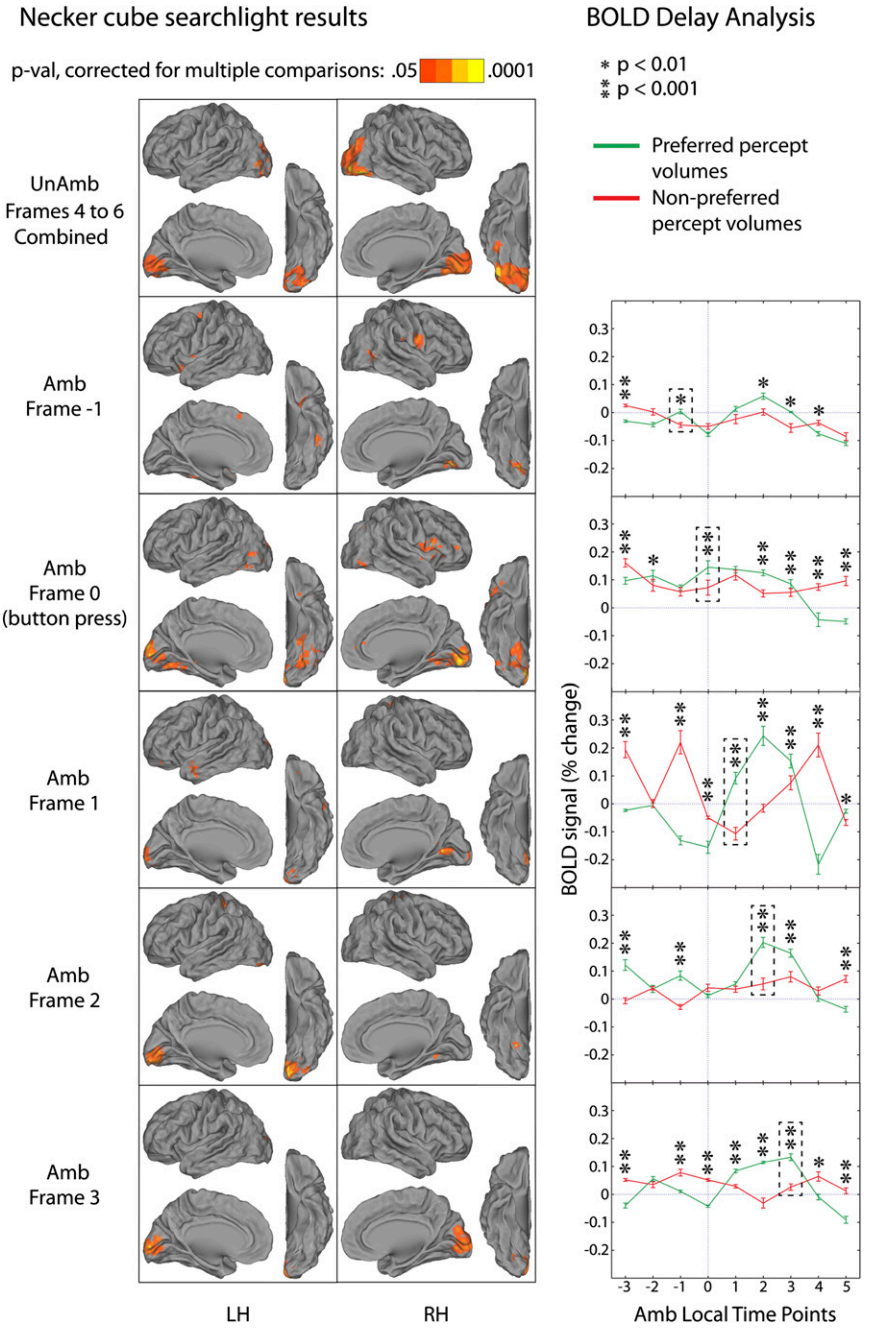

Fig. 2. Searchlight MVPA results for Necker cube in the UnAmb and Amb conditions and the hemodynamic delay control analysis. (*Left*) Searchlight MVPA group analysis results for the UnAmb condition and different frames of the Amb condition. Maps were thresholded at $P < 0.05$, corrected for multiple comparisons. LH, left hemisphere; RH, right hemisphere. (*Right*) Percept-selective voxels were chosen from searchlight results of each Amb frame and separated according to their preferred UnAmb images. fMRI time courses were averaged across each voxel group for button presses indicating preferred percept (green) vs. nonpreferred percept (red). Data were pooled across the four selective voxel groups (preferring face, vase, the two perspectives of Necker cube, respectively) and averaged across subjects ($n = 11$). Dashed boxes indicate the corresponding frame of searchlight results from which the selective voxels were chosen. Time point 0 is the frame containing the button press. Error bars denote SEM across subjects.

Next, we conducted searchlight MVPA across the whole brain to identify regions whose activity patterns reflected fluctuating percepts during bistable perception in the Amb condition. Searchlight decoding was carried out for five fMRI frames (herein we use "frame" and "volume" interchangeably, repetition time = 2.0 s) surrounding the button presses separately, from one frame before (frame −1) to three frames after (frame 3) the button press, with frame 0 defined as the one including the button press. We found that, even in the frame before the button press, brain activity patterns in medial and orbitofrontal cortices, precentral/central sulci, and ventral temporal and insular regions were able to decode the upcoming perceptual change. As time progressed relative to the button presses, percept-reflective activity moved from frontal and anterior temporal regions to posterior visual cortices. Both bistable images produced qualitatively similar results (for Necker cube, see Fig. 2, left column; for Rubin face-vase, see Fig. S1). Locations of the voxels with the highest decoding accuracy (averaged across subjects) in each frame are shown in Fig. S2.

Due to the effect of hemodynamic delay in the blood oxygen level-dependent (BOLD) fMRI signal, it is important to ensure that activity patterns identified above contained information about the current percept indicated by the button press during frame 0, and not the percept immediately preceding the button press. To this end, we conducted a control analysis (for details, see *SI Methods*, *Control for Hemodynamic Delay*). Briefly, for each bistable image, and at each fMRI frame analyzed, we separated the predictive voxels identified by searchlight MVPA into two groups, each corresponding to having higher BOLD activity for one of the percepts (determined by a t test using UnAmb data). If a given frame contained activity for the percept indicated by the current button press, then, for each voxel group, a higher BOLD activity would be expected if that button press corresponded to its preferred percept compared with the opposite percept. The contrary would be true, however, if the activity analyzed reflected the previous percept. The results from this analysis verified that, for all five frames, regions identified by the MVPA analysis contained information about the current percept indicated by the button press at frame 0 (Fig. 2, right column). This analysis confirmed that activity patterns in higher-order brain regions are able to predict the upcoming perceptual change up to 2 s before the button press.

To compare brain activity patterns underlying simple, unambiguous perception with those underlying bistable perception, we combined decoding results in the Amb condition across the five frames and overlaid the resulting image with the UnAmb decoding result (Fig. 3*A*). Although UnAmb and Amb conditions shared substantial overlap in visual cortices, there were many more frontoparietal, anterior temporal, and insular regions recruited in the Amb condition, suggesting that bistable perception requires many more brain resources than simple, unambiguous perception (also, see Fig. S5).

Lastly, we investigated whether the fine-grained representation within a region is similar between the UnAmb and Amb conditions. To this end, we trained the searchlight classifier on the UnAmb data set and tested it on different frames of the Amb data surrounding the button presses. This analysis was carried out across the whole brain. We found that isolated regions in frontoparietal, anterior, and ventral temporal cortices (Fig. S3) were able to cross-decode, suggesting that the fine-grained representations in these regions are similar across the two conditions. Notably, although activity pattern in V1 was able to decode the percepts in both UnAmb and Amb conditions (Fig. 3*A*), it was not able to cross-decode (Fig. S3), consistent with the fact that low-level features of the physical stimuli are different between these two conditions.

Brain Activity Patterns Involved in Intermittent Bistable Perception. We then applied searchlight MVPA to data from the discontinuous (Disc) condition, decoding the frame with image presentation (frame 0) and the three ensuing blank frames (frames 1~3) separately. Only blank periods between consecutive image presentations with the same button response were analyzed, in which case the content of the perceptual memory trace during the intervening blank period could be unequivocally determined. The decoding results for the different frames are shown in Fig. S4, and the locations of the voxels with the highest decoding accuracy in each frame are plotted in Fig. S2. Due to the effect of hemodynamic delay, it is not possible to clearly dissociate brain activity patterns underlying bistable perception and those underlying perceptual memory in this case. However, because the Disc condition contains both bistable perception and perceptual memory, whereas the Amb condition contains only bistable perception, the difference in decoding results between them may reflect brain activity patterns underlying perceptual memory. To compare these two conditions, we combined decoding results across the four frames in the Disc condition and across the five frames in the Amb condition and overlaid the resulting images (Fig. 3B). Compared with the Amb condition, the Disc condition involved many additional higher-order regions in the prefrontal

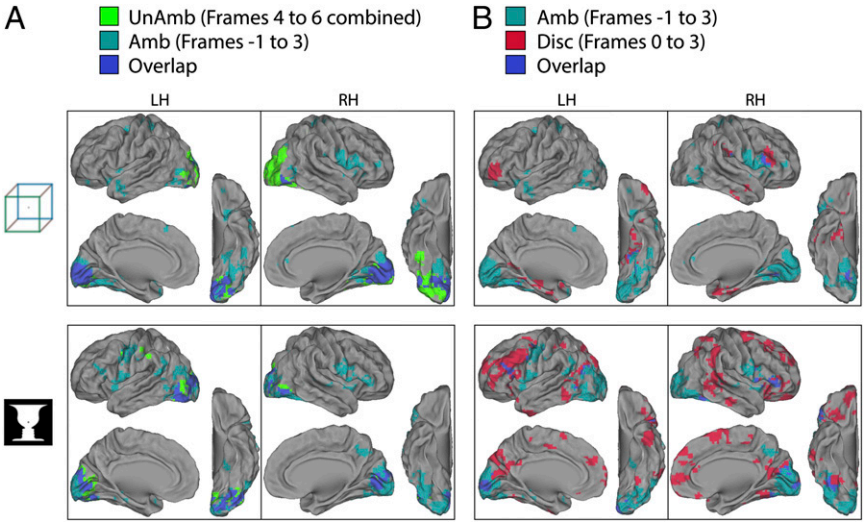

Fig. 3. Comparison of searchlight MVPA results between conditions. (*A*) Comparison between UnAmb and Amb conditions. (*B*) Comparison between Amb and Disc conditions. Results from Necker cube and Rubin face-vase stimuli are shown in the *Upper* and *Lower* row, respectively. All results are from group analysis, thresholded at $P < 0.05$, corrected for multiple comparisons. The Amb condition results were combined across frames −1 to 3. The Disc condition results were combined across frames 0 to 3. LH, left hemisphere; RH, right hemisphere.

cortex, temporoparietal junction (TPJ), and anterior temporal cortex (Fig. 3B; also see Fig. S5).

Changes in Interregional Directed Influences Across Perceptual Conditions. We used GC analysis to assess changes in directed influences between brain regions across the three experimental conditions. Although it may not be possible to infer interregional absolute causal relations with GC analysis applied to fMRI data due to heterogeneous hemodynamic delay across the brain (46–48), changes in GC patterns across experimental conditions are not subject to this confound, as a given region's hemodynamic response profile is independent of task conditions (49).

Based upon the MVPA results (combined across the three conditions), we defined 21 and 24 Regions of Interest (ROIs) for the Necker cube and Rubin face-vase stimulus, respectively (Fig. 4*A* and Table S1). ROIs were ordered from posterior to anterior according to their Talairach coordinates. Because visual sensory regions are located posteriorly, we used GC influences in the posterior-to-anterior direction to approximate "bottom-up" influences, and the converse to approximate "top-down" influences. We emphasize that this is an approximation, not only because locations on the posterior–anterior axis provide a very crude correspondence to hierarchy, but also because there are many parallel pathways in the brain without any clear hierarchical relationship. Given that the ROIs were selected as clusters of voxels whose fine spatial patterns contained information about the perceptual content, we performed voxel-wise GC analysis between every pair of ROIs (see *SI Methods*, *Granger Causality Methods*). The percentages of significant voxel pairs for every ROI pair in both directions of influence are shown in Fig. 4*B*. Across the three experimental conditions and both bistable stimuli, there was consistently greater recurrent connectivity among the most posterior ROIs compared with anterior ROIs. In Amb and Disc conditions, there appeared to be more top-down (anterior-to-posterior, upper-right triangle in each matrix) than "bottom-up" (posterior-to-anterior, lower-left triangle in each matrix) influences. As mentioned above, an absolute interpretation of single-condition GC result is difficult; we therefore focused on contrasts between conditions, as reported below.

