Supplementary Materials for

Assessing the information content of ERP signals in schizophrenia using multivariate decoding methods

Contents

1. Decoding procedure for point-by-point decoding

The point-by-point decoding was largely identical to the time-averaged decoding described in the main manuscript, except that the decoding was conducted separately for each time point in the window of interest and some additional preprocessing was performed using methods derived from previous decoding studies (Bae and Luck, 2018, 2019a, 2019b). Specifically, we low-pass filtered each epoch at 6 Hz (using the EEGLAB *eegfilt*() routine) to remove potential influence of alpha band (8-12 Hz) activity, which can carry spatial information independently of the sustained ERPs (Bae and Luck, 2018) but can also impair decoding by adding substantial trial-by-trial voltage fluctuations. The filtered EEG segments were then reduced to a time window of -200 to 1000 ms and resampled at 50 Hz (20-ms per sample), which increased the analysis speed. This procedure produced 61 time points per EEG epoch. The decoding analysis was performed (independently) at each time point, but the statistical analysis was limited to our a priori time window of 400 to 1000 ms. After the decoding was performed, the averaged decoding accuracy values at each time point were smoothed using a 5-point moving window to minimize noise. As described previously (Bae and Luck, 2019b, 2019a, 2018), the temporal precision resulting from the entire EEG processing and decoding analysis pipeline was approximately ±50 ms.

Note that the 0.1 Hz high-pass filter that was applied as part of our standard EEG preprocessing pipeline was chosen to reduce trial-by-trial variation due to low-frequency drifts while producing minimal temporal distortion of the ERP signals (Kappenman and Luck, 2010; Tanner et al., 2015). High-pass filters can cause signals during one latency range to produce artifactual decoding during earlier and later time ranges, especially with higher cutoff frequencies (van Driel et al., 2019). However, such artifacts should have a minimal impact on the present study given our relatively low high-pass cutoff, and would be expected to impact both PSZ and HCS equivalently.

2. Statistical analysis of point-by-point decoding accuracy

The statistical analyses for the point-by-point decoding were more complex than for the time-average decoding because a separate decoding accuracy value was available for each time point, leading to the need to control for multiple comparisons. To achieve this, we used a nonparametric cluster-based permutation technique that is analogous to the cluster-based mass univariate approach (Maris & Oostenveld, 2007; Groppe et al., 2011) and has previously been applied to ERP decoding data (Bae and Luck, 2019b, 2019a, 2018). First, we will describe how this approach was used to determine which set of time points yielded above-chance decoding, separately for HCS and PSZ. Then we will describe how HCS and PSZ were compared.

The statistical approach for the point-by-point decoding involved three steps. In Step 1, we used one-sample *t* tests at each individual time point to determine whether decoding accuracy at that time point was significantly greater than chance (0.5). We used one-tailed tests here because below-chance decoding accuracy was not meaningful. In Step 2, we found clusters of contiguous time points ($\geqq$ 2 contiguous time points) for which the single-point *t* tests were significant ($p < .05$) and computed the cluster-level *t* mass (the sum of the *t* scores within each cluster). In Step 3, we tested whether a given cluster mass was greater than the mass that would be expected by chance. The null distribution for the *t* mass was determined via permutation testing (see below). This approach controls the Type I error rate at the cluster level, yielding a probability of .05 that one or more clusters would be significant if true decoding accuracy were at chance throughout the time window (Groppe et al., 2011).

The distribution of chance-level cluster *t* mass values was constructed using permutation testing under the null hypothesis that the observed cluster *t* mass values (Step 2) were actually samples from a random classifier that had no information about the side of the stimulus. For each iteration of the permutation procedure, we shuffled the true labels of the side of the stimulus at a time point in a given trial before determining the classifier was correct. This approach was similar to that used in previous EEG decoding studies (Fahrenfort et al., 2017; Foster et al., 2017, 2016, in press), but we permuted at the stage of testing the decoder output rather than at the stage of training the decoder to reduce computing time (Bae and Luck, 2019b, 2019a). To reflect the temporal auto-correlation of the continuous EEG data (Linkenkaer-Hansen et al., 2001) (Linkenkaer-Hansen et al., 2001), we used the same shuffling of target labels for all time points in a given iteration.

