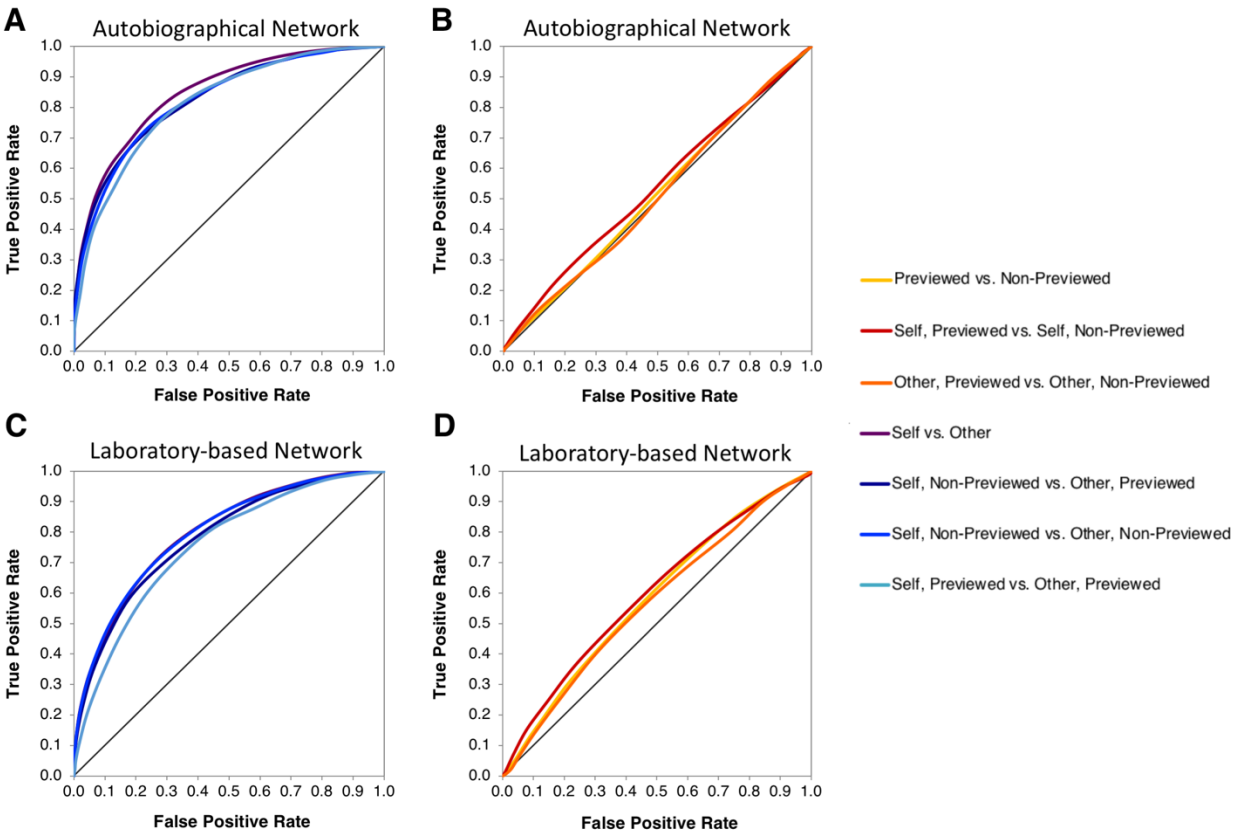
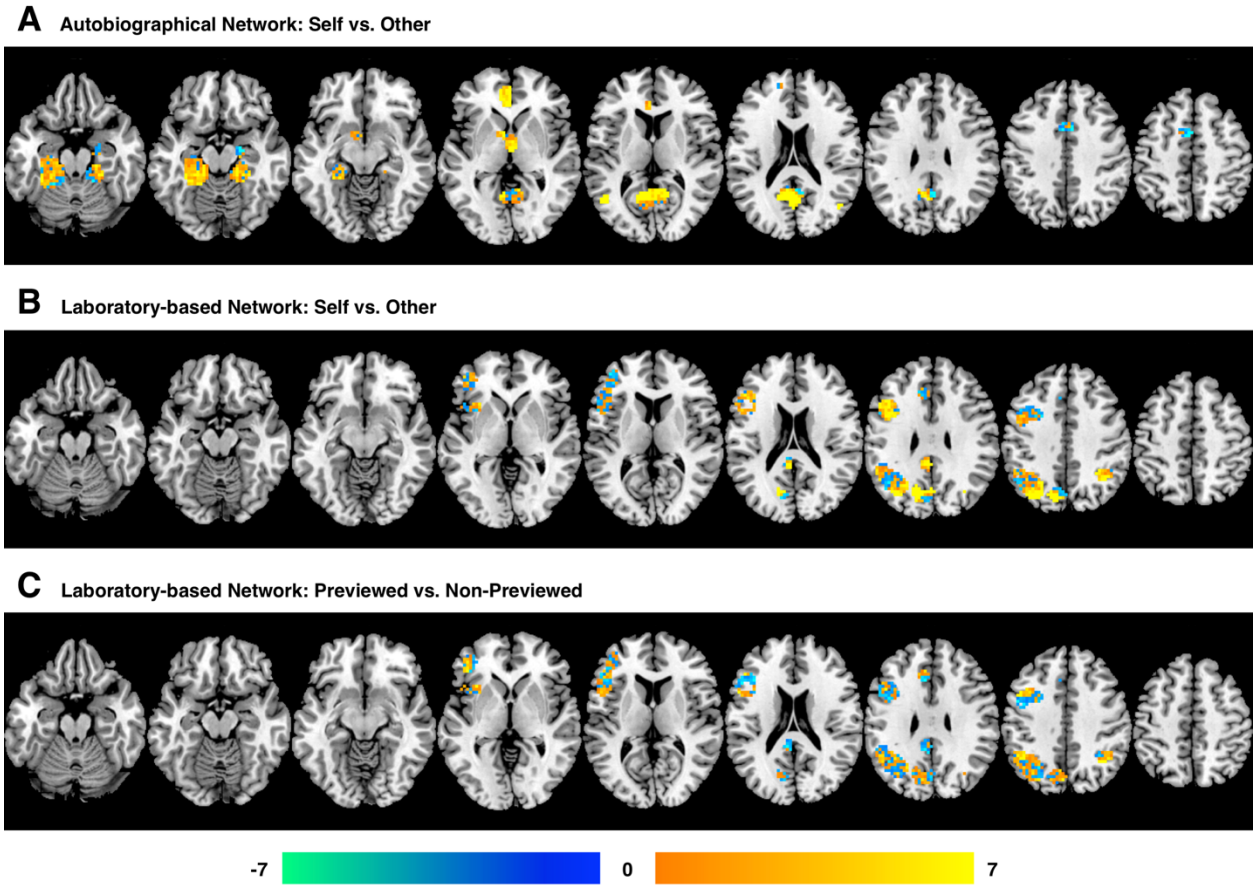