Compared with the UnAmb condition, we found in the Amb condition a dramatic elevation in connectivity across the brain in both posterior-to-anterior and anterior-to-posterior directions (Fig. 4*C*, *Left*), indicating that both bottom-up and top-down influences are strengthened in the Amb condition. Overall, 95% of all ROI pairs for both Necker cube and Rubin face-vase images showed significantly higher connectivity in the Amb compared with UnAmb condition, whereas less than 1% of all ROI pairs for each bistable image exhibited significantly lower connectivity in the Amb condition [assessed by McNemar test, $P < 0.05$, false-discovery rate (FDR) corrected].

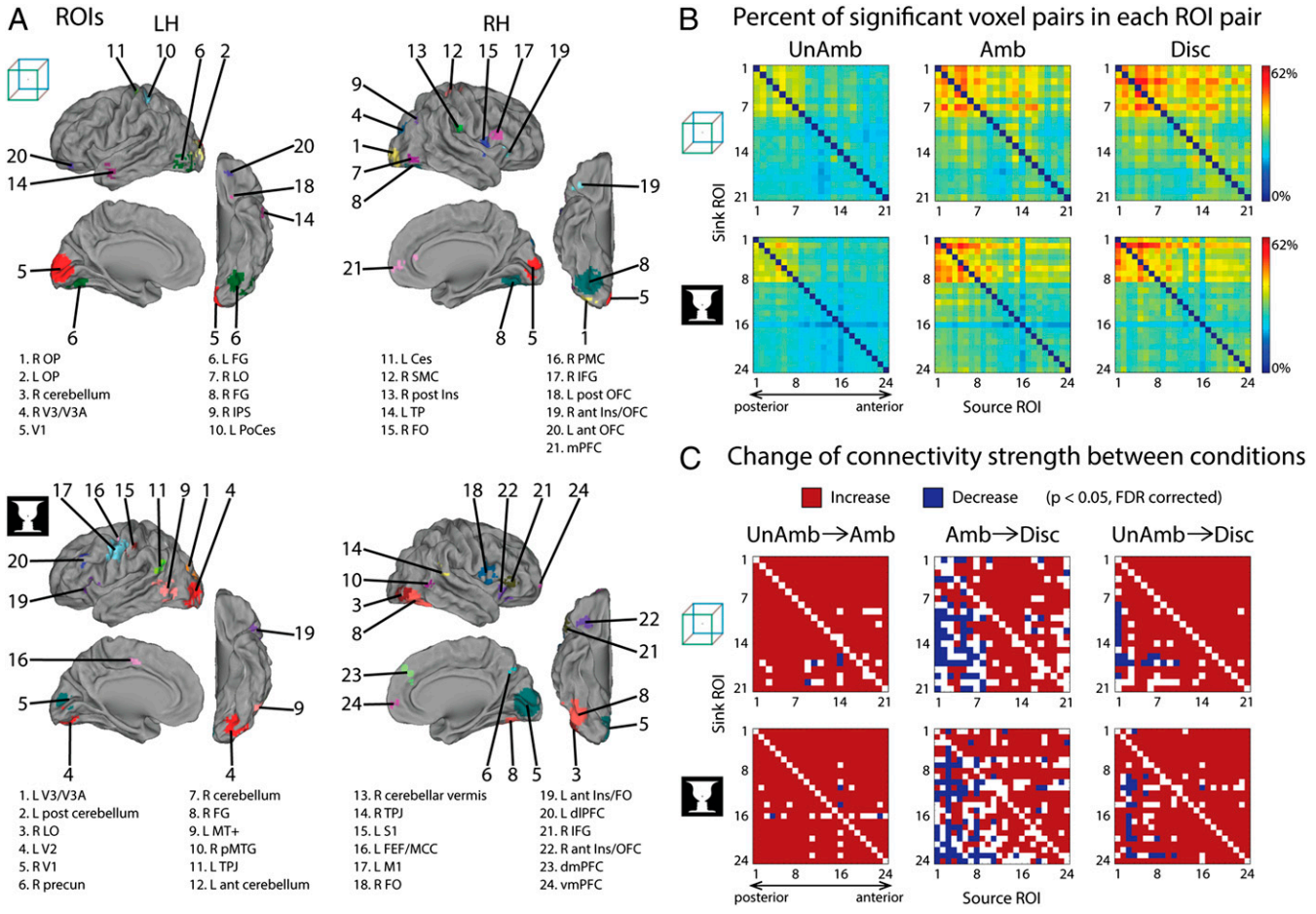

Fig. 4. ROIs and GC analysis results. (*A*) All ROIs used for GC analysis are plotted on a standard brain surface. ROIs are ordered according to their posterior–anterior position in the Talairach space. Their abbreviated names are shown on the bottom of each graph. For ROI details, see Table S1. (*B*) Percentage of significant voxel pairs for each ROI pair in each direction under UnAmb (*Left*), Amb (*Center*), and Disc (*Right*) conditions. Direction of GC influence is from the source ROI to the sink ROI. (*C*) Changes in connectivity strengths between conditions. Changes from UnAmb to Amb (*Left*), Amb to Disc (*Center*), and UnAmb to Disc (*Right*) conditions with significant increases (red) and decreases (blue) of connectivity strength (McNemar test; $P < 0.05$, FDR corrected).

When contrasting the Amb and Disc conditions, we found that 62% (Necker cube) and 59% (Rubin face-vase) of all ROI pairs demonstrated higher connectivity in the Disc condition whereas only 16% of all ROI pairs for either stimulus showed lower connectivity in the Disc condition (Fig. 4*C*, *Center*; $P < 0.05$, FDR corrected). Remarkably, the vast majority (84% for cube, 86% for face-vase) of directed influences having lower strength in the Disc condition resided in the posterior-to-anterior direction (lower-left triangles). Moreover, most of these diminished connectivity originated from the eight most posterior ROIs for both stimuli, which included mostly early and ventral visual regions. These results suggest that there were enhanced top-down influences in the Disc condition whereas bottom-up influences were weakened, consistent with the MVPA results showing that the Disc condition recruited mainly higher-order regions.

Lastly, a comparison between the UnAmb and Disc conditions showed that the majority of ROI pairs had higher connectivity in the Disc condition (88% for Necker cube; 87% for Rubin face-vase) whereas only 5% of all ROI pairs for either stimulus showed lower connectivity in the Disc condition (Fig. 4*C*, *Right*; $P < 0.05$, FDR corrected). Consistent with earlier results, most of the reduced connectivity in the Disc condition was from posterior visual regions to more anterior ROIs.

Graph-Theoretic Analysis. To further quantify the interregional interaction patterns revealed by the GC analysis, for each ROI, we obtained several metrics used commonly in graph-theoretic analyses: out-degree, representing the net influence a given ROI has on the rest of the network (network defined as an abstract graph including all ROIs); in-degree, representing the net influence a given ROI receives from the rest of the network; and out-in degree, the difference between out-degree and in-degree as a measure of the net causal outflow from an ROI (50–52).

We first applied this analysis to a binary connectivity matrix, where a connection is considered "present" if its percentage of significant voxel pairs exceeds a threshold. For a given ROI, the out-degree was defined as the number of ROIs to which it sends influence; similarly, the in-degree represented the number of ROIs from which it received influence. Three different thresholds (40%, 50%, and 60%) were tested, and they yielded similar results. For the sake of brevity, we report only the results using the 50% threshold below.

A two-factor ANOVA (factors: ROI and task condition) was carried out for each of the above metrics (Fig. 5 and Fig. S6). Although the effect of ROI was highly significant across the three conditions and both stimuli ($P < 0.0001$, except for out-degree, cube: $P = 0.025$), we focused on the changes of GC patterns across task conditions. The effect of condition by itself was not significant in any of the ANOVA results. However, the interaction effect of ROI × condition was highly significant for both in-degree ($P < 0.005$) and out-in degree ($P < 0.001$) across both stimuli. The in-degree increased substantially more from the UnAmb to Amb to Disc condition in posterior compared with anterior ROIs (Fig. S6). The out-in degree results (Fig. 5) suggest that, in the UnAmb condition, the input and output of each region were roughly balanced. However, during the Amb condition, posterior regions tended to have a net in-flow, indicating that they were receiving more directed influences from the rest of the network than sending out. By contrast, anterior ROIs tended to have a net out-flow, indicating that they were sending out more influences than receiving. This pattern was further intensified in the Disc condition. All of the above results were consistent across both Necker cube and Rubin face-vase stimuli.

Because the above analysis transformed the connectivity matrix into a binary matrix that did not account for the difference in connectivity strength once a connection passed the threshold, we conducted an additional analysis using weighted matrices, whereby each connection passing a threshold (50%) was weighted by its percentage of significant voxel pairs. The results from this analysis (Fig. S7) were highly similar to the above unweighted analysis.

Top-Down vs. Bottom-Up Influences During Perceptual Switching and Maintenance. Lastly, we characterized GC patterns accompanying perceptual switching and perceptual maintenance, respectively,

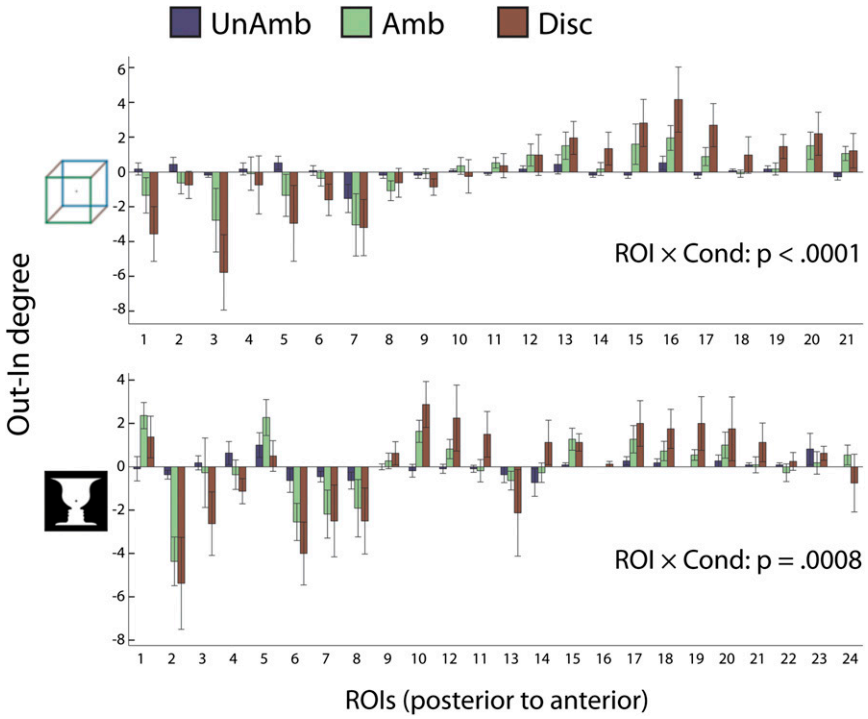

Fig. 5. Total causal flow (out-in degree) for each ROI in each condition. ROIs are ordered posterior (left-most) to anterior (right-most). Out-in degrees were computed for each ROI using binary connectivity matrices thresholded at 50%. The mean and SEM across subjects are plotted ($n = 11$, 11, and 8 for UnAmb, Amb, and Disc condition, respectively). *P* values of the ROI × condition interaction effect from a two-way ANOVA are indicated in the graph.

in the Amb condition. To this end, we defined short, 6-s trials centered around perceptual switches or during perceptual maintenance. We then conducted a voxel-wise GC analysis on the two groups of trials separately for all ROI pairs under both bistable stimuli. The raw connectivity matrices showing the percentage of significant voxel pairs for each ROI pair are presented in Fig. S8. Inspired by earlier studies (8, 9, 13–15), we were particularly interested in the potential disparity between top-down and bottom-up influences during either perceptual switching or maintenance. We thus compared the percentage of significant voxel pairs between the posterior-to-anterior and anterior-to-posterior directions across all ROI pairs (Fig. 6). During perceptual switching, there was greater anterior-to-posterior connectivity than in the opposite direction for both stimuli ($P < 0.00002$, Wilcoxon signed-rank test), implicating greater top-down influences. By contrast, there was no significant difference between the two directions under perceptual maintenance ($P > 0.6$ for both stimuli).