Decoding accuracy for a given permutation iteration was computed for the same number of decoding attempts as in the actual decoding analysis (300 values, derived from 3 cross-validations x 2 stimulus sides x 50 iterations). The mean accuracy across these decoding attempts was used as the decoding accuracy at that time point for a given permutation iteration. And as in our main decoding analysis, the time course of the permutation-based decoding accuracy was then smoothed with a 5-point running average filter. We then used the decoding accuracy values to compute the cluster-level *t* mass (with a mass of zero if there were no significant *t* values). If there was more than one cluster of individually significant *t* values, we took the mass of the largest cluster.

This procedure was iterated 1000 times to produce a null distribution for the cluster mass (with a resolution of $p = 10^{-3}$). To compute the *p* value corresponding to a cluster-level *t* mass in the actual data set, we simply found where this value fell within the null distribution. The *p* value for a given cluster was set based on the nearest percentiles of the null distribution (using linear interpolation; we report $p < 10^{-3}$ if the observed mass was greater than all masses in the null distribution). We rejected the null hypothesis and concluded that the decoding was above chance for any observed cluster if the observed *t* mass for that cluster was in the top 95% of the null distribution (one-tailed). Statistical comparisons were limited to time points within 400 ms ~ 1000 ms to match the time window used for the CDA time window in the original paper (Leonard et al., 2013). Separate null distributions were computed for each memory load and each group of participants, and the analyses asked whether decoding accuracy was above chance in any clusters of time points for a given memory load, separately for HCS and PSZ.

The procedure was modified slightly to compare decoding accuracy in HCS and PSZ at each time point. First, we compared decoding accuracy between HCS and PSZ using a two-tailed independent-samples *t* test at each time point (rather than comparing decoding accuracy to chance) prior to computing the cluster *t* mass. Second, the empirical null distribution of the cluster *t* mass was constructed by randomly shuffling the participants' group membership after the single-participant decoding accuracy values were computed rather than by shuffling the target-side labels prior to computing the decoding accuracy. This was based on the null hypothesis that the individuals from the two groups were sampled from the same population. Third, because the number of participants differed between the PSZ and HCS groups, we created permuted groups with 21 values in one group and 24 values in the other.

3. Point-by-Point Decoding Results

We next examined decodability at each individual time point (every 20 ms) during the 400-1000 ms time window. Figure S1 shows point-by-point decoding accuracy at each memory load for HCS and PSZ. Clusters of time points with above-chance decoding (after controlling for multiple comparisons with a permutation test) are shown with gray shading, and clusters of

points that differ significantly between groups are shown with red shading. At memory load one, HCS showed only a hint of above-chance decoding, and no clusters of time points were significantly greater than chance during the time window of interest (Figure S1A). In contrast, PSZ showed above-chance decoding during most of the 400-1000 ms window at memory load one, with a large cluster of significant time points (Figure S1D, gray region). Moreover, there was a cluster of time points that produced significantly greater decoding accuracy for PSZ than for HCS (after controlling for multiple comparisons; Figure S1D, red region). At memory load three, both HCS and PSZ produced above-chance decoding during the entire time window, and there was a cluster of time points that produced significantly greater decoding accuracy for HCS than PSZ (Figure S1B, red region). At memory load five, decoding accuracy was again above chance during the entire time window in both groups, and there was no significant difference between groups (Figures S1C and S1F). These results show how the high temporal resolution of ERPs can be used to isolate the specific time points at which the information content of neural activity differs between PSZ and HCS.

