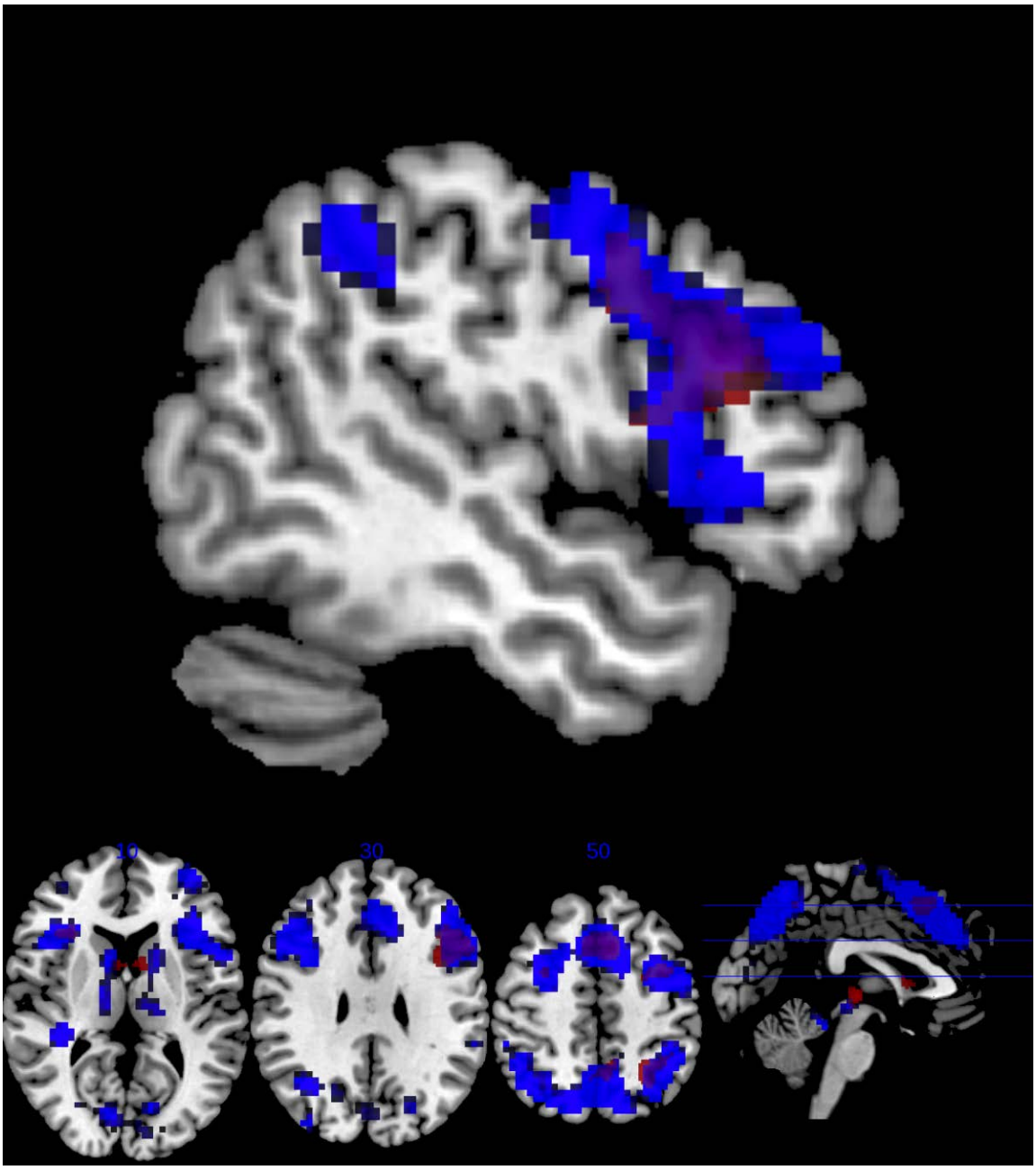
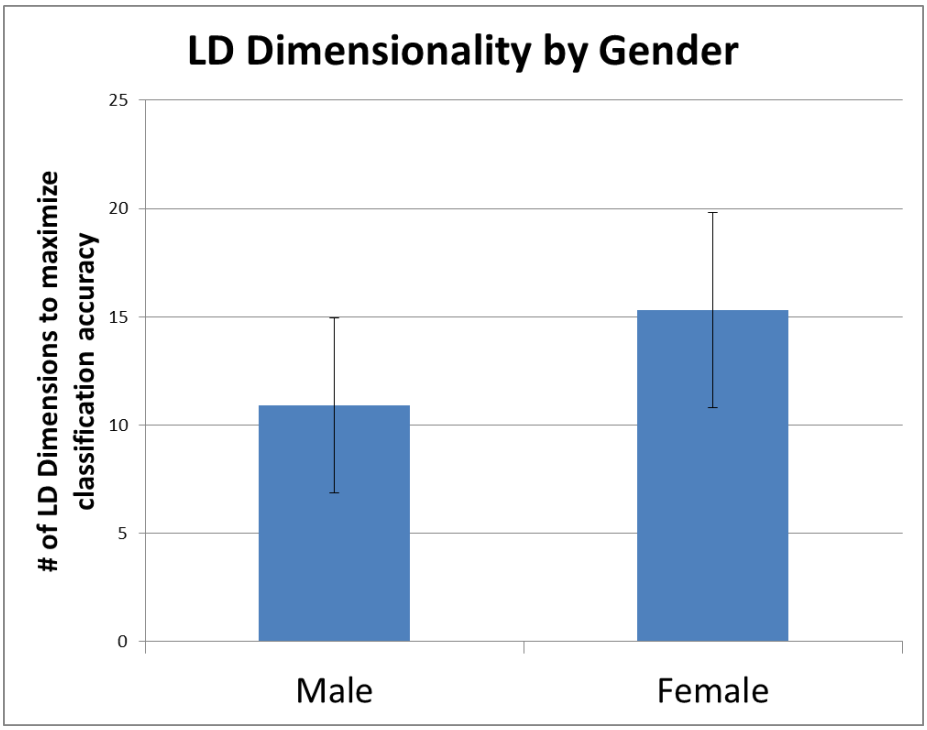