Discussion

Summary of Findings. In sum, we report the following main findings: (*i*) During simple unambiguous perception, activity patterns reflecting perceptual content are localized mainly to posterior visual regions whereas bistable perception in the presence of ambiguous stimuli involves both visual regions and higher-order frontoparietal and temporal regions. Interestingly, intermittent viewing of ambiguous images recruits additional higher-order frontoparietal and temporal regions. (*ii*) Compared with simple unambiguous perception, bistable perception elicits dramatically increased top-down as well as bottom-up influences. Intermittent bistable perception in turn evokes even stronger top-down influence, but relatively weakened bottom-up influence.

Mechanisms of Bistable Perception. Although researchers of bistable perception have fiercely debated the involvement of top-down (4, 53, 54) vs. bottom-up (55, 56) mechanisms for over a century, recent views support the existence of both mechanisms (2, 5, 6, 57, 58). Our results provide direct experimental support for such a view by demonstrating that both top-down and bottom-up influences are strongly elevated during bistable perception compared with simple unambiguous perception.

Previous neuroimaging and neurophysiological studies on bistable perception have respectively emphasized perceptually reflective activity patterns in lower-level visual areas (25–27, 59) and higher-order temporal and frontal regions (35, 60–63). Our findings showing that perceptual content can be decoded in these different regions in the same subjects with the same technique bring these previous results together. Further, the results from our whole-brain searchlight decoding analysis might provide guidance for future neurophysiological investigations.

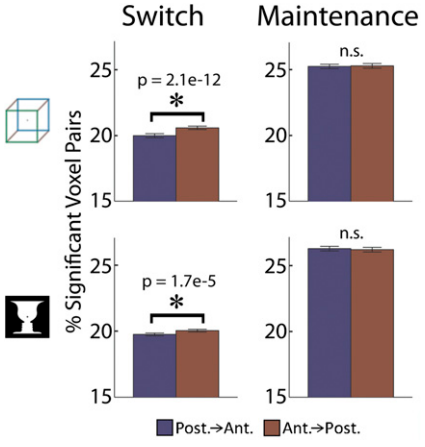

Fig. 6. GC patterns during perceptual switch vs. maintenance. Percentages of significant voxel pairs were compared between the putative bottom-up (posterior-to-anterior) and top-down (anterior-to-posterior) directions across all ROI pairs by a Wilcoxon signed-rank test (*P* values are indicated in the graph). The bar graphs plot the mean and SEM across ROI pairs.

Our finding that top-down influences outweigh bottom-up influences during perceptual switching but not maintenance (Fig. 6) is consistent with earlier neuroimaging (7–9), electrophysiological (11), TMS (13–15), and neuropsychological (16–18) findings suggesting that frontoparietal regions might initiate perceptual switching. To the extent that the attentional load might be larger during perceptual switching than maintenance, these results are also consistent with previous GC results on visual spatial attention (34). Interestingly, we did not find the right parietal region implicated in some of these prior studies in our MVPA results (Fig. 3*A*), suggesting that this region might not directly encode the perceptual content itself. Identifying switch-related regions (e.g., those with higher activity during perceptual switches) and content-related regions (as revealed by the MVPA analysis herein) in the same study and investigating the interactions between them is an important topic for future studies.

Simple, Unambiguous Perception. Our results are decidedly consistent with the predictive coding framework (1, 29–33). In the UnAmb condition, there was markedly less involvement of higher-order regions (Fig. 3*A*) and substantially weaker information flow in both bottom-up and top-down directions (Fig. 4). This observation is consistent with predictive-coding ideas, which suggest that reciprocal interactions between higher-order and lower-order regions weaken or discontinue when the "model" instantiated by the higher-order region is validated by sensory inputs and the perceptual ambiguity thereby resolved. By striking contrast, both top-down and bottom-up influences are sustained and much amplified in the Amb condition (Fig. 4) when perceptual ambiguity persists for as long as the image is being viewed.

An earlier study has reported similar modulations of frontal neuronal firing under physical stimuli alternation compared with binocular flash suppression (35). Contrastingly, we did not find perceptually reflective activity patterns in the frontal cortex in the UnAmb condition (Fig. 3*A*). This apparent difference might result from two possibilities: (*i*) In their paradigm, the alternating stimuli were checker board vs. monkey face, which convey very different conceptual and emotional values, whereas the different images in our UnAmb condition were considerably more similar (two different perspectives of the Necker cube; face vs. vase). (*ii*) Our results were corrected for multiple comparisons across the whole brain. Thus, it remains possible that our statistical power was not sufficient to detect perceptually reflective activity patterns in the frontal cortex in the UnAmb condition.

Intermittent Bistable Perception. Remarkably, our MVPA and GC analyses revealed that the Disc condition, during which the ambiguous images were viewed only a quarter of the time, recruited many additional higher-order frontoparietal and temporal regions not involved in the continuous viewing of the Amb condition (Fig. 3*B*, red) and elicited even stronger top-down influences than the Amb condition (Fig. 4*C*). These results may stem from the fact that the Disc condition invokes a strong presence of perceptual memory during the blank periods. They further raise the fascinating possibility that, after online disambiguation of bistable images, perceptual memory is transferred to a different set of regions. The detailed evolution of this process should be an interesting topic for future investigations. Our results also argue against a purely bottom-up mechanism for bistable perception, which would predict perceptual memory to be encoded solely within lower-level visual areas (42). The involvement of higher-order regions in intermittent bistable perception is consistent with an

earlier psychophysical experiment (64) and a previous fMRI activation study showing the involvement of the fusiform gyrus and frontoparietal regions in perceptual memory (41). Moreover, our finding of significant decoding in the orbitofrontal cortex (Fig. 3*B*) echoes an earlier suggestion that this region is involved in top-down facilitation of object recognition (65).

A recent study showed that the content of visual working memory can be stored in the activity patterns of early visual areas (66). Importantly, working memory is distinct from perceptual memory. Visual working memory requires active maintenance and attention and is accessible to conscious awareness (67). By contrast, perceptual memory formed during intermittent presentations of ambiguous images shares important characteristics with priming by unambiguous images (40) and is a form of implicit, unconscious memory (36).

A Tentative Conceptual Model. Our GC results can be conceptually summarized in Fig. 7*A*. In the UnAmb condition, there is recurrent processing among posterior visual regions (abstractly represented as region A), but limited top-down and bottom-up influences, as well as limited recurrent processing among higher-order regions (region B) (Fig. 4*B*, *Left*). In the Amb condition, all of the above interactions are strengthened (Fig. 4*C*, *Left*). Top-down influences are further enhanced in the Disc condition whereas bottom-up influences are weakened compared with the Amb condition (Fig. 4*C*, *Center* and *Right*).

To incorporate our MVPA findings into the picture, we separated both the higher-order region B and the lower-order region A into two respective populations, each activated by one of the percepts. A parsimonious conceptual model that can explain our results is outlined in Fig. 7*B*. Although mutual inhibition in the lower-order region cannot be ruled out, for the sake of parsimony, we placed mutual inhibition within the higher-order region only. In our view, mutual inhibition within higher-order regions could best explain our finding that, in the Disc condition, there were enhanced top-down but weakened bottom-up influences. Mutual inhibition restricted to lower-level regions also cannot explain contextual effects in binocular rivalry (e.g., see figure 2c in ref. 6).

In the Amb condition, this model could potentially explain why fMRI signals in the LGN and V1 reflect changing percepts (27, 59) but this effect is much weaker or nonexistent in neuronal firing (28, 61, 68). Because neuronal firing reflects the output of neurons in a local region (including both local recurrent processing and output to other regions) whereas the fMRI signal reflects their inputs (from both local and distant neurons) as well (69–71), the differential feedback received by lower-order regions accompanying different perceptual contents would be better reflected in the fMRI signal than neuronal firing (Fig. 7*B*, *Middle*) (72–74).

Along the same reasoning, in the Disc condition, because most lower-order visual regions cannot decode the content of perception/perceptual memory (Fig. 3*B*) despite the presence of strong top-down influences (Fig. 4*C*), a likely scenario is that the target of top-down influences does not differentiate between perceptual (/memory) states; i.e., top-down influences are diffuse (Fig. 7*B*, *Bottom*). Speculatively, this scenario could be due to the requirement of the presence of bottom-up activity to establish recurrent processing with specific top-down influences; the absence of bottom-up activity would unveil the top-down activity in the "default" state, which is more diffuse. We look forward to future experimental testing of these predictions.

To our knowledge, this conceptual model is consistent with previous neuroimaging and neurophysiological findings (Fig. 7*B*, tables on the right). At present, neurophysiological studies using paradigms similar to our UnAmb and Disc conditions are still limited. Nonetheless, our model makes predictions about what one might find in neurophysiological experiments of intermittently presented ambiguous images (Fig. 7*B*, *Bottom*): Activity patterns underlying perceptual memory content might be preferentially localized to frontal and anterior temporal regions instead of lower-order visual regions—a prediction that is consistent with preliminary findings (75). Lastly, because mutual inhibition is implemented in higher-order regions, this model can also explain why contextual effects modulate dominance but not suppression

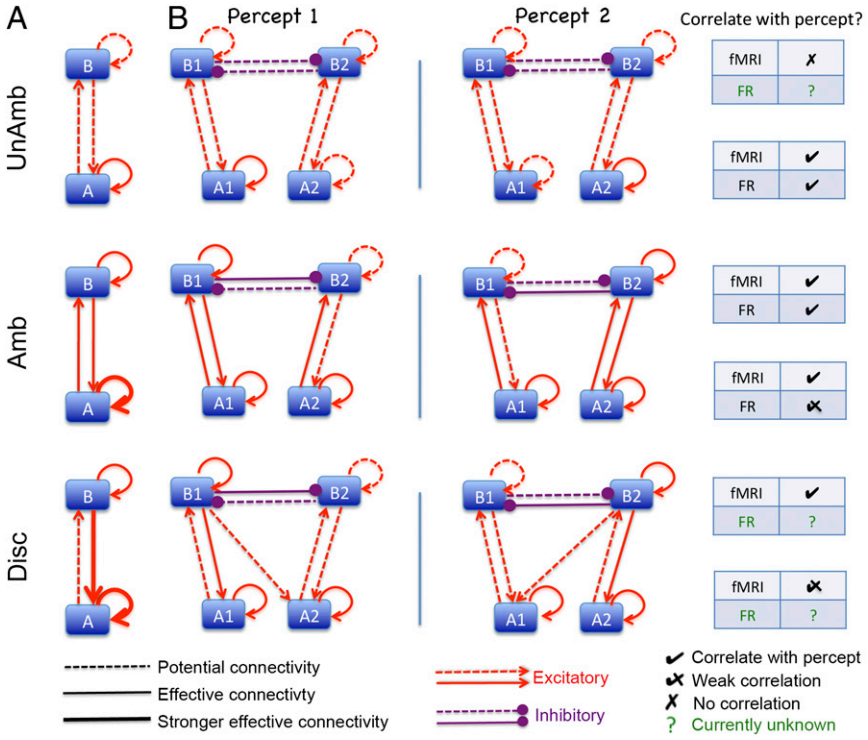

Fig. 7. A conceptual model that can account for our results. (*A*) Summary of the GC results in the three conditions. Regions A and B represent abstract lower-order and higher-order regions, respectively. Dashed lines, potential (e.g., anatomical) but weak or absent directed influence (as measured by GC). Solid lines indicate GC influences. Thicker lines indicate stronger GC influences. (*B*) A model that can explain both our MVPA and GC results. *Left* column, percept 1 is dominant. *Right* column: percept 2 is dominant. Red arrows, excitatory connections; purple lines, inhibitory connections. Tables on the right describe known (black) or currently unknown (green) fMRI and firing rate (FR) observations about whether activity in lower- or higher-level regions correlates with subjective percept in the three experimental conditions. The fMRI observations come from the MVPA results reported herein. In this graph, region A represents roughly early visual areas and region B represents roughly frontal and anterior temporal regions. We note that this two-level model is highly abstracted; in reality, there are many levels of brain regions in the hierarchy.

durations whereas stimulus strength modulates suppression but not dominance durations (6).

One potential future avenue for experimental testing of this model is to separate neuronal populations with different preferences within each region and examine the interactions between these different neuronal groups under different perceptual states. It is important to note that the model put forward here is only a conceptual one. For concrete circuit mechanisms, it would need to be formalized with quantitative computational modeling. Many previous computational models of bistable perception already exist (for a review, see ref. 76). However, our findings stress the importance of hierarchical models (43, 77–82). In particular, we hope that the present results delineating differential involvements of lower-level vs. higher-level regions in simple perception, bistable perception, and intermittent bistable perception, as well as the dramatic changes in directed influences between them across these perceptual conditions, will help constrain future hierarchical models of bistable perception. Lastly, further work on spiking models of bistable perception (83, 84) could potentially help reconcile different findings made by fMRI and neurophysiology.

