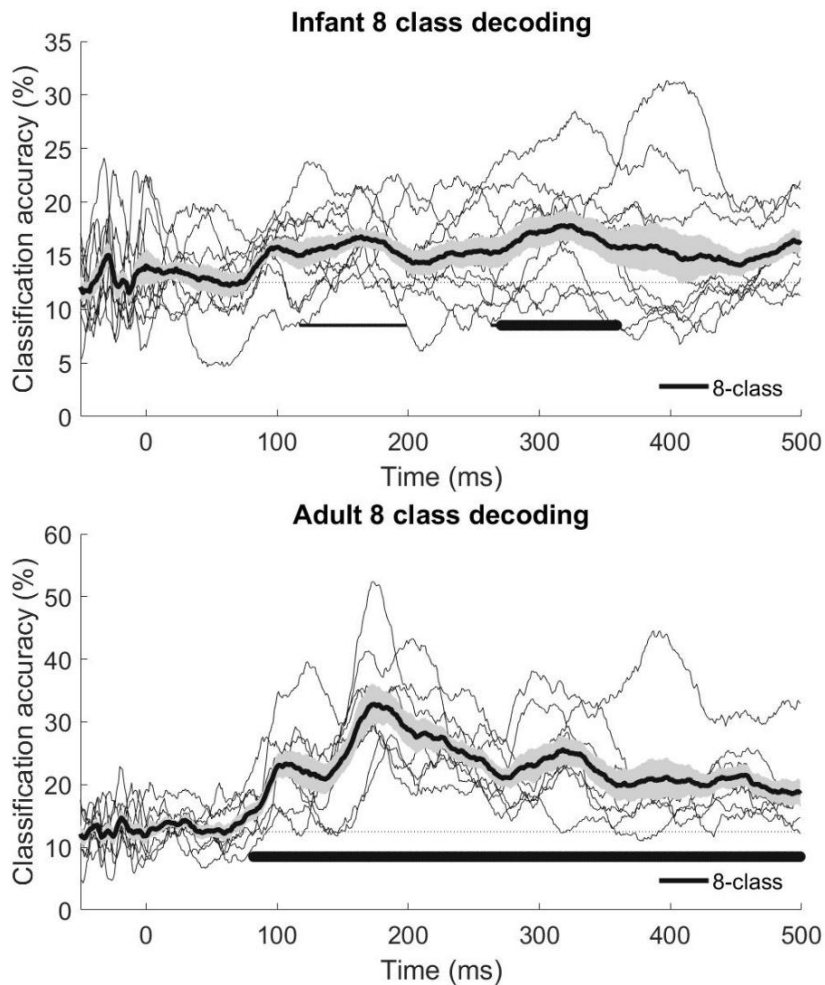