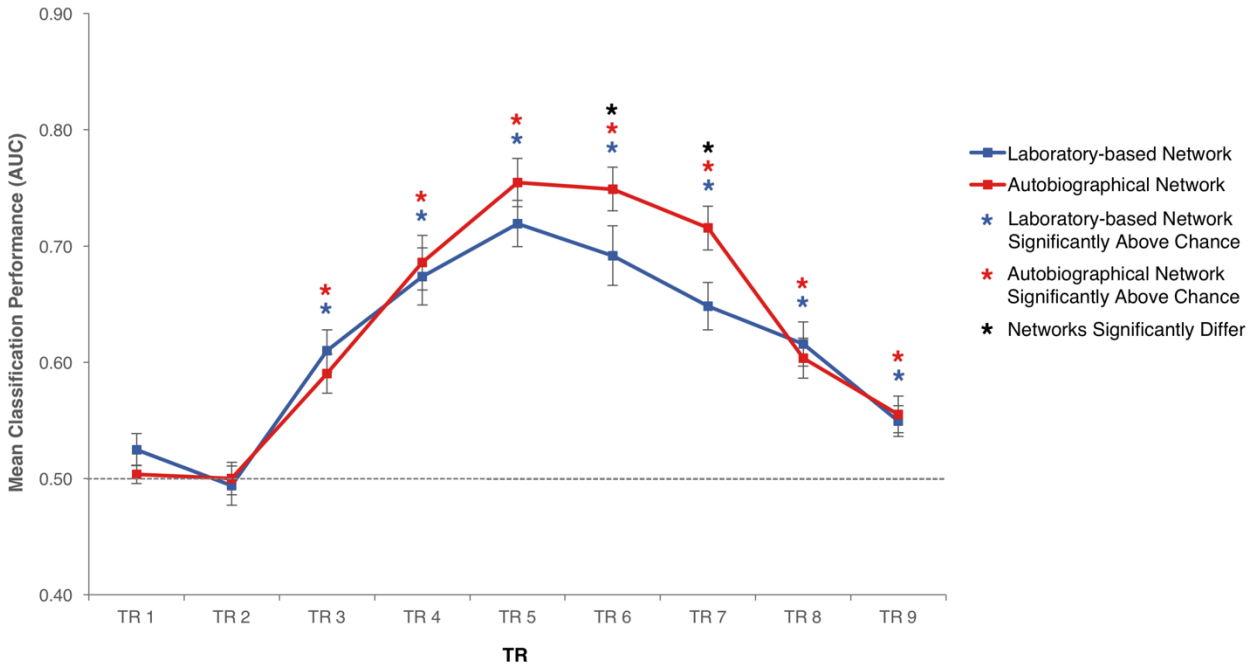
Supplementary Figure 1: Receiver operating characteristic (ROC) curves reflecting classifier performance within the Autobiographical and Laboratory-based Networks. ROC curves (shown averaged across participants) were created based on ranking how strongly each classification output favored one class over the other (e.g., Class A vs. Class B) and graphing the relationship between the true positive rate (the probability of classifying a member of Class A correctly) and false positive rate (the probability of incorrectly classifying a member of Class B as Class A). ROC curves allow for the calculation of area under the curve (AUC), which summarizes the mean accuracy with which a randomly selected pair of trials from each class could be correctly labeled. Decoding performance is shown in cooler colors for Self vs. Other classifications (panel A: Autobiographical Network; panel C: Laboratory-based Network) and warmer colors for Previewed vs. Non-Previewed classifications (panel B: Autobiographical Network; panel D: Laboratory-based Network). Chance performance (AUC = 0.5) is indicated by the diagonal black line.