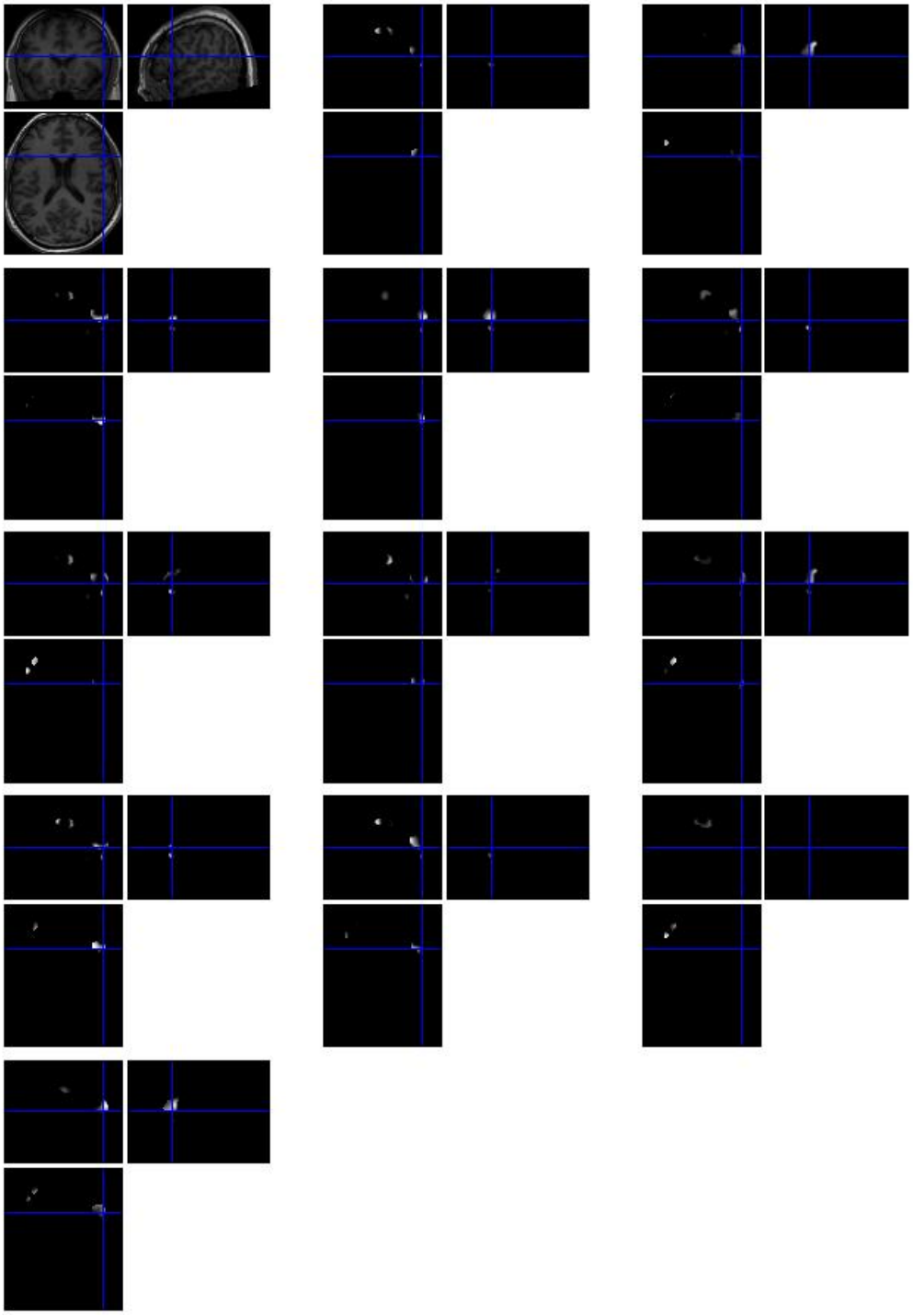