Timing of Predictive Brain Activity. The fact that we were able to decode the upcoming perceptual switch up to 2 s before the button press (Fig. 2 and Fig. S1) might come as a surprise given earlier results showing that fMRI signals lagged the button press indicating perceptual switch by 1–4 s (20, 59, 85). Our control analysis demonstrated that the activity patterns we decoded indeed reflected the perceptual content indicated by the current button press rather than the delayed activity reflecting the previous percept (Fig. 2). The difference between our results and these previous studies is likely due to a difference in brain regions investigated, as these previous studies analyzed only visual areas, whereas the early predictive activity in our results resided mainly in frontal and anterior temporal regions (Fig. 2 and Fig. S1). Interestingly, primate neurophysiology studies have observed that firing rate changes in area V4 precede button presses indicating perceptual switches by up to 1 s (4). Our results are also reminiscent of reports showing that fMRI activity in prefrontal cortex encodes free decisions well before subjects' button presses (86) and that brain activity precedes the awareness of conscious volition (87, 88).

Motor-Related Activity. Because each percept was mapped to a particular button, some of the activity we decoded could be motor-related. This effect is especially evident in the left motor cortex as all subjects pressed the button with their right hand (Fig. 3). However, because the motor response is identical across the three conditions, the dramatic differences between conditions in the results from both MVPA and GC analyses cannot be attributed to the motor response.

Ambiguous Images vs. Binocular Rivalry. Bistable perception can be elicited both by viewing ambiguous images as used in the present study and by presenting dissimilar images to the two eyes (i.e., binocular rivalry). Previous theoretical and computational studies have often assumed that the underlying mechanisms are similar between them. Currently, considerable evidence suggests the importance of top-down influences in both forms of bistable perception (4, 5). Nonetheless, whether our present results can be generalized to binocular rivalry awaits future investigation. For example, a previous psychophysical study found that bistable perception elicited by ambiguous images is more susceptible to attentional modulation than that elicited by binocular rivalry and thus may have a stronger top-down component (89).

Broader Implications. Our results strongly support the predictive coding ideas in visual perception (1, 29, 31–33). They further reveal specific large-scale network mechanisms underlying simple vs. bistable perception, as well as those potentially underlying perceptual memory elicited in intermittent bistable perception. As Carandini et al. suggested, "The ultimate test of any theory of the neural basis of visual perception is its ability to predict neuronal responses during natural vision" (90). Because natural vision is marked by the needs to resolve ambiguities imposed by complex natural scenes as well as the ever-present noise and incompleteness of the retinal image (1, 3, 5), it might reside somewhere between the unambiguous simple perception and ambiguous bistable perception studied herein. The mechanisms underlying perceptual memory might also contribute to natural vision, given the seamlessly flowing nature of our visual consciousness despite unstable and incomplete retinal images.

Methods

The experiment was approved by the Institutional Review Board of the National Institute of Neurological Disorders and Stroke. Thirteen healthy right-handed volunteers between 19 and 37 y of age (5 females) with normal or corrected-to-normal vision participated in the study. All subjects provided written informed consent. Two subjects were excluded due to excess movement in the scanner. Three additional subjects were excluded from the Disc condition analyses due to an insufficient number of perceptual switches required for the MVPA analysis.

Additional methods can be found in *SI Methods*.

ACKNOWLEDGMENTS. We thank David Leopold and Panagiota Theodoni for comments on a previous draft of the manuscript and Avi Snyder and Don Zhang for sharing cortical parcellation ROIs. This research was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Neurological Disorders and Stroke.

1. Olshausen BA, Field DJ (2005) How close are we to understanding v1? *Neural Comput* 17(8):1665–1699.
2. Long GM, Toppino TC (2004) Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychol Bull* 130(5):748–768.
3. Albright TD (2012) On the perception of probable things: Neural substrates of associative memory, imagery, and perception. *Neuron* 74(2):227–245.
4. Leopold DA, Logothetis NK (1999) Multistable phenomena: Changing views in perception. *Trends Cogn Sci* 3(7):254–264.
5. Sterzer P, Kleinschmidt A, Rees G (2009) The neural bases of multistable perception. *Trends Cogn Sci* 13(7):310–318.
6. Blake R, Logothetis NK (2002) Visual competition. *Nat Rev Neurosci* 3(1):13–21.
7. Sterzer P, Kleinschmidt A (2007) A neural basis for inference in perceptual ambiguity. *Proc Natl Acad Sci USA* 104(1):323–328.
8. Lumer ED, Friston KJ, Rees G (1998) Neural correlates of perceptual rivalry in the human brain. *Science* 280(5371):1930–1934.
9. Kleinschmidt A, Büchel C, Zeki S, Frackowiak RS (1998) Human brain activity during spontaneously reversing perception of ambiguous figures. *Proc Biol Sci* 265(1413):2427–2433.
10. Sterzer P, Russ MO, Preibisch C, Kleinschmidt A (2002) Neural correlates of spontaneous direction reversals in ambiguous apparent visual motion. *Neuroimage* 15(4):908–916.
11. Britz J, Landis T, Michel CM (2009) Right parietal brain activity precedes perceptual alternation of bistable stimuli. *Cereb Cortex* 19(1):55–65.
12. Knapen T, Brascamp J, Pearson J, van Ee R, Blake R (2011) The role of frontal and parietal brain areas in bistable perception. *J Neurosci* 31(28):10293–10301.
13. de Graaf TA, de Jong MC, Goebel R, van Ee R, Sack AT (2011) On the functional relevance of frontal cortex for passive and voluntarily controlled bistable vision. *Cereb Cortex* 21(10):2322–2331.
14. Kanai R, Carmel D, Bahrami B, Rees G (2011) Structural and functional fractionation of right superior parietal cortex in bistable perception. *Curr Biol* 21(3):R106–R107.
15. Zaretskaya N, Thielscher A, Logothetis NK, Bartels A (2010) Disrupting parietal function prolongs dominance durations in binocular rivalry. *Curr Biol* 20(23):2106–2111.
16. Windmann S, Wehrmann M, Calabrese P, Güntürkün O (2006) Role of the prefrontal cortex in attentional control over bistable vision. *J Cogn Neurosci* 18(3):456–471.
17. Meenan JP, Miller LA (1994) Perceptual flexibility after frontal or temporal lobectomy. *Neuropsychologia* 32(9):1145–1149.
18. Ricci C, Blundo C (1990) Perception of ambiguous figures after focal brain lesions. *Neuropsychologia* 28(11):1163–1173.
19. Hesselmann G, Kell CA, Eger E, Kleinschmidt A (2008) Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proc Natl Acad Sci USA* 105(31):10984–10989.
20. Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21(4):753–759.
21. Polonsky A, Blake R, Braun J, Heeger DJ (2000) Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nat Neurosci* 3(11):1153–1159.

22. Hsieh PJ, Caplovitz GP, Tse PU (2006) Bistable illusory rebound motion: Event-related functional magnetic resonance imaging of perceptual states and switches. *Neuroimage* 32(2):728–739.
23. Wunderlich K, Schneider KA, Kastner S (2005) Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nat Neurosci* 8(11):1595–1602.
24. Muckli L, et al. (2002) Apparent motion: Event-related functional magnetic resonance imaging of perceptual switches and States. *J Neurosci* 22(9):RC219.
25. Haynes JD, Rees G (2005) Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol* 15(14):1301–1307.
26. Lee SH, Blake R, Heeger DJ (2007) Hierarchy of cortical responses underlying binocular rivalry. *Nat Neurosci* 10(8):1048–1054.
27. Haynes JD, Deichmann R, Rees G (2005) Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature* 438(7067):496–499.
28. Keliris GA, Logothetis NK, Tolias AS (2010) The role of the primary visual cortex in perceptual suppression of salient visual stimuli. *J Neurosci* 30(37):12353–12365.
29. Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360(1456):815–836.
30. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66(3):241–251.
31. Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis* 20(7):1434–1448.
32. Yuille A, Kersten D (2006) Vision as Bayesian inference: Analysis by synthesis? *Trends Cogn Sci* 10(7):301–308.
33. Bar M (2007) The proactive brain: Using analogies and associations to generate predictions. *Trends Cogn Sci* 11(7):280–289.
34. Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M (2008) Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J Neurosci* 28(40):10056–10061.
35. Panagiotaropoulos TI, Deco G, Kapoor V, Logothetis NK (2012) Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron* 74(5):924–935.
36. Tulving E, Schacter DL (1990) Priming and human memory systems. *Science* 247(4940): 301–306.
37. Magnussen S, Greenlee MW (1999) The psychophysics of perceptual memory. *Psychol Res* 62(2-3):81–92.
38. Leopold DA, Wilke M, Maier A, Logothetis NK (2002) Stable perception of visually ambiguous patterns. *Nat Neurosci* 5(6):605–609.
39. Orbach J, Zucker E, Olson R (1966) Reversibility of the necker cube. VII. Reversal rate as a function of figure-on and figure-off durations. *Percept Mot Skills* 22:615–618.
40. Pearson J, Brascamp J (2008) Sensory memory for ambiguous vision. *Trends Cogn Sci* 12(9):334–341.
41. Sterzer P, Rees G (2008) A neural basis for percept stabilization in binocular rivalry. *J Cogn Neurosci* 20(3):389–399.
42. Noest AJ, van Ee R, Nijs MM, van Wezel RJ (2007) Percept-choice sequences driven by interrupted ambiguous stimuli: A low-level neural model. *J Vis* 7(8):10.
43. Gigante G, Mattia M, Braun J, Del Giudice P (2009) Bistable perception modeled as competing stochastic integrations at two levels. *PLOS Comput Biol* 5(7):e1000430.
44. de Jong MC, Kourtzi Z, van Ee R (2012) Perceptual experience modulates cortical circuits involved in visual awareness. *Eur J Neurosci* 36(12):3718–3731.
45. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103(10):3863–3868.
46. Handwerker DA, Gonzalez-Castillo J, D'Esposito M, Bandettini PA (2012) The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *Neuroimage* 62(2):1017–1023.
47. de Zwart JA, et al. (2005) Temporal dynamics of the BOLD fMRI impulse response. *Neuroimage* 24(3):667–677.
48. Roebroeck A, Formisano E, Goebel R (2011) The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *Neuroimage* 58 (2):296–302.
49. Bressler SL, Seth AK (2011) Wiener-Granger causality: A well established methodology. *Neuroimage* 58(2):323–329.
50. Seth AK (2010) A MATLAB toolbox for Granger causal connectivity analysis. *J Neurosci Methods* 186(2):262–273.
51. Sridharan D, Levitin DJ, Menon V (2008) A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci USA* 105(34):12569–12574.
52. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10(3):186–198.
53. von Helmhotz H (1910) *Treatise on Physiological Optics* (Optical Society of America, New York).
54. Walker P (1978) Orientation-selective inhibition and binocular rivalry. *Perception* 7(2): 207–214.
55. Bishop PO (1973) Neurophysiology of binocular single vision and stereopsis. *Handbook of Sensory Physiology*, ed Jung R (Springer, New York), Vol 7.
56. Jung R (1961) Neuronal integration in the visual cortex and its significance for visual information. *Sensory Communication*, ed Rosenblith WA (Wiley, New York).
57. Kornmeier J, Bach M (2012) Ambiguous figures: What happens in the brain when perception changes but not the stimulus. *Front Hum Neurosci* 6:51.
58. Kleinschmidt A, Sterzer P, Rees G (2012) Variability of perceptual multistability: From brain state to individual trait. *Philos Trans R Soc Lond B Biol Sci* 367(1591):988–1000.
59. Tong F, Engel SA (2001) Interocular rivalry revealed in the human cortical blind-spot representation. *Nature* 411(6834):195–199.
60. Logothetis NK, Schall JD (1989) Neuronal correlates of subjective visual perception. *Science* 245(4919):761–763.
61. Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379(6565):549–553.
62. Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci USA* 94(7):3408–3413.
63. Kreiman G, Fried I, Koch C (2002) Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc Natl Acad Sci USA* 99(12):8378–8383.
64. Maier A, Wilke M, Logothetis NK, Leopold DA (2003) Perception of temporally interleaved ambiguous patterns. *Curr Biol* 13(13):1076–1085.
65. Bar M, et al. (2006) Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA* 103(2):449–454.
66. Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458(7238):632–635.
67. Baddeley A (2003) Working memory: Looking back and looking forward. *Nat Rev Neurosci* 4(10):829–839.
68. Lehky SR, Maunsell JH (1996) No binocular rivalry in the LGN of alert macaque monkeys. *Vision Res* 36(9):1225–1234.
69. Logothetis NK (2008) What we can do and what we cannot do with fMRI. *Nature* 453(7197):869–878.
70. He BJ, Raichle ME (2009) The fMRI signal, slow cortical potential and consciousness. *Trends Cogn Sci* 13(7):302–309.
71. Raichle ME, Mintun MA (2006) Brain work and brain imaging. *Annu Rev Neurosci* 29:449–476.
72. Maier A, et al. (2008) Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat Neurosci* 11(10):1193–1200.
73. Wilke M, Logothetis NK, Leopold DA (2006) Local field potential reflects perceptual suppression in monkey visual cortex. *Proc Natl Acad Sci USA* 103(46):17507–17512.
74. Wilke M, Mueller KM, Leopold DA (2009) Neural activity in the visual thalamus reflects perceptual suppression. *Proc Natl Acad Sci USA* 106(23):9465–9470.
75. Tsuchiya N, Maier A, Logothetis N, Leopold D (2009) Neuronal activity in area MT during perceptual stabilization of ambiguous structure-from-motion. *J Vis* 9(8):756.
76. Braun J, Mattia M (2010) Attractors and noise: Twin drivers of decisions and multistability. *Neuroimage* 52(3):740–751.
77. Freeman AW (2005) Multistage model for binocular rivalry. *J Neurophysiol* 94(6): 4412–4420.
78. Wilson HR (2003) Computational evidence for a rivalry hierarchy in vision. *Proc Natl Acad Sci USA* 100(24):14499–14503.
79. Dayan P (1998) A hierarchical model of binocular rivalry. *Neural Comput* 10(5): 1119–1135.
80. Lago-Fernandez LF, Deco G (2002) A model ofbinocular rivalry based on competition in IT. *Neurocomputing* 44-46:503–507.
81. Lumer ED (1998) A neural model of binocular integration and rivalry based on the coordination of action-potential timing in primary visual cortex. *Cereb Cortex* 8(6): 553–561.
82. Tong F, Meng M, Blake R (2006) Neural bases of binocular rivalry. *Trends Cogn Sci* 10(11):502–511.
83. Theodoni P, Kovács G, Greenlee MW, Deco G (2011) Neuronal adaptation effects in decision making. *J Neurosci* 31(1):234–246.
84. Laing CR, Chow CC (2002) A spiking neuron model for binocular rivalry. *J Comput Neurosci* 12(1):39–53.
85. Haynes JD, et al. (2007) Reading hidden intentions in the human brain. *Curr Biol* 17(4): 323–328.
86. Soon CS, Brass M, Heinze HJ, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11(5):543–545.
87. Schurger A, Sitt JD, Dehaene S (2012) An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc Natl Acad Sci USA* 109(42): E2904–E2913.
88. Kornhuber HH, Deecke L (1965) [Changes in the brain potential in voluntary movements and passive movements in man: Readiness potential and reafferent potentials]. *Pflugers Arch Gesamte Physiol Menschen Tiere* 284:1–17.
89. Meng M, Tong F (2004) Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *J Vis* 4(7):539–551.
90. Carandini M, et al. (2005) Do we know what the early visual system does? *J Neurosci* 25(46):10577–10597.