Supplementary Figure 2. Group-averaged classification importance maps. The Self vs. Other classification maps are shown for the Autobiographical Network (panel A) as well as the Laboratory-based Network (panel B). The Previewed vs. Non-Previewed classification map is shown for the Laboratory-based Network in panel C. The mean importance values are depicted in arbitrary raw units. For each [Class A] vs. [Class B] classification, warmer colors denote voxels in which increased activation biased classifier predictions towards Class A, while cooler colors denote voxels in which increased activation biased classifier predictions towards Class B. Although we depict these group-averaged importance maps for illustrative purposes, we caution readers to avoid over-interpreting them, as classifiers were trained and tested on the BOLD data from individual participants, and thus individual classifiers may have had weight distributions quite different than those depicted here. Furthermore, given the multivariate nature of the classifications, these importance maps should not be considered an exhaustive assessment of which voxels are individually informative about the distinction of interest.

Classification importance maps were derived based upon the logistic regression β weights yielded during each classifier training cycle; these β weights were averaged across each of the 5 cross-validation iterations, and then across each of the 20 rounds of trial-count-balanced classifications. By convention, a positive weight value indicates that a voxel's activity magnitude on each trial was positively correlated with the probability of that trial being from Class A,

whereas a negative weight value indicates the opposite relationship (i.e., increased activity leading to a prediction of Class B). These β weights were then multiplied by each voxel's mean activity level for Class A trials (which, owing to our trial balancing and z-scoring procedure, is always the additive inverse of its mean activity level for Class B trials) and rescaled by a constant factor of 10,000 (to aid in visualization). Voxels with positive values for both activity and weight were given positively signed importance values, voxels with negative activity and weight were given negatively signed importance values, and voxels for which the activity and weight had opposite signs were assigned importance values of zero.



Supplementary Figure 3: Mean classification performance for Self vs. Other using information only from individual TRs. A one-sample, two-tailed *t*-test was used to assess whether mean classification performance was above chance while a paired-samples, two-tailed *t*-test was used to compare mean classification performance between the two networks for every TR (both *t*-tests were corrected for 9 comparisons using Holm-Bonferroni Sequential Correction such that $p < 0.05$). The dashed line represents chance performance. Asterisks indicate significant results: a blue asterisk denotes that decoding performance was significantly above chance for the Laboratory-based Network, a red asterisk indicates that decoding performance was significantly above chance for the Autobiographical Network, and a black asterisk represents a significant difference in mean classification performance between the two networks (in this case, where the Autobiographical Network outperforms the Laboratory-based Network). The two networks were able to reliably decode Self vs. Other status starting at the 3rd TR, with the Autobiographical Network outperforming the Laboratory-based Network at the 6th and 7th TRs. The Autobiographical Network also outperformed the Laboratory-based Network at the 5th TR (p = 0.010, uncorrected), but this did not survive correction.