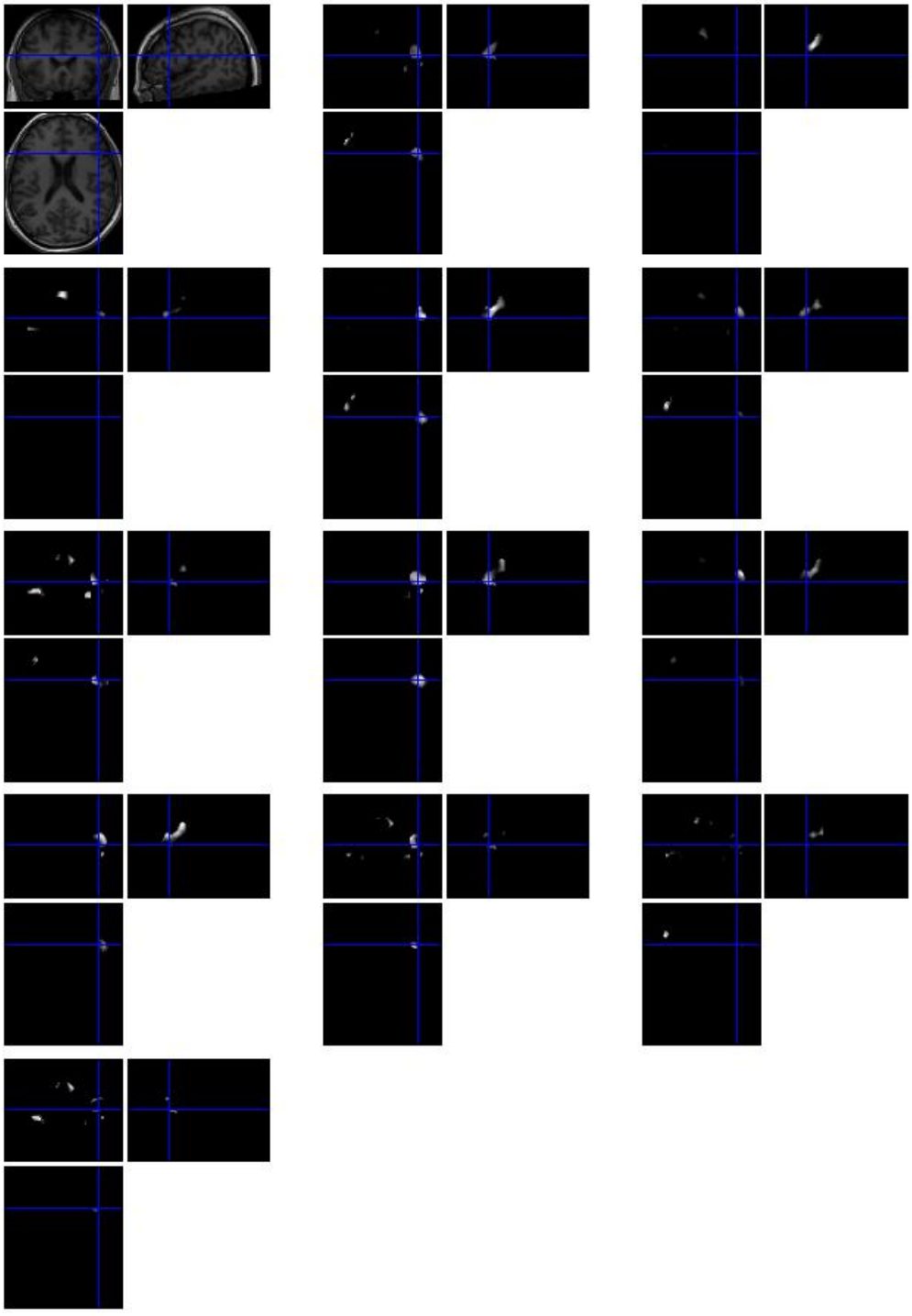
Supplementary Figure S1. Overlapping activation between our current results in red, and previous results. The feature-selected network that was used has considerable similarity/overlap with results from previous research using a similar task[7,8]. This overlap suggests that the network that we have used for feature selection maps onto and replicates previous work using a similar task. The activation from[7] is more robust because there were more than twice as many trials and 6 more participants. These results are presented in radiological coordinates, where right on the display is the left side on the participant. The left inferior frontal gyrus is highlighted above.