Supporting Information

Wang et al. 10.1073/pnas.1221945110

SI Methods

Task Design. Classic Necker cube and Rubin face-vase ambiguous images were used to study bistable perception [in the ambiguous (Amb) condition] and intermittent bistable perception [in the discontinuous (Disc) condition] (Fig. 1*A*, *Lower*). To study simple unambiguous perception [in the unambiguous (UnAmb) condition], we manipulated the content, outlines, and shading of the ambiguous images to accentuate one of the two percepts (Fig. 1*A*, *Upper*). Images were presented via a back-projection display (1024 × 728 resolution, 60 Hz refresh rate) with a uniform background (white for all cube images, black for the original bistable face-vase image, white for modified face image and gray for modified vase image). The Necker cube images were presented at 15.9 × 13.2° of visual angle, and face-vase images were presented at 19.4 × 14.2° of visual angle (measured from the edges of the vase).

Each subject completed nine fMRI runs, including three repeating sets of three consecutive runs in the following order: UnAmb, Amb, and Disc conditions (Fig. 1*B*). Each UnAmb run contained 16 blocks, with each block consisting of 4 s of written instruction, 2 s of rest (fixation on a crosshair in the center of the screen), 16 s of image presentation, and 4 s of rest. The four unambiguous images (Fig. 1*A*, *Upper*) were presented four times each in a pseudorandom order. Subjects were asked to indicate their percept via one of two buttons at each image presentation. Occasionally, they indicated perceptual changes with multiple button presses during a single image presentation; these trials were excluded from the analyses.

Amb runs contained six task blocks each, with each block consisting of 4 s of written instruction, 2 s of rest, 60 s of image presentation, and 6 s of rest. Each ambiguous image (Necker cube or Rubin face-vase; Fig. 1*A*, *Lower*) was presented three times in a pseudorandom order. Subjects reported every spontaneous perceptual switch via one of two buttons throughout the course of image presentation. It was emphasized that the subject should not attempt to intentionally switch or hold percepts.

Each Disc run also contained six blocks, with each block consisting of 3 s of written instruction, 1 s of rest, nine repetitions of 2-s image presentation followed by a 6-s blank period, and lastly 4 s of rest. Subjects were asked to indicate their dominant percept during each image presentation via a button press. Perceptual switching during the 2-s image presentation was very rare and was excluded from analyses. The two ambiguous images were presented in alternating blocks.

Subjects were instructed to fixate upon a crosshair at the center of the screen at all times to avoid the potential influence of gaze on perception (1). In both Amb and Disc conditions, subjects were also given the choice to press a third button if they were unsure of their percept. Before entering the scanner, subjects performed practice runs until they were comfortable with the task and the buttons corresponding to each percept.

fMRI Data Acquisition. Functional and anatomical imaging was conducted on a General Electric 3T scanner with an 8-channel head coil. Anatomical images were obtained using a sagittal magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) sequence with a resolution of 1 × 1 × 1 mm^3. An axial T2-weighted structural scan was acquired with TR = 4,200 ms, TE = 120 ms, and a resolution of 3 × 3 × 3 mm^3. BOLD-contrast functional images were obtained using a single-shot gradient echo sequence, with 39 contiguous transverse slices covering the whole brain (slice thickness = 3 mm; in-plane resolution, 3 × 3 mm^2, TR = 2,000 ms, TE = 27 ms, flip angle = 90°).

fMRI Data Preprocessing. fMRI data were preprocessed as follows: (*i*) compensation of systematic, slice-dependent time shifts; (*ii*) elimination of systematic odd-even slice intensity difference due to interleaved acquisition; (*iii*) rigid body correction for inter-frame head motion within and across runs; and (*iv*) intensity scaling to yield a whole-brain mode value of 1,000 (with a single scaling factor for all voxels). Atlas registration was achieved by computing affine transforms connecting the fMRI run first frame (averaged over all runs after cross-run realignment) with the T2- and T1- weighted structural images (2). Our atlas representative template included MP-RAGE data from 12 normal individuals and was made to conform to the 1988 Talairach atlas (3). Data were resampled to 3 × 3 × 3 mm^3 voxels after atlas registration. In all following analyses, data from Necker cube and Rubin face-vase image presentations were analyzed separately.

Multivariate Pattern Analysis. Searchlight multivariate pattern analysis (MVPA) was iteratively carried out for every voxel across the whole brain to find local patterns of brain activity reflective of perceptual content. Exploratory analyses limited to training data revealed that a searchlight radius of 13 mm (corresponding to a sphere of 341 voxels) was the optimal size. Each searchlight sphere was evaluated by a gray-matter mask defined on the atlas image, and voxels outside the gray-matter mask were removed from the analyses.

MVPA was performed using the PyMVPA software package (4). Classifier decoding of the subject's percepts was achieved through a linear support vector machine (C = 1.0) using the LIBSVM implementation (www.csie.ntu.edu.tw/~cjlin/libsvm/). Preprocessing was performed on the raw fMRI data by removing linear trends and normalization to rest volumes on a per voxel basis. To prevent any bias in classifier decoding, the number of samples for each percept across every training fold was equalized. Prediction accuracies were obtained using an *N*-fold cross-validation classification regime, where *N* was equal to the number of blocks for each image. In the Amb and Disc conditions, $n = 9$, as each image was presented in three blocks during each of the three runs. For the UnAmb condition, $n = 12$, as each pair of Necker cube or Rubin face-vase images was presented four times in each of the three runs. An accuracy value was obtained for the center voxel of each searchlight sphere and then averaged across folds. For the UnAmb condition, frames 4 through 6 (image onset is aligned with the onset of frame 1) were averaged and then decoded together. For the Amb condition, frames −1 to 3 (with frame 0 being the one including the button press) surrounding the button presses were separately decoded (e.g., training and testing on frame −1 exclusively). For the Disc condition, frames 0 to 3 (with frame 0 being image presentation, and frames 1∼3 being blank period) were separately decoded (Fig. 1*B*).

For the Amb and Disc conditions, blocks were considered valid only if there was at least one button press for each percept. "Unsure" button presses were excluded from analyses. To ensure adequate spacing of reported percepts during the Amb condition, we discarded the first button press for button presses occurring less than 4 s apart. As the majority of percept durations were greater than 4 s (Fig. 1*C*), a sufficient number of trials were preserved for classifier training and testing. Button presses occurring within the last three frames of each block were discarded to allow a complete analysis of postbutton-press frames. In the Disc condition, subjects occasionally failed to provide a button response during the 2-s stimulus presentation period. In these situations, it was not clear what the perceptual state would have

been. Also, if the indicated percepts were different between consecutive image presentations, the nature of the perceptual memory content during the intervening blank period could not be determined. Therefore, only blank periods between consecutive image presentations with the same button response, and their preceding frame containing image presentation, were analyzed.

For the group analysis, each subject's accuracy map was first smoothed with a Gaussian kernel (FWHM = 5 mm). A one-sample t test was then performed against chance level (accuracy = 0.5) for every voxel to create a probability map. To correct for multiple comparisons, we followed the cluster-wise correction procedure used in Johnson et al. (5) and applied a threshold of $P < 0.01$ and cluster size >30 voxels, which yielded clusters that survived $P < 0.05$.

Control for Hemodynamic Delay. Due to the effect of hemodynamic delay, it is important to ensure that the BOLD activity we decoded is related to the percept indicated by the button press at frame 0 and not the preceding percept. This issue is especially important in the context of MVPA. Because there are only two responses used for decoding (i.e., "labels") and the previous button press is always the opposite of the current button press, a classifier that is based on BOLD activity related to the previous button press can perfectly decode the current button press, simply by adding a negative sign to all of the classifier weights (compared with decoding the previous button press). Because the training and testing data sets are subject to the same hemodynamic delay, this problem cannot be solved by MVPA applied to Amb-condition data alone. In other words, if the classifier is relying on BOLD activity corresponding to the previous button press, it can decode the current button press just as well as the previous button press, by switching the sign of classifier weights.

To address this issue, and to confirm that the BOLD activity we decoded corresponded to the current button press, a control analysis was conducted. First, a two-sample t test was performed on the peak BOLD responses to the presentations of unambiguous images (averaged across frames 4 and 5) between percepts from the same bistable image (e.g., face vs. vase) in a voxel-wise manner. Voxels were considered "preferring voxels" if the t test was significant ($P < 0.05$). These voxels were then separated into respective percept-selectivity based on the sign of the t test (e.g., face-preferring voxels and vase-preferring voxels).

Second, for each subject, we defined a set of masks as the conjunction of voxels selected using the following criteria: (*i*) belonging to one of the four groups of preferring voxels as defined above; (*ii*) having an accuracy greater than 0.7 for the individual subject Amb condition searchlight result for a specific frame (one mask for each of frames −1 through 3); (*iii*) confined within a gray-matter mask defined on the atlas image; and (*iv*) having less than 0.1% change BOLD variance across the Amb runs, to remove voxels contaminated by noise. This process yielded a total of 20 masks per subject: 4 preferring-voxel groups (criterion *i*) multiplied by 5 frames (criterion *ii*). The numbers of voxels included in each mask are listed in Table S2.

Last, for each mask defined above, the BOLD signal from the Amb condition was extracted from three frames before to five frames after each button press and then averaged across button presses indicating the same percept. BOLD activity accompanying preferred button presses (e.g., face-indicating button press for face-preferring voxels) and nonpreferred button presses (e.g., vase-indicating button press for face-preferring voxels) were then separately averaged across the four preferring-voxel groups (corresponding to face, vase, and two percepts of Necker cube, respectively) and across subjects. A paired t test was used to compare the BOLD signal time courses at each time point around preferred vs. nonpreferred button presses across subjects.

