Decoding subjectively correct "Yes/No" thoughts in the human brain

Zhi Yang[1,2], Javier Gonzalez-Castillo[2], Zirui Huang[1], Rui Dai[1], Georg Northoff[3], and Peter Bandettini[2]

[1]Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [2]National Institute of Mental Health, Bethesda, MD, United States, [3]University of Otawa, Otawa, ON, Canada

PURPOSE. Multi-variate pattern analysis (MVPA) applied to BOLD fMRI data has proven successful at decoding different aspects of cognitive function (e.g., observed stimuli, presence or absence of memories). In recent years, BOLD fMRI has also permitted basic communication with a subgroup of locked-in syndrome patients using activity patterns for distinct covert tasks (motor imagery vs. spatial navigation) as response signatures[1]. To advance such a fMRI-based brain-computer interface, we focus the current study on examining whether "Yes/No" thoughts in response to binary "Yes/No" questions can be decoded from BOLD-fMRI signals using MVPA and whether they can be decoded regardless of intentions.

METHODS. Two fMRI experiments were conducted in this work. Each experiment consisted of an anatomical scan followed by several functional scans. Acquisition parameters for the functional scans are summarized in Table 1. In Exp. 1, we used a task-cueing paradigm and a MVPA searchlight approach to investigate if, and where and when in the brain, subjectively correct "Yes/No" answers can be decoded. Ten subjects participated in this first experiment. Each trial of the paradigm starts with a visual cue (2s) instructing subjects whether they should respond the subsequent question honestly or dishonestly. The cue is followed by a simple common-knowledge question (4s; e.g., Is the Statue of Liberty in Beijing?), a random delay period (2–6s) during which subjects no longer see the question, and a final instruction to provide a motor response (2s) using an MRI compatible response box. The task was split into ten scans, each containing 32 trials. After basic pre-processing (slice time correction, motion correction, conversion to signal-percent-change), spatiotemporal patterns of hemodynamic response associated with subjectively "Yes" and "No" thoughts were estimated (AFNI *3dDeconvolve*) for each scan of each subject. These patterns were subsequently input to Gaussian Naïve Bayesian (GNB) classifiers in a searchlight manner (leave-two-out cross validation) across all voxels in the grey matter. We did this analysis separately for all time points in a range of 12s starting at the onset of the intention cue. Group-level *t*-tests were performed to examine the significance of the decoding accuracy against 50%. Areas with accuracy significantly above chance are shown in panel A of Fig. 1. In Exp. 2, we scanned three subjects on a 7T scanner to confirm, with an independent dataset,

Fig. 1

	Scanner	TR	Voxel size	TE	FA	#Slices
Experiment 1	3T	2s	3×3×5	30ms	90°	33
Experiment 2	7T	2s	2×2×2	25ms	50°	54

Table 1

that the regions detected in Exp. 1 indeed contain sufficient information to accurately decode "Yes/No" thoughts. We also used this second experiment to evaluate the effect of increasing temporal signal-to-noise ratio (TSNR) in classification accuracy. The paradigm for Exp. 2 was similar to Exp. 1 except: (1) delay between offset of question display and motor response cue was extended to 8s to allow direct use of individual trial responses (i.e., no regression needed); and 2) which button signaled "Yes" and which one signaled "No" was randomly changed on each trial (i.e., to avoid the possibility of subjects encoding "Yes/No" thoughts as motor responses during the question display and random delay periods). After pre-processing (same as in Exp. 1), we trained GNB classifiers using response patterns generated by averaging an increasing number of randomly selected trials (N_{avg}), which ranged from 2 to 18 trials. We conducted a leave-two-trials-out cross-validation scheme. In each cross-validation iteration for a given ROI and N_{avg} level, we trained classifiers based on the time points 2s, 4s and 6s after question onset in training trials to separately predict the labels of the three time points in the test trials (left-out from training). The final "Yes" or "No" label for a test trial was decided by the vote of its three predicted labels (one per time point). This analysis was conducted on each subject separately.

RESULTS. We observed nine regions (ROIs) with decoding accuracy significantly above chance level in Exp. 1 at three different time points after question onset (Fig 1.A). These were: left parahippocampal gyrus (2s after question onset), left supra-marginal gyrus (4s), left middle frontal gyrus (4s), right superior temporal gyrus (4s), left and right inferior frontal gyrus (4s), left medial frontal gyrus (4s), left superior temporal gyrus (6s), and left middle frontal gyrus (6s). Panels B-D in Fig. 1 show how decoding accuracy changes with increasing numbers of averaged trials (N_{avg}) for each ROI (colored lines). Notably, the left middle frontal gyrus ROI in both 4s and 6s, and the left medial frontal gyrus ROI in 4s after question onset (in Exp. 1) showed consistent increasing trend in all three subjects; reaching over 85% accuracy for N_{avg}=18 in all subjects. As controls, we also include results for two additional ROIs not shown to accurately decode "Yes/No" thoughts in Experment 1 (right insular cortex and left primary visual cortex). Moreover, we also conducted a permutation analysis for each ROI by randomizing labels during classifier training. Results from this permutation analysis are shown in grey color for each ROI. In all control cases, the decoding accuracy did not increase with N_{avg}. Semi-transparent bands that accompany each line in panels B-D indicate 95% confidence levels of the accuracy estimation.

CONCLUSIONS. With MVPA searchlight, we identified a set of brain regions showing group-level above-chance accuracy in decoding the subjectively corrected "Yes/No" answers to binary questions. Our results from 7T scans further verified that three regions can be used to robustly decode the "Yes/No" answers (regardless of intentions) with high accuracy, given sufficiently high TSNR (which can be achieved by means of ultra-high field scanners and trial-averaging). These findings suggest that subjectively correct answers can be accurately decoded with fMRI in the spatial-temporal patterns of prefrontal cortex, providing a basis for fMRI-based brain-computer interface.

REFERENCES

1. Owen A, Coleman M, Boly M, et al.et al. Detecting Awareness in the Vegetative State. Science 2006;313(5792):1402.