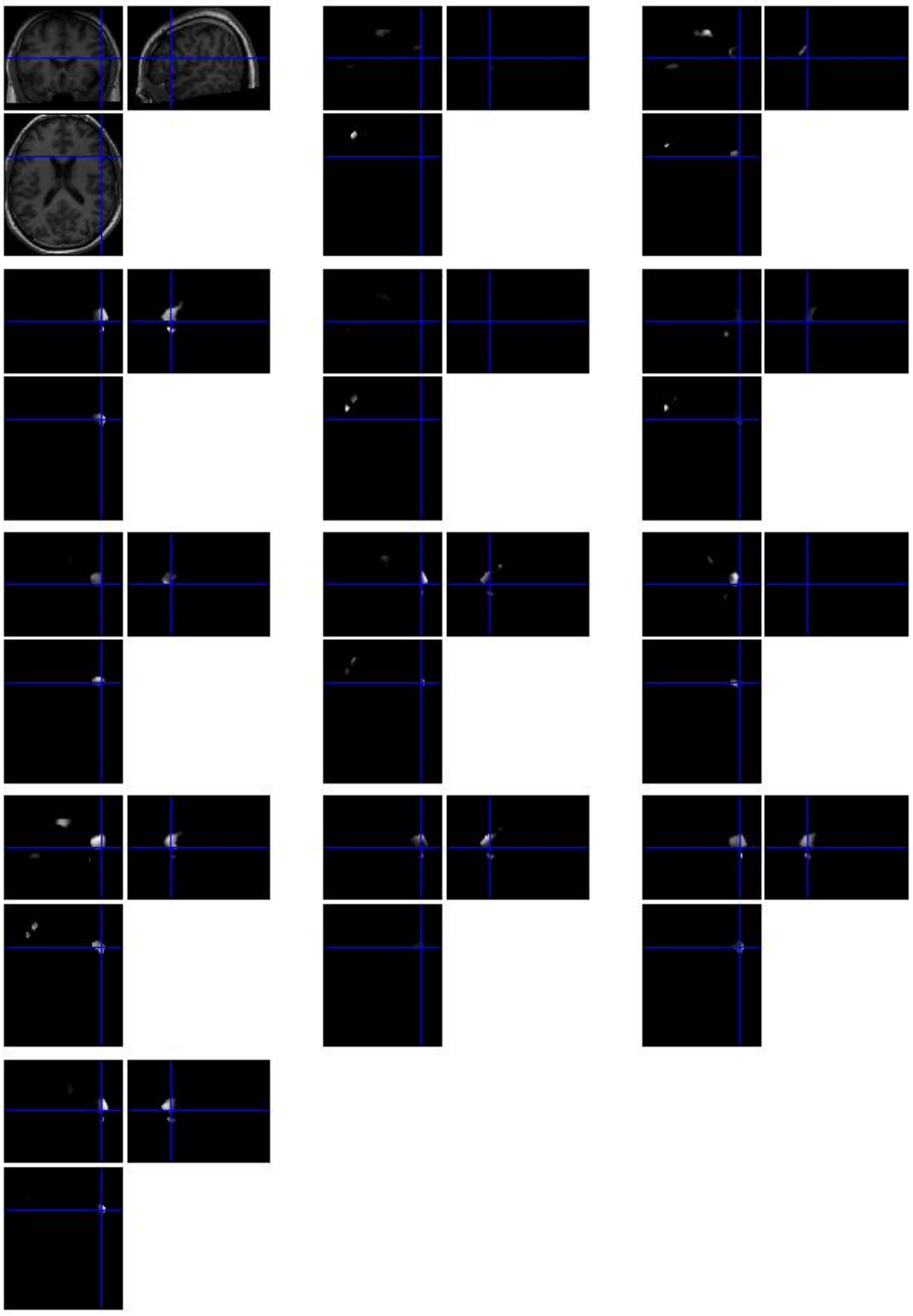
Supplementary Figure S2. LD dimensionality to maximize classification accuracy between lure vs. control trials for male and female participants. Error bars represent standard errors of the mean.