The results of this analysis (Fig. 2, right column) suggest that, at each frame around the button presses in the Amb condition, for the voxels that significantly decoded between the two percepts, their activity was higher if the button press at frame 0 corresponded to their "preferred percept" as determined by the UnAmb data. For example, a face-preferring voxel has higher activity if the button press at frame 0 indicated the switching from "vase" percept to "face" percept. Importantly, this result holds for voxels from all frames decoded by MVPA (frame −1 to 3; Fig. 2). These results convincingly demonstrate that the BOLD activity we decoded is related to the current button press, and that, 2 s before the button press, activity patterns in higher-order brain areas could already predict the upcoming percept.

UnAmb–Amb Cross-Decoding. To investigate whether the fine-grained representation within a region is similar between the UnAmb and Amb conditions, we performed cross-decoding using searchlight MVPA as described above. For each subject, the classifier was trained on all valid trials of the UnAmb condition, with frames 4~6 of each trial averaged together. The classifier was then tested on all valid trials of the Amb condition. The valid trials used in the UnAmb and Amb conditions were the same set of trials used for MVPA analyses described above. The accuracy maps from each subject were then entered into group analysis and corrected for multiple comparisons as described above.

Summary of MVPA Results Using Cortical Parcellation. To summarize the MVPA results across the three conditions, we first partitioned the cerebral cortex into six disjoint cortical regions. Five of these regions were defined following methods outlined in Zhang et al. (6), and covered respectively: (*i*) prefrontal cortex; (*ii*) premotor and motor cortex; (*iii*) somatosensory cortex; (*iv*) parietal and occipital cortex; and (*v*) temporal cortex. In addition, an insular region covering the bilateral insular cortex was hand segmented based on our atlas image using Analyze (AnalyzeDirect, Inc.). The six cortical regions are shown in Fig. S5A. Next, for each condition, we obtained the number of voxels within each of the above six cortical regions that were significant for MVPA decoding ($P < 0.05$, group analysis, corrected for multiple comparisons). The results were averaged over the two ambiguous images, and reported in Fig. S5B.

Construction of Regions of Interest. To obtain Regions of Interest (ROIs), the searchlight group analysis results were first thresholded at a $P < 0.01$ (uncorrected) level and converted to $\log_{10}$ P values. Then, results from the UnAmb condition, Amb condition (frames 0 through 3), and Disc condition (frame 3) were combined additively. The resulted image was blurred with a 5-mm-radius sphere kernel and subjected to an automatic peak search. Only clusters with a peak value above 3 and cluster size >30 voxels were considered. Peaks closer than 30 mm were consolidated by algebraically averaging their coordinates. Voxels from the above image that resided within a sphere of 15-mm radius around each peak were defined as an ROI. This procedure resulted in 24 ROIs for the Rubin face-vase stimulus and 21 ROIs for the Necker cube stimulus.

Plotting of searchlight results and ROIs on the cortical surface was done in Caret, using the Humans PALS-B12 atlas (http://brainvis.wustl.edu/wiki/index.php/Caret:About).

Granger Causality Methods. Granger causality (GC) analysis was performed to assess linear directed influences between ROIs. According to the GC model, time series X_1 Granger-causes X_2 if knowing the past of X_1 helps to predict the future of X_2 better than using the past of X_2 alone. In this example, X_1 is the "Source" and X_2 is the "Sink." Because our ROIs were identified by searchlight MVPA, the fine spatial patterns within a given ROI were informative. Thus, we could not average across an ROI. Instead, GC analysis was carried out on every voxel pair for a given ROI pair, following the method used in ref. 7.

GC analysis was performed using the Granger Causal Connectivity Analysis toolbox (8). Bivariate trial-based GC analysis was used, with each trial spanning the duration of image presentation for the UnAmb (16-s) and Amb (60-s) conditions and the nine repetitions of image and blank presentations for the Disc condition (72-s). Preprocessing steps included linear detrending and temporal- and ensemble-mean removal. We confirmed that our data were covariance stationary by the augmented Dickey-Fuller test (8). Using Bayesian Information Criterion, a model order of 2 was chosen. Model validity was confirmed by Durbin–Watson residual whiteness test and the statistical percent consistency test (8).

GC Analysis. The percentage of significant ($P < 0.05$) voxel pairs from the GC analysis was determined for every ROI pair in both directions (pooled across subjects), for the three experimental conditions and two stimuli separately. In the UnAmb condition, the two modified images of Necker cube were analyzed together, as were the two modified images of Rubin face-vase. McNemar analyses were carried out for each ROI pair in each direction to compare GC strengths between experimental conditions. The McNemar test determines whether one of the two conditions has a significantly higher percentage of significant voxel pairs. McNemar test results were considered significant after correction for multiple comparisons using the false-discovery rate method (9).

To compare GC patterns during perceptual switching and perceptual maintenance in the Amb condition, trials of three-frame length were identified. For perceptual switching, the trials included all instances in which the center frame was a valid button press, and the frames before and after did not include a button press. For perceptual maintenance, periods of three-frame length demonstrating perceptual maintenance (i.e., without a button press) were included. The maintenance trials were allowed to overlap. Preprocessing steps included temporal- and ensemble-mean removal. Then, voxel-wise GC analysis was performed similarly as above using the same set of ROIs. Given the short duration of each trial, the model order was set to 1, and model validity was confirmed by the Durbin–Watson residual whiteness test and the statistical percent consistency test. The percentage of significant voxel pairs was compared between posterior→anterior (i.e., the more posterior ROI is the source) ROI pairs and anterior→posterior ROI pairs using the Wilcoxon signed-rank test.

Graph-Theoretic Analysis. The raw results from voxel-wise GC analyses are described in a 24×24 connectivity matrix for Rubin face-vase stimulus and a 21×21 connectivity matrix for Necker cube stimulus, with each element corresponding to the percentage of significant voxel pairs from the source ROI to the sink ROI (Fig. 4*B* and Fig. S8). Note that, because GC assesses directed influences, the connectivity matrix is asymmetrical. This matrix was thresholded at 50% and turned into a binary matrix. Then, for each ROI, the following metrics were determined (8, 10): (*i*) out-degree: the number of connections from a given ROI to any other ROI in the network; (*ii*) in-degree: the number of connections to a given ROI from any other ROI in the network; and (*iii*) out-in degree: the difference between out-degree and in-degree as a measure of the net causal outflow from an ROI.

A two-factor ANOVA (factors: ROI and task condition) was conducted for each of the above metrics. To evaluate the robustness of the results to the arbitrarily chosen threshold, we repeated the analysis using thresholds of 40% and 60%.

A weighted analysis was also carried out, in which each connection surviving the threshold of 50% was weighted by the percentage of significant voxels. Thus, instead of being turned into a binary matrix, the full range of values above the threshold was used.

All analyses, unless otherwise noted, were conducted using custom-written codes in C++, MATLAB, and Python.

1. Long GM, Toppino TC (2004) Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychol Bull* 130(5):748–768.
2. Ojemann JG, et al. (1997) Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage* 6(3):156–167.
3. Talairach J, Tournoux P (1988) *Co-Planar Stereotaxic Atlas of the Human Brain* (Georg Thieme Verlag, Stuttgart).
4. Hanke M, et al. (2009) PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7(1):37–53.
5. Johnson JD, McDuff SG, Rugg MD, Norman KA (2009) Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron* 63(5):697–708.
6. Zhang D, et al. (2008) Intrinsic functional relations between human cerebral cortex and thalamus. *J Neurophysiol* 100(4):1740–1748.
7. Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M (2008) Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J Neurosci* 28(40):10056–10061.
8. Seth AK (2010) A MATLAB toolbox for Granger causal connectivity analysis. *J Neurosci Methods* 186(2):262–273.
9. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc, B* 57:289–300.
10. Sridharan D, Levitin DJ, Menon V (2008) A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci USA* 105(34):12569–12574.

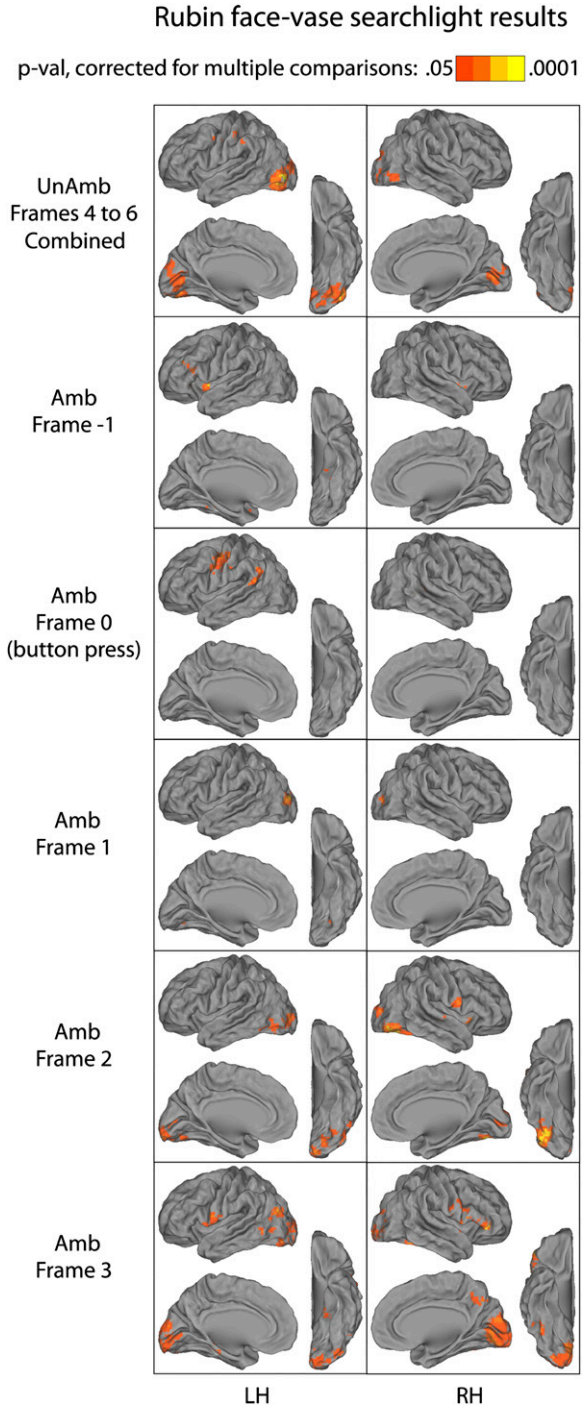

Fig. S1. Searchlight MVPA results for Rubin face-vase image in the Amb condition. Format is the same as in Fig. 2.

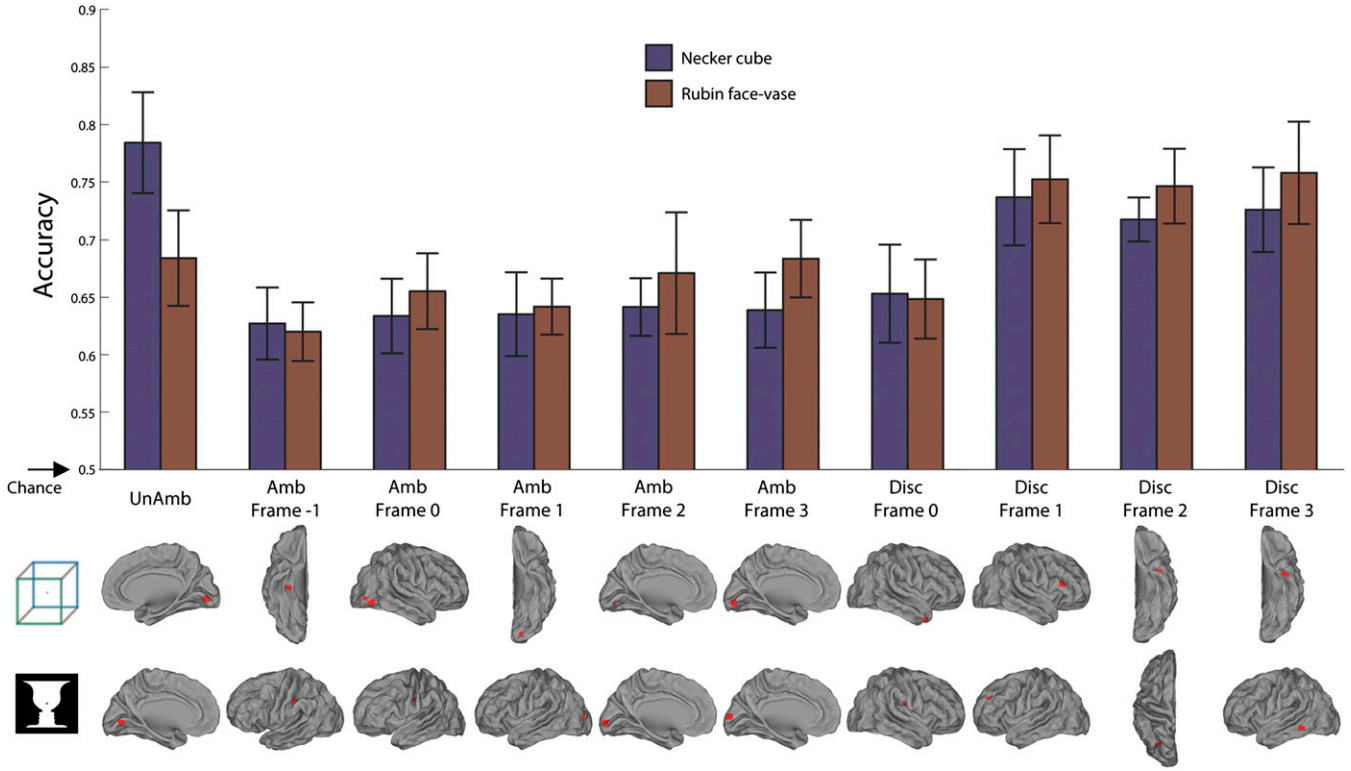

Fig. S2. For each searchlight MVPA analysis (UnAmb, the five frames in the Amb condition, and the four frames in the Disc condition), the accuracy map was averaged across subjects. The peak voxel with the highest accuracy value was obtained and plotted on the brain surface using a 6-mm-radius sphere centered on that voxel (*Lower*). The bar graph shows the mean and SEM of this voxel's decoding accuracy across subjects. We note that, if the peak voxels were defined for each individual subject, the decoding accuracy would be much higher.