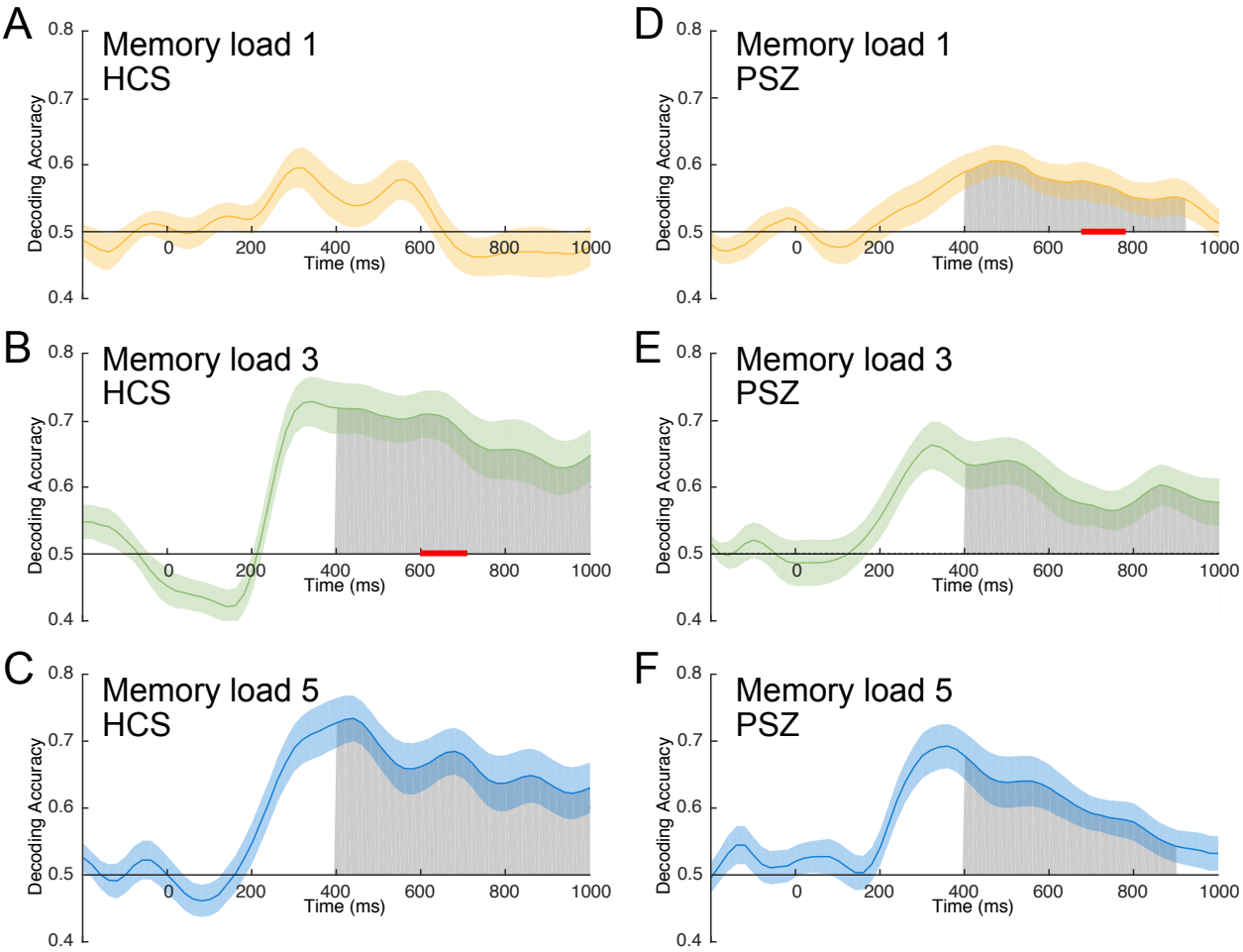

Figure S1. Point-by-point decoding accuracy at each memory load for HCS and PSZ. The colored shading along a given curve represents ±1 S.E.M. Grey areas represent clusters of time points in which decoding was significantly greater than chance (after controlling for multiple comparisons). Red lines indicate clusters of time points that were significantly greater in that group compared to the other group (after controlling for multiple comparisons).