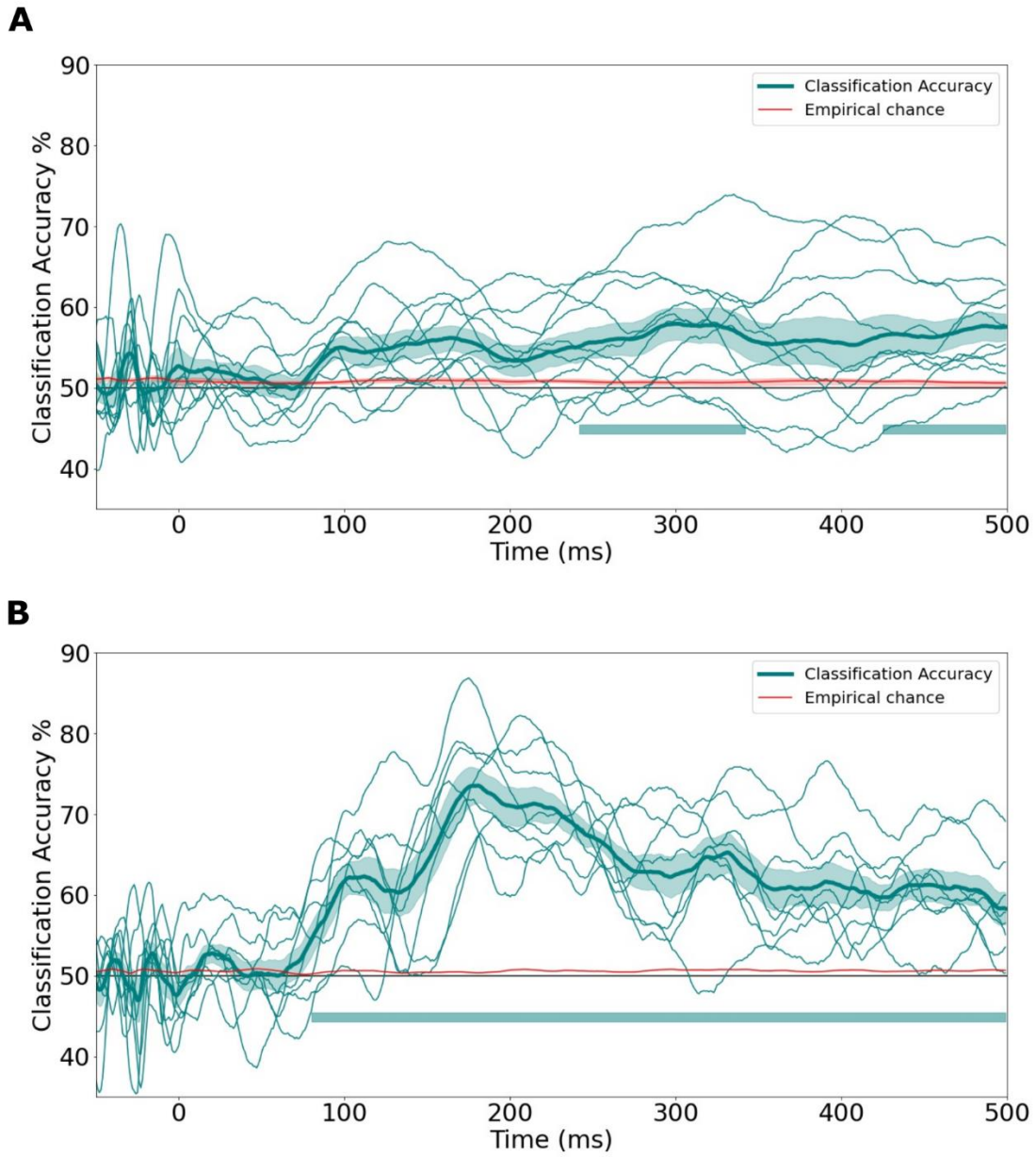