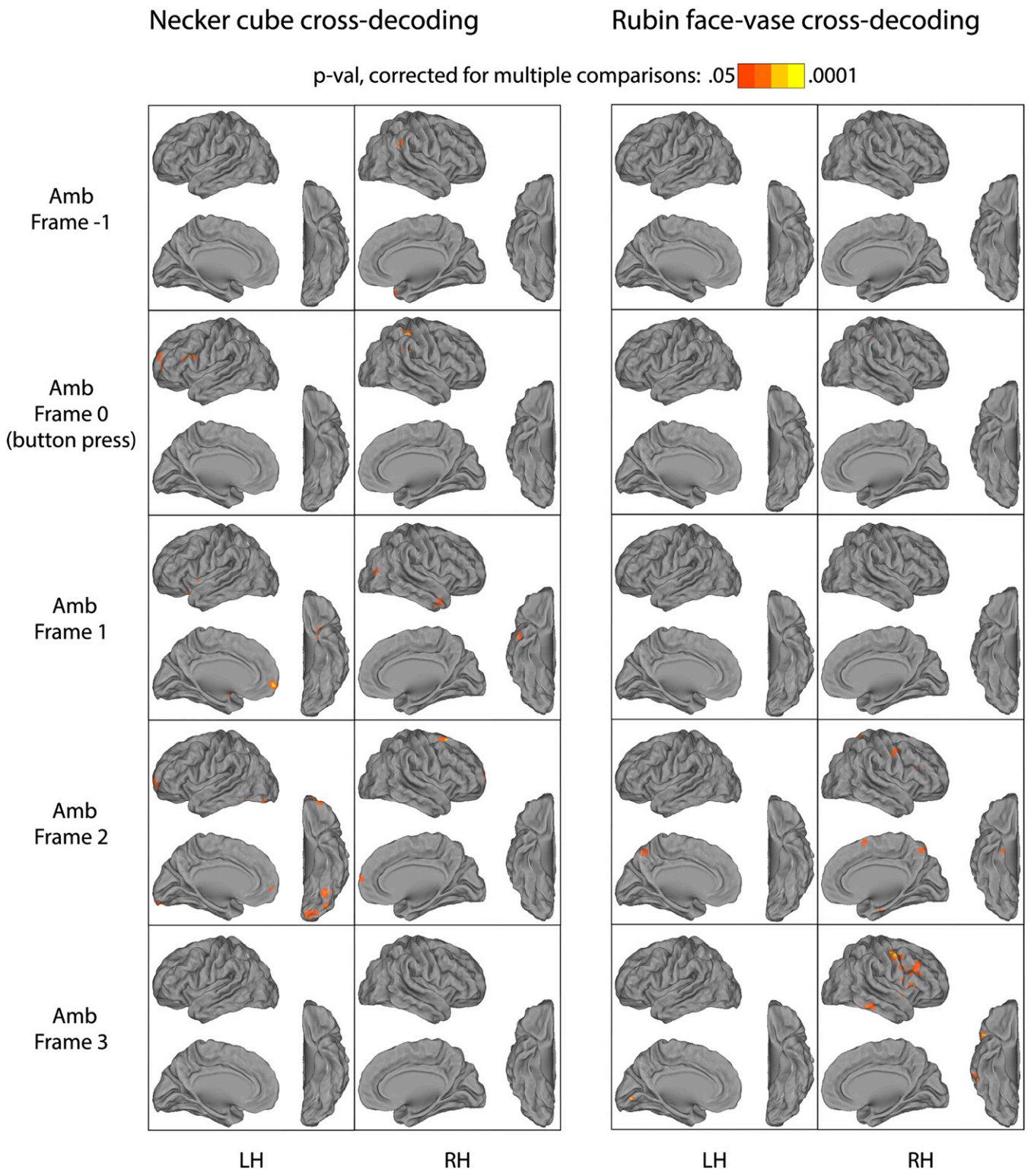

Fig. S3. UnAmb–Amb cross-decoding results for both ambiguous images. Searchlight MVPA classifier was trained on UnAmb data (frames 4~6 of each trial were averaged together), and tested on frames −1 to 3 surrounding the button press (contained in frame 0) in the Amb condition. Group analysis results were corrected for multiple comparisons and thresholded at $P < 0.05$.

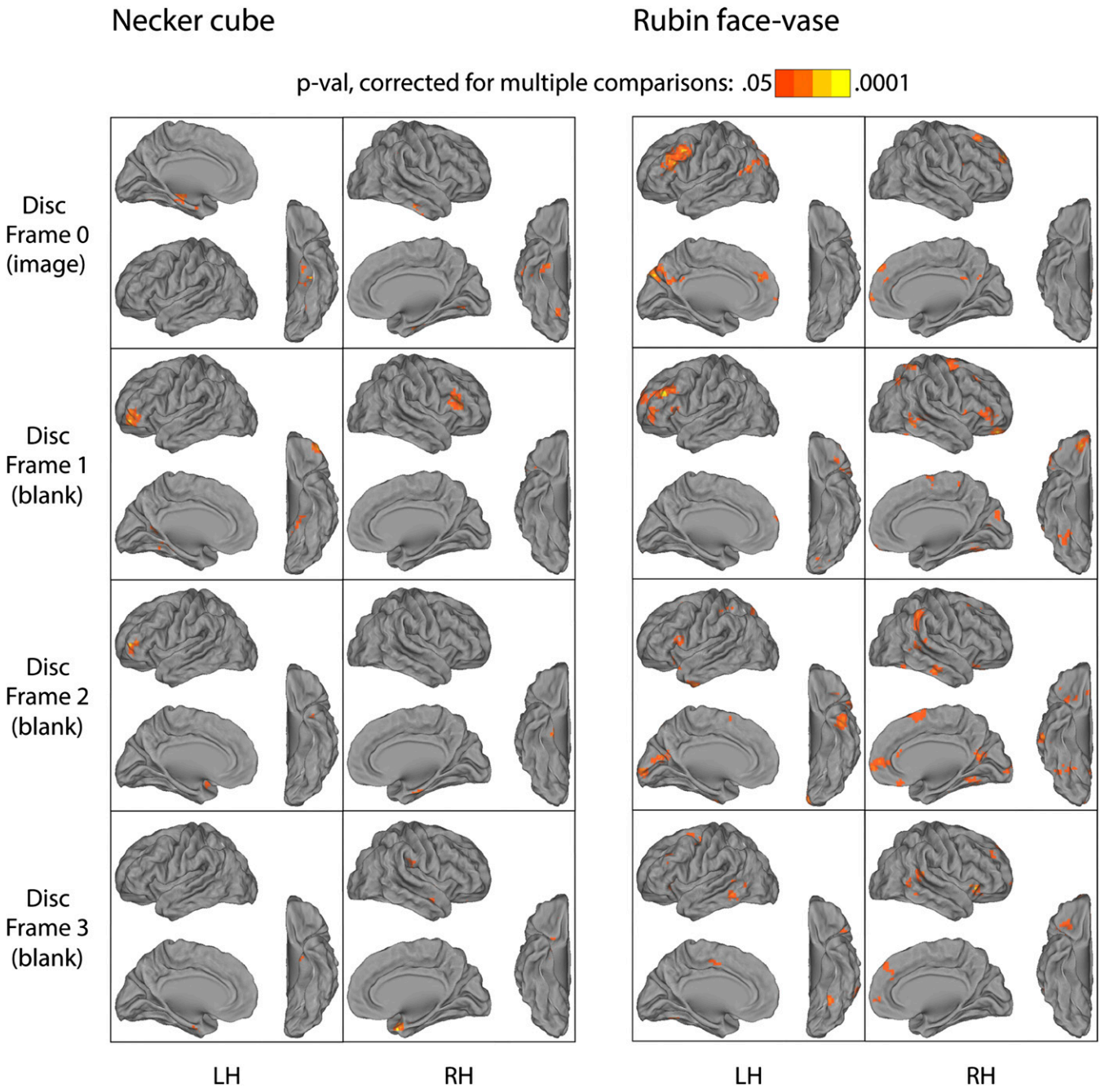

Fig. S4. Searchlight MVPA results for both ambiguous images in the Disc condition. Different frames were decoded separately. Group analysis results were corrected for multiple comparisons and thresholded at $P < 0.05$.

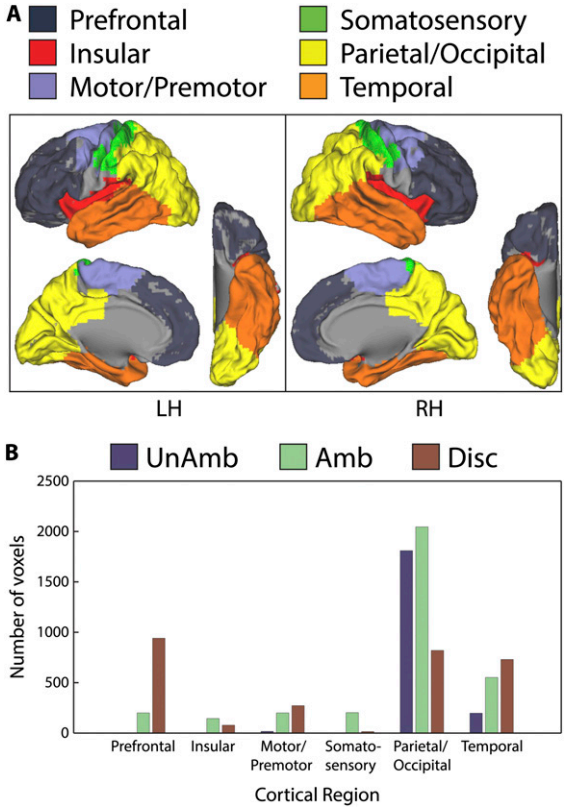

Fig. S5. Cortical parcellation summary of MVPA results. (*A*) The locations of the six cortical regions. (*B*) The number of voxels within each cortical region that were significant in the searchlight MVPA group analysis ($P < 0.05$, corrected for multiple comparisons), for each condition separately. Results are averaged across Necker cube and Rubin face-vase stimuli.