4. CNR analysis using EOG channels

To assess whether EOG voltages created by eye movements differed between PSZ and HCS, we applied our CNR approach to the data from the bipolar vertical and horizontal EOG channels. We computed the CNR using the method described in the main text, and we fed the results into an ANOVA to determine whether any significant differences were present. The main effects were not significant (HCS vs PSZ: $F(1,43) = .931$, $p = .34$; Set size: $F(2,86) = 1.418$, $p = .248$), nor was the two-way interaction significant ($F(2,86) = 0.268$, $p = .765$). To more directly assess the evidence for a lack of difference between PSZ and HCS, we computed Bayes factors comparing PSZ and HCS at each memory load, using the default JZS scaling factor of .707 (Rouder et al., 2009). We found that the data was more consistent with the null hypothesis than with a difference between groups at all three set sizes (Set size 1: $BF_{01} = 2.43$; Set size 3: $BF_{01} = 2.55$; Set size 5: $BF_{01} = 2.47$). These Bayes factors are not strong evidence for the null, but they provide reasonable assurance that the group differences in decoding accuracy were not a consequence of differences in eye movements.

5. References for Supplementary Materials

Bae, G.Y., Luck, S.J., 2019a. Reactivation of previous experiences in a working memory task. Psychol. Sci. 30, 587–595.

Bae, G.Y., Luck, S.J., 2019b. Decoding motion direction using the topography of sustained ERPs and alpha oscillations. NeuroImage 184, 242–255.

Bae, G.Y., Luck, S.J., 2018. Dissociable Decoding of Working Memory and Spatial Attention from EEG Oscillations and Sustained Potentials. J. Neurosci. 38, 409–422.

Fahrenfort, J.J., Grubert, A., Olivers, C.N., Eimer, M., 2017. Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. Sci. Rep. 7, 1886.

Foster, J.J., Bsales, E.M., Jaffe, R.J., Awh, E., 2017. Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. Curr. Biol. 27, 3216–3223.

Foster, J.J., Sutterer, D.W., Serences, J.T., Vogel, E.K., Awh, E., 2016. The topography of alpha-band activity tracks the content of spatial working memory. J. Neurophysiol. 115, 168–177.

Foster, J.J., Sutterer, D.W., Serences, J.T., Vogel, E.K., Awh, E., in press. Alpha-Band Oscillations Enable Spatially and Temporally Resolved Tracking of Covert Spatial Attention. Psychol. Sci.

Kappenman, E.S., Luck, S.J., 2010. The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology 47, 888–904.

Leonard, C.J., Kaiser, S.T., Robinson, B.M., Kappenman, E.S., Hahn, B., Gold, J.M., Luck, S.J., 2013. Toward the neural mechanisms of reduced working memory capacity in schizophrenia. Cereb. Cortex 23, 1582–1592. https://doi.org/10.1093/cercor/bhs148

Linkenkaer-Hansen, K., Nikouline, V.V., Palva, J.M., Ilmoniemi, R.J., 2001. Long-Range Temporal Correlations and Scaling Behavior in Human Brain Oscillations. J. Neurosci. 21, 1370–1377. https://doi.org/10.1523/JNEUROSCI.21-04-01370.2001

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. 16, 225–37. https://doi.org/16/2/225 [pii] 10.3758/PBR.16.2.225

Tanner, D., Morgan-Short, K., Luck, S.J., 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology 52, 997–1009.

van Driel, J., Olivers, C.N.L., Fahrenfort, J.J., 2019. High-pass filtering artifacts in multivariate classification of neural time series data. bioRxiv 530220. https://doi.org/10.1101/530220