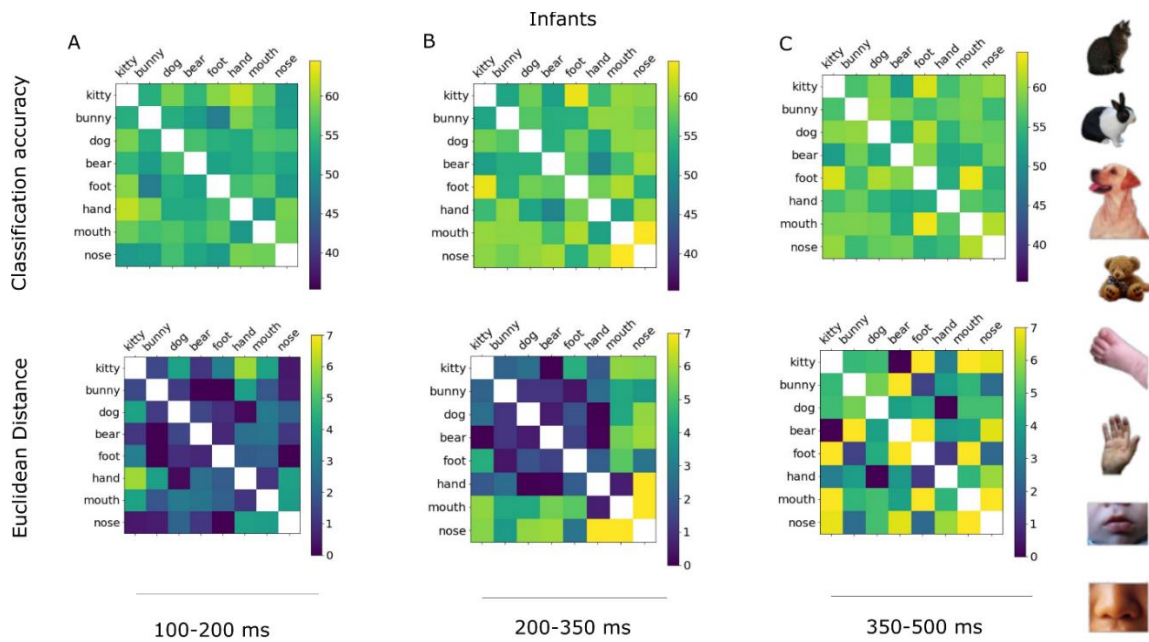
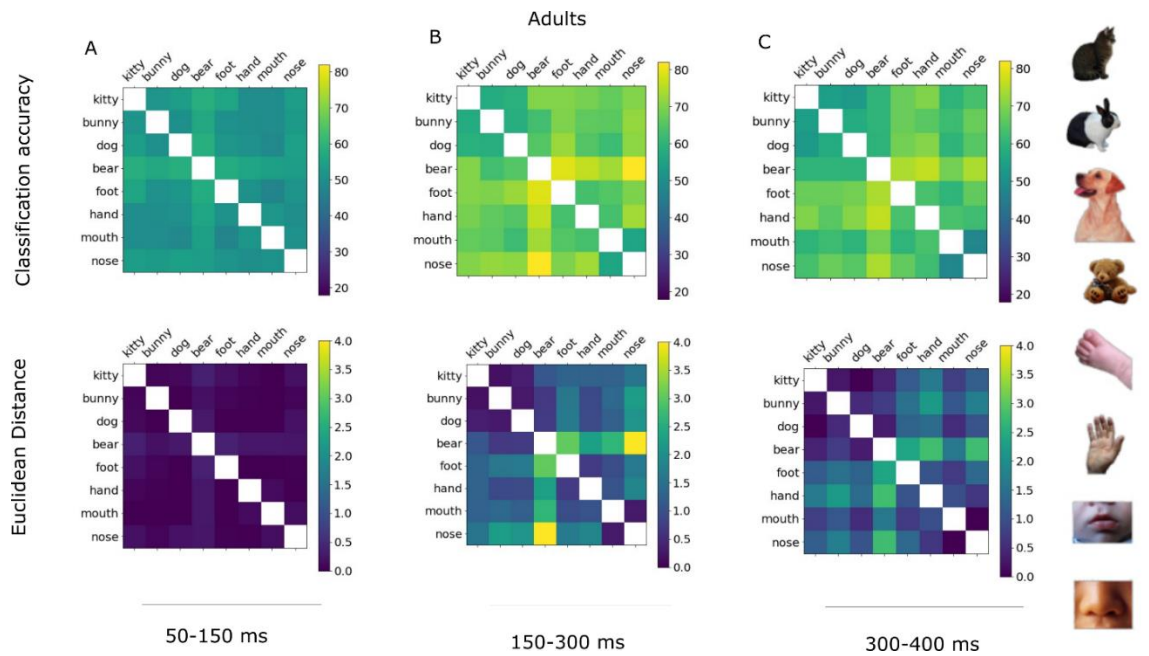
## Supplementary Materials