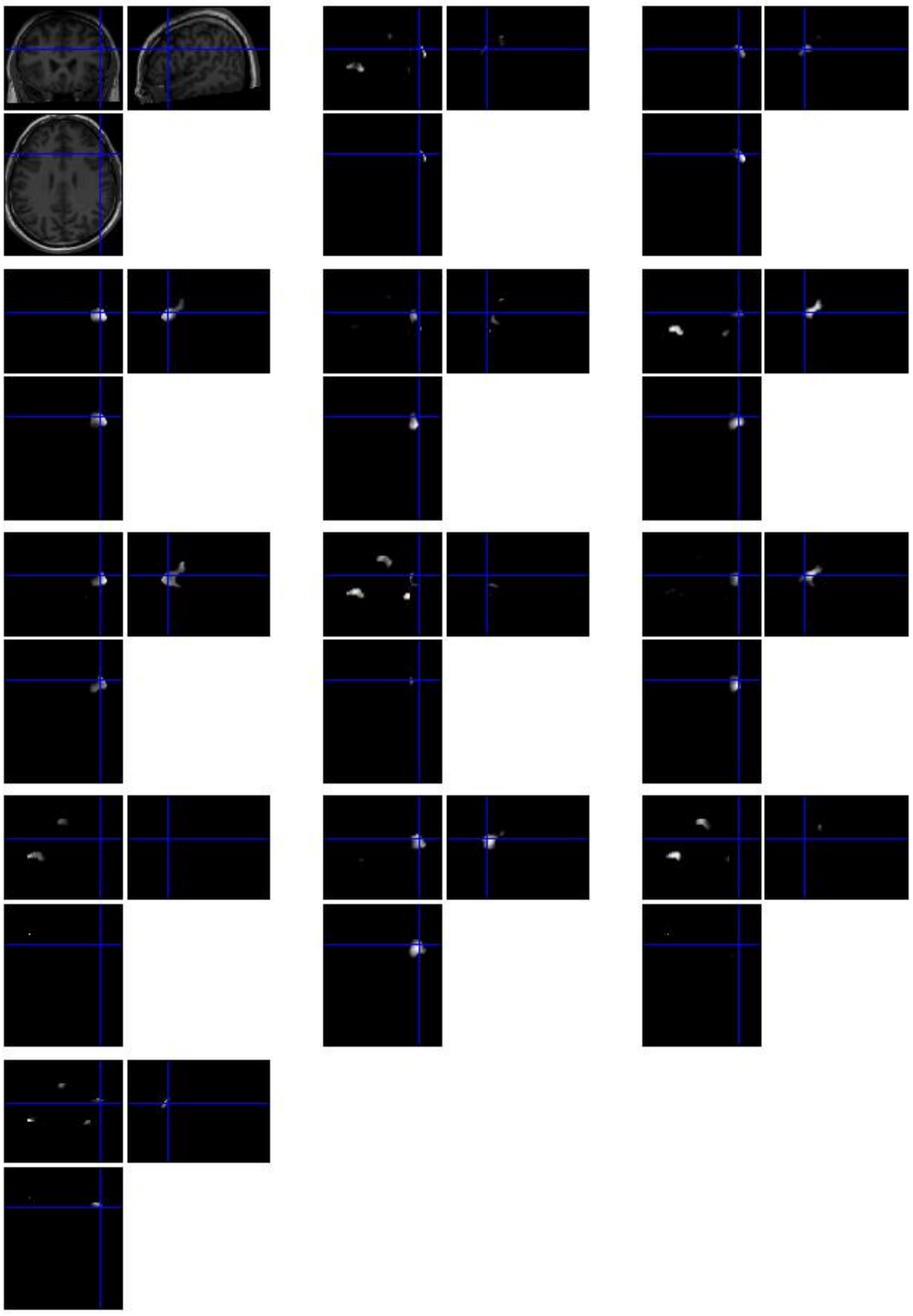
Supplementary Figure S3. High Delay PC 1 for each participant. Results are displayed in radiological coordinates, with a single participant's anatomical image to orient the PCs anatomically.