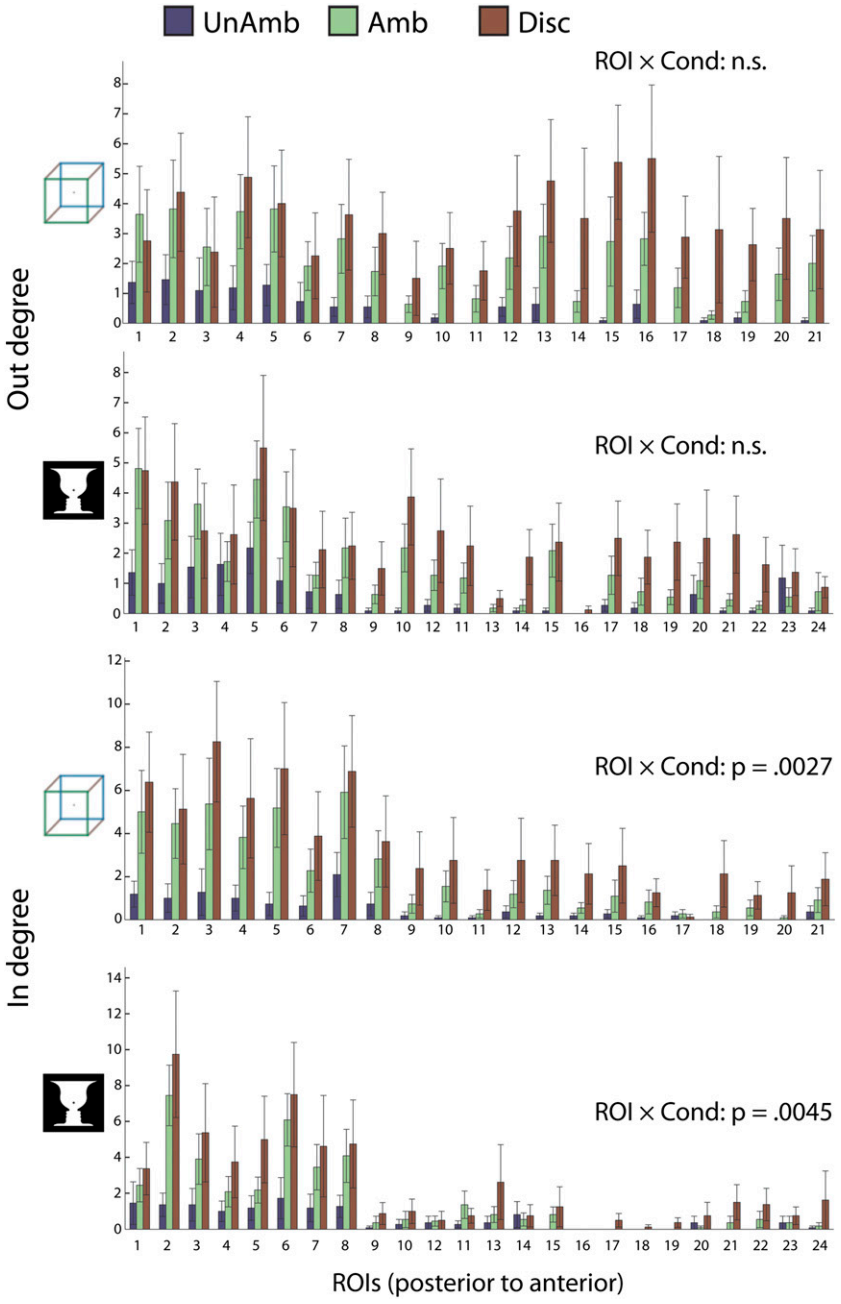

Fig. S6. Out-degree and in-degree for each ROI in each condition, from the unweighted analysis using a threshold of 50% (as in Fig. 5). Format is the same as in Fig. 5.

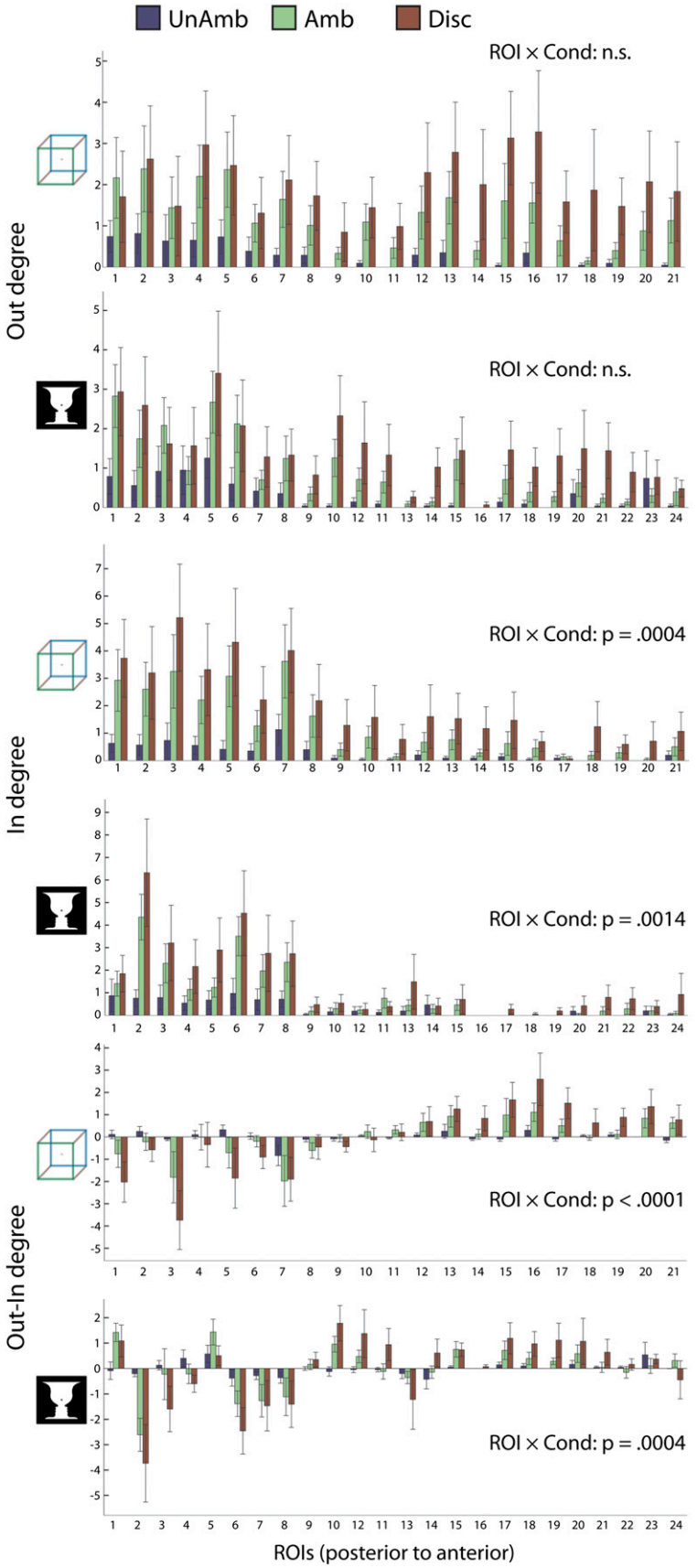

Fig. S7. Format is the same as in Fig. 5 and Fig. S6. Results from weighted analysis (threshold 50%).

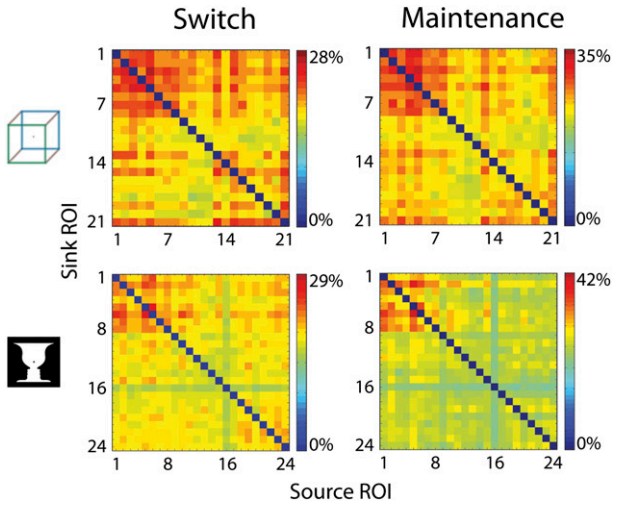

Fig. S8. Percentage of significant voxel pairs for each ROI pair in each direction, during perceptual switch (*Left*) and perceptual maintenance (*Right*) in the Amb condition. The direction of GC influence is from the source ROI to the sink ROI. Format is the same as Fig. 4*B*.

Table S1. Detailed information of all Regions of Interest (ROIs) used in Granger causality analysis

ROI #	Abbreviation	Full name	Talairach coordinates (mm)			No. of voxels
			x	y	z	
Necker cube						
1	R OP	Right occipital pole	30.8	−100.8	−9.2	86
2	L OP	Left occipital pole	−31.2	−100.1	3.3	52
3	R cerebellum	Right cerebellum	7.0	−95.0	−33.5	45
4	R V3/V3A	Right V3/V3A	20.3	−93.1	31.9	70
5	V1	V1	−5.0	−91.8	1.1	463
6	L FG	Left fusiform gyrus	−30.2	−75.7	−8.6	373
7	R LO	Right lateral occipital	57.0	−73.4	−4.5	38
8	R FG	R fusiform gyrus	27.3	−70.8	−9.6	402
9	R IPS	Right intraparietal sulcus	43.9	−59.2	38.1	44
10	L PoCes	Left postcentral sulcus	−27.1	−35.7	57.7	116
11	L Ces	Left central sulcus	−22.7	−33.7	56.4	124
12	R SMC	Right sensorimotor cortex	27.6	−27.8	60.4	84
13	R post Ins	Right posterior insular	50.7	−25.8	15.9	61
14	L TP	Left temporal pole	−63.9	−0.9	−19.8	34
15	R FO	Right frontal operculum	64.8	0.2	3.0	49
16	R PMC	Right premotor cortex	47.6	9.8	32.4	29
17	R IFG	Right inferior frontal gyrus	49.3	13.4	18.6	47
18	L post OFC	Left posterior orbitofrontal cortex	−21.2	18.3	−23.7	20
19	R ant Ins/OFC	Right anterior insular and orbitofrontal cortex	31.3	30.8	−2.1	69
20	L ant OFC	Left anterior orbitofrontal cortex	−18.2	40.5	−16.9	16
21	mPFC	Medial prefrontal cortex	2.0	44.7	7.1	42
Rubin face–vase						
1	L V3/V3A	Left V3/V3A	−23.9	−97.5	25.3	43
2	L post cerebellum	Left posterior cerebellum	−6.3	−97.1	−28.9	42
3	R LO	Right lateral occipital	48.6	−87.6	−4.1	60
4	L V2	Left V2	−24.3	−85.6	−6.5	415
5	R V1	Right V1	9.5	−84.3	4.5	498
6	R precun	Right precuneus	6.8	−76.8	42.5	41
7	R cerebellum	Right cerebellum	34.3	−69.9	−33.0	32
8	R FG	Right fusiform gyrus	42.0	−66.8	−14.5	292
9	L MT+	Left middle temporal area	−44.6	−64.5	0.3	170
10	R pMTG	Right posterior medial temporal gyrus	60.3	−62.6	13.4	46
11	L TPJ	Left temporoparietal junction	−54.8	−56.4	22.8	33
12	L ant cerebellum	Left anterior cerebellum	−35.4	−56.2	−36.4	92
13	R cerebellar vermis	Right cerebellar vermis	10.2	−35.4	−36.2	80
14	R TPJ	Right temporoparietal junction	54.9	−32.6	25.7	33
15	L S1	Left somatosensory cortex	−59.4	−19.2	47.3	51
16	L FEF/MCC	Left frontal eye field and mid–cingulate cortex	−17.9	−16.3	42.4	45
17	L M1	Left motor cortex	−43.5	−9.4	40.9	288
18	R FO	Right frontal operculum	47.6	8.8	16.7	137
19	L ant Ins/FO	Left anterior insular and frontal operculum	−43.4	17.4	−1.1	62
20	L dlPFC	Left dorsolateral prefrontal cortex	−36.5	19.7	33.3	56
21	R IFG	Right inferior frontal gyrus	50.6	30.1	17.0	49
22	R ant Ins/OFC	Right anterior insular and orbitofrontal cortex	26.1	30.2	−3.3	74
23	dmPFC	Dorsal medial prefrontal cortex	0.7	45.0	29.1	65
24	vmPFC	Ventral medial prefrontal cortex	6.9	60.5	−5.2	32

Table S2. Number of voxels in the masks used for the hemodynamic delay control analysis (Fig. 2)

Preferring-voxel groups	Frame -1	Frame 0	Frame 1	Frame 2	Frame 3
Green front	16.9 ± 4.0	15.0 ± 4.0	12.5 ± 3.3	12.5 ± 3.1	15.4 ± 4.1
Blue front	68.0 ± 32.8	52.9 ± 28.5	56.7 ± 33.9	39.5 ± 24.4	23.4 ± 10.4
Face	12.1 ± 5.0	24.4 ± 8.1	19.0 ± 7.7	31.4 ± 12.5	46.2 ± 20.0
Vase	15.0 ± 4.8	15.2 ± 3.7	15.5 ± 3.8	21.6 ± 9.3	29.7 ± 11.6

Values are mean ± SEM across subjects. For details see *SI Methods*, *Control for Hemodynamic Delay*.