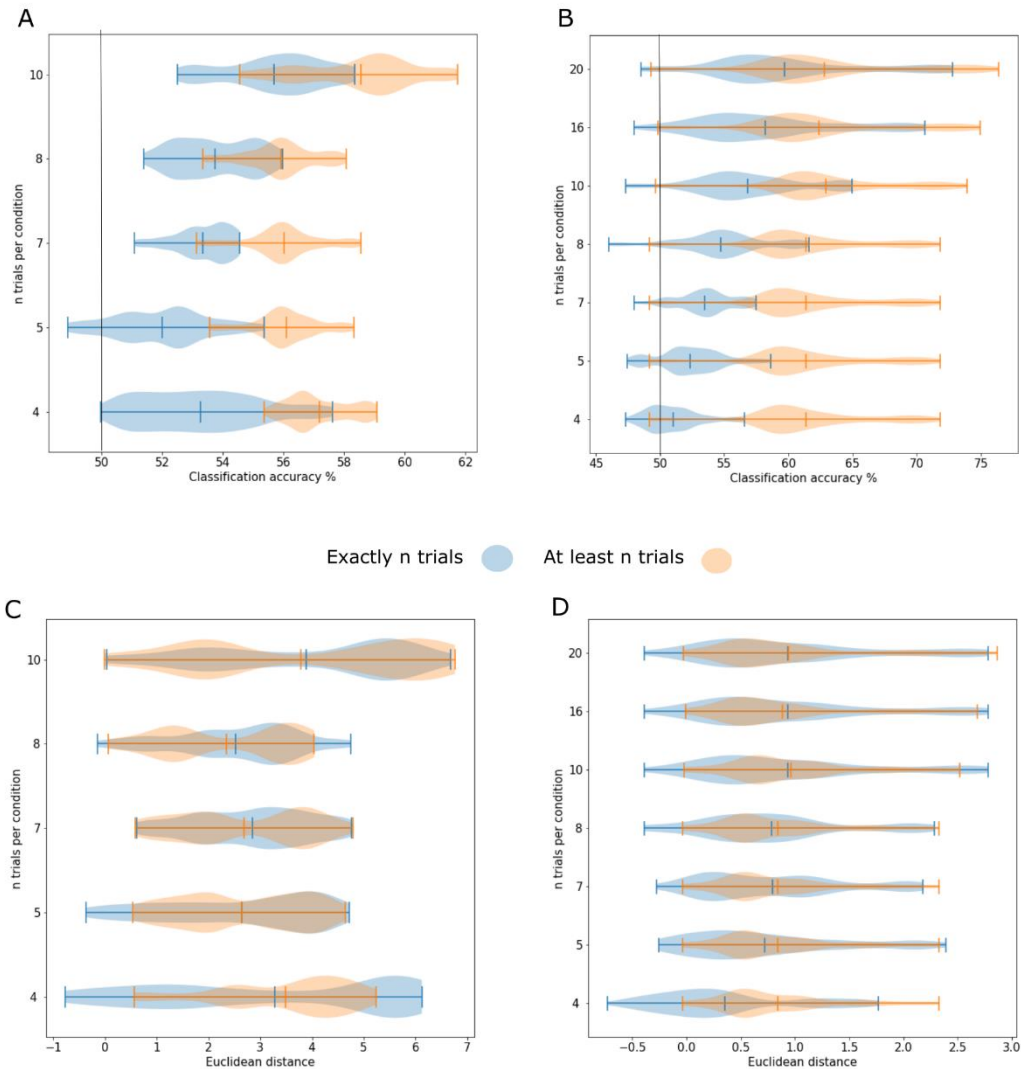
**Figure S1.** Decoding accuracy over the time series within the subset of participants that were included in Bayet et al (2020) (Infants  $n=10$ , Adults  $n=8$ ). Horizontal bars indicate above chance classification accuracy. Average accuracy time series were significantly correlated with those obtained using pairwise classification in both adults (Spearman's  $r = 0.95$ ,  $p < 0.001$ ) and infants (Spearman's  $r = 0.65$ ,  $p < 0.001$ ).