Supplementary Figure S4. Low Delay PC 1 for each participant. Results are displayed in radiological coordinates, with a single participant's anatomical image to orient the PCs anatomically.



Supplementary Figure S5. High Delay PC 2 for each participant. Results are displayed in radiological coordinates, with a single participant's anatomical image to orient the PCs anatomically.



Supplementary Figure S6. Low Delay PC 2 for each participant. Results are displayed in radiological coordinates, with a single participant's anatomical image to orient the PCs anatomically.

Supplementary Table S1. BOLD activation magnitudes for low and high delayers

Area	Cluster Size (# Voxels)	Peak T-value	Location of Peak x,y,z (in mm)
High > Low			
Left subcallosal Gyrus/BA 25	19	3.6	-6, 12, -10
Right Inferior Frontal Gyrus/BA 47	24	3.43	36, 18, -16
Left Superior Temporal Gyrus/BA 38	30	3.33	-38, 12, -18
Low > High			
Left Cerebellum	21	3.54	-10, -54, -6

Activation differences for High > Low delay and Low > High delay for the Lure – Control contrast, a statistical threshold of $p < .005$ for 10 contiguous voxels was employed.

Supplementary Table S2 Participant demographic information

Delay Group	N	Number of Females	Age	Centered Delay Score from age 4 (*M* ± *SD*)	Raven's Progressive Matrices (*M* ± *SD*)
High Delaying	12	9	44.7 ± 2	279 ± 145.7	24.92 ± 5.48
Low Delaying	12	4	44.3 ± 1.76	-210 ± 144.9	23 ± 4.17[1]

[1] We were missing Raven's data from one low-delay participant

Supplementary Methods

Additional Participant Information

When our participants were adolescents, participants' self-control was assessed by parental ratings on a modified version of the California Child Q-set (e.g., "[My child] is planful", "[My child] thinks ahead"). When they were in their 20s and 30s, participants provided self-reports ratings on the same California Child Q-set items. (e.g., "I am planful; I think ahead", "I am persistent in activities; do not give up easily").