**Figure S2.** Decoding accuracy over the time series within the originally included subset of participant data (**A.** Infants  $n=10$ , **B.** Adults  $n=8$ ). Horizontal bars indicate above chance classification accuracy as compared to an empirical null average at each time point. Empirical chance was calculated by running classification on data with randomly permuted labels over 100 permutations.

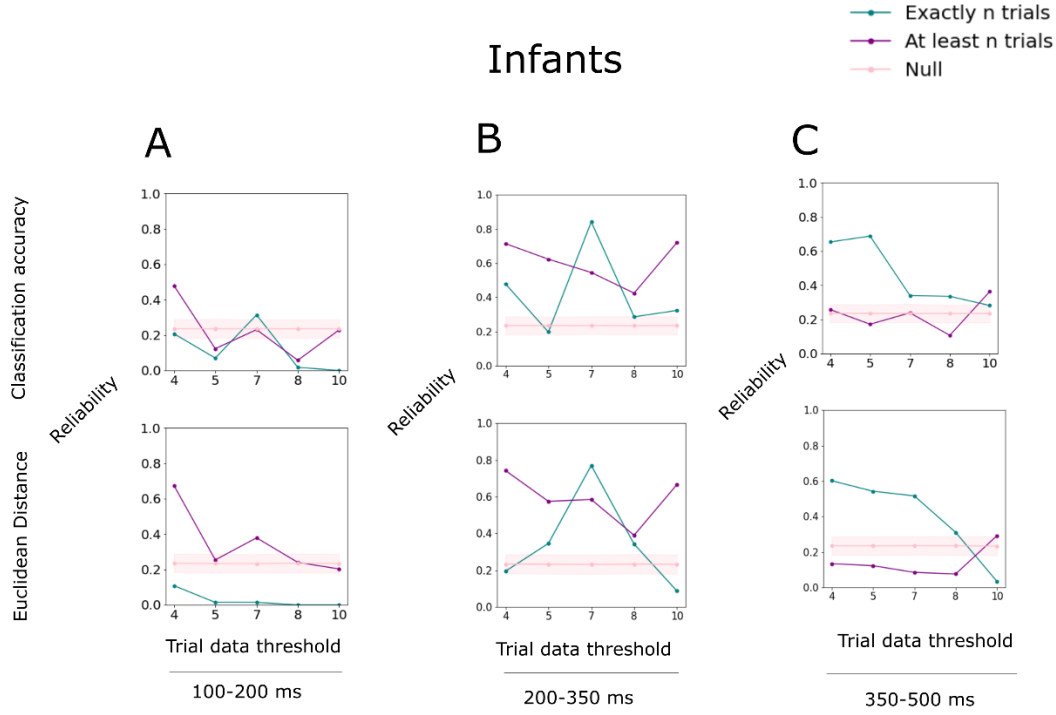


**Figure S3.** Representational Dissimilarity Matrices (RDMs) of pairwise classification accuracy and cross validated Euclidean distance for the subsets of adults (n=8) and infants (n=15) with highest overall RDM reliability. RDMs calculated in the time windows during which classification accuracy rises above chance (**A**), during the window of highest classification accuracy (**B**) and following the window of highest classification accuracy (**C**).

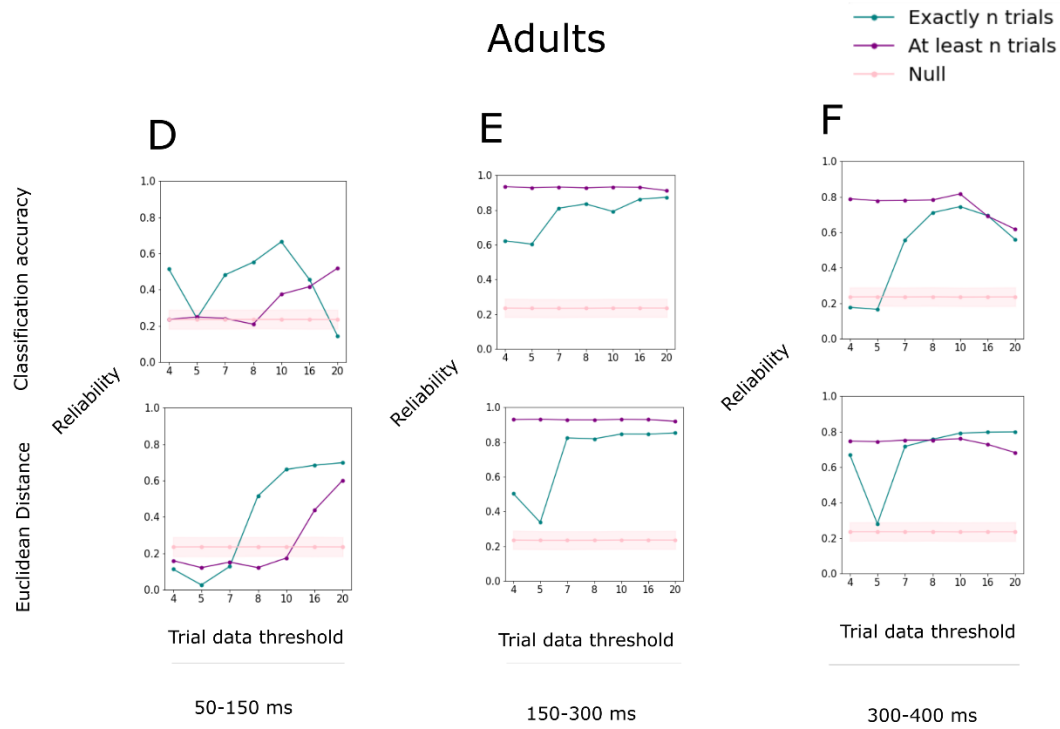


**Figure S4.** Average classification accuracy at different trial thresholds with (**A**) infants (time window 100-500 ms) and (**B**) adults (time window 50-500 ms) and Euclidean distance with (**C**) infants (time window 100-500 ms) and (**D**) adults (time window 50-500 ms). Blue denotes the distribution when the number of trials included was cut off at the threshold, and orange denotes when all trials were included for all participants who met the threshold of trials per condition.

## Infants



## Adults



**Figure S5.** Average split-half reliability of the group-level Representational Dissimilarity Matrices of both classification accuracy and Euclidean distance obtained at each trial number threshold, with corresponding average and 2.5-97.5 percentiles of the null split-half noise ceiling calculated in the time windows during which classification accuracy rises above chance (Infants: **A**, Adults: **D**), during the window of highest classification accuracy (Infants: **B**, Adults: **E**), and following the window of highest classification accuracy (Infants: **C**, Adults: **F**).

	Pre-peak Infant: 100-200 ms Adult: 50-150 ms	Peak Infant: 200-350 ms Adult: 150-300 ms	Post-peak Infant: 350-500 ms Adult: 300-400 ms
Infants, at least n trials	$r = -0.39, p = 0.720$	$r = 0, p = 1$	$r = 0.30, p = 0.720$
Infants, exactly n trials	$r = -0.30, p = 0.720$	$r = 0.79, p = 0.300$	$r = 0.89, p = 0.222$
Adults, at least n trials	<b><math>r = 0.79, p = 0.035 *</math></b>	<b><math>r = 0.91, p = 0.008 *</math></b>	<b><math>r = 0.79, p = 0.035 *</math></b>
Adults, exactly n trials	<b><math>r = 0.93, p = 0.006 *</math></b>	<b><math>r = 1.0, p &lt; 0.001 *</math></b>	<b><math>r = 0.96, p = 0.001 *</math></b>

**Table S1.** Spearman correlations between group average classification accuracy and trial number threshold in all subsets. All  $p$ -values are FDR corrected across time windows and type of subset (i.e., at least vs. exactly n trials).

	Pre-peak Infant: 100-200 ms Adult: 50-150 ms	Peak Infant: 200-350 ms Adult: 150-300 ms	Post-peak Infant: 350-500 ms Adult: 300-400 ms
Infant at least n	$r = -0.89, p = 0.080$	$r = -0.30, p = 0.744$	$r = 0, p = 1$
Infant exactly n	<b><math>r = -0.97, p = 0.015 *</math></b>	$r = -0.30, p = 0.744$	<b><math>r = -0.99, p &lt; 0.001 *</math></b>
Adult at least n	$r = 0.75, p = 0.078$	$r = -0.53, p = 0.252$	$r = -0.36, p = 0.430$
Adult exactly n	<b><math>r = 0.96, p = 0.002 *</math></b>	<b><math>r = 0.89, p = 0.014 *</math></b>	<b><math>r = 0.96, p = 0.002 *</math></b>

**Table S2.** Spearman correlations between the reliability of Euclidean distance RDMs and trial number threshold in all subsets. All  $p$ -values are FDR corrected across time windows and type of subset (i.e., at least vs. exactly n trials).

	Pre-peak Infant: 100-200 ms Adult: 50-150 ms	Peak Infant: 200-350 ms Adult: 150-300 ms	Post-peak Infant: 350-500 ms Adult: 300-400 ms
Infants, at least n trials	$r = -0.49, p = 0.780$	$r = 0, p = 1$	$r = 0.01, p = 1$
Infants, exactly n trials	$r = -0.70, p = 0.564$	$r = -0.09, p = 1$	$r = -0.89, p = 0.222$
Adults, at least n trials	$r = 0.75, p = 0.156$	$r = -0.53, p = 0.030$	$r = -0.50, p = 0.030$
Adults, exactly n trials	$r = -0.25, p = 0.588$	$r = 0.86, p = 0.084$	$r = 0.64, p = 0.240$

**Table S3.** Spearman correlations between the reliability of classification accuracy RDMs and trial number threshold in all subsets. All  $p$ -values are FDR corrected across time windows and type of subset (i.e., at least vs. exactly n trials).