Therefore, the participants in the present study comprise two distinct groups: those for whom we had consistent evidence of high self-control ability (i.e., those who were above average in their self-control abilities from preschool through their 30s), and those for whom we had consistent evidence of relatively low self-control ability (i.e., those who were below average in their self-control from preschool to their 30s). We refer to these groups as "high delayers" and "low delayers," respectively. Other than differences in self-control abilities, these samples were approximately equivalent (e.g., they did not differ in current IQ as measured with Raven's Progressive Matrices, $t(21) = 0.94$, n.s.). We received institutional review board approval from Stanford University and all participants provided consent before testing and scanning. All participants were scanned at the Lucas Center for Imaging at Stanford University. Group demographics are presented in Supplementary Table S2. Centered Delay Scores at age 4 were calculated by the time spent delaying a reward at age 4 years [see ref. [3] for review] which was recorded in seconds and normalized by calculating a difference score of the participant's

delay time relative to the average delay time of children in the same experimental condition [see ref. [24] for details]

fMRI Acquisition Parameters

Participants were scanned with a General Electric Signa 3.0T fMRI scanner (General Electric Medical Systems, Milwaukee, WI) with an 8-channel head coil. A high-resolution, T1-weighted anatomical spoiled gradient sequence ([SPGR] 256 x 256 in-plane resolution, 240-mm field of view [FOV], 136 x 1.2-mm axial slices) was used to acquire an anatomical scan for each subject for transformation and localization of data to Talairach grid space. A spiral in/out T2*-weighted sequence[25] was used to acquire five runs of functional data (TR = 2000ms, echo time = 30, flip angle = 90, skip 0, 64 x 64 matrix) with 31 4mm slices per volume.

fMRI Preprocessing and GLM Analysis Parameters

All functional data were de-spiked with AFNI's spike correction algorithm. Functional images were then corrected for differences in slice timing using 4-point sinc-interpolation in SPM 5[26] and corrected for head movement using MCFLIRT[27]. The high-resolution T1 SPGR anatomical images were co-registered to the first functional scan using a structural overlay in functional space as an intermediate. Afterwards, the images were segmented using SPM5 (Wellcome Department of Cognitive Neurology, London) to separate gray and white matter voxels using the International Consortium of Brain Mapping (ICBM) tissue probability maps; affine normalization parameters were calculated from those maps, which were in standard MNI

space. These parameters were then applied to the functional images to normalize all participants' brains into the same space with a resolution of 2×2×2 mm. Functional images were then spatially smoothed with a Gaussian kernel of 8mm.

Functional images were entered into a general linear model in SPM5 in which each of the 3 different probe types (yes, lure, and control) were modeled with the canonical hemodynamic response function (HRF). Error trials, the stimulus display, 10 seconds of baseline fixation at the beginning and end of each run, and the forget cue were modeled separately, but were not analyzed. Furthermore, 24 motion parameters were added into our model as regressors, which included the linear, squared, derivative, and squared derivative of the six rigid-body movement parameters[22]. In addition to these movement parameters, heart rate and two respiratory physiological variables were also added to the analysis as regressors to covary out these variables.

Whole brain voxel-wise analyses were performed using both standard univariate and regression analyses. For all fMRI analyses, a voxel-level threshold of $p < .005$ (uncorrected for multiple comparisons) and a cluster-size threshold of 10 contiguous voxels were employed to reduce and balance type I and type II errors[20] and also to increase the number of voxels in the feature selected network (for the subsequent multivariate analyses) because a voxel may not contribute strongly to a univariate effect, but could contribute to a multivariate effect.

Supplementary References

24. W. Mischel, Y. Shoda, P. K. Peake, *J.Pers.Soc.Psychol.* **54**, 687 (Apr, 1988).
25. G. H. Glover, M. E. Thomason, *Magn. Reson. Med.* **51**, 863 (Apr, 2004).
26. A. V. Oppenheim, R. W. Schafer, J. R. Buck, *Discrete-time signal processing.* (Prentice Hall, Upper Saddle River, NJ, ed. 2nd, 1999).
27. M. Jenkinson, P. Bannister, M. Brady, S. Smith, *Neuroimage* **17**, 825 (Oct, 2002).