Decoding behavioral responses from fMRI without learning behavioral responses from fMRI

Poster No:

3720

Submission Type:

Abstract Submission

Authors:

Joram Soch[1,2], John-Dylan Haynes[1,2,3,4,5,6,7,8]

Institutions:

[1]Bernstein Center for Computational Neuroscience, Berlin, Germany, [2]Berlin Center for Advanced Neuroimaging, Berlin, Germany, [3]Berlin School of Mind and Brain, Berlin, Germany, [4]Clinic for Neurology, Charité – Universitätsmedizin, Berlin, Germany, [5]Department of Psychology, Humboldt University, Berlin, Germany, [6]EXC NeuroCure, Charité – Universitätsmedizin, Berlin, Germany, [7]EXC Science of Intelligence, Technical University, Berlin, Germany, [8]CRC Volition and Cognitive Control, Technical University, Dresden, Germany

Introduction:

The data acquired during a functional magnetic resonance imaging (fMRI) experiment can usually be categorized into experimental design X (e.g. experimental conditions, modulator variables), physiological data Y (i.e. the measured hemodynamic signals) and behavioral data Z (e.g. button presses, stimulus ratings). In multivariate pattern analysis (MVPA) of fMRI data [1,2], button presses are typically decoded by training a classifier to distinguish the recorded responses Z based on the measured data Y (*conventional response decoding*, CRD). Here we show that this can be achieved without constructing an explicit mapping from fMRI signals to behavioral data. In fact, button presses can be equally well decoded when first reconstructing the experimental design X from measured data Y and then predicting behavioral responses Z from the reconstructed design X (*neurobehavioral decoding*, NBD).

Methods:

Decoding from the design [3]: First, we estimate a model of the behavioral data, given the experimental design: $Z = g(X)$. In the case of discrete experimental conditions and response options, this simply means to calculate condition-to-response transition probabilities.
Conventional response decoding: Second, we establish a mapping from measured signals to behavioral data: $Z = h(Y)$. This is used to predict button presses from fMRI signals.
Neurobehavioral decoding: Finally, we establish a mapping from measured signals to ex-perimental design: $X = f^{-1}(Y)$. Together with the behavioral model g, this allows to predict button presses via experimental conditions: $Z = g(f^{-1}(Y))$ (see Figure 1A).
All these models are estimated in a cross-validated fashion, using leave-one-out cross-validation over fMRI recording sessions. Behavioral models (g) were estimated with trial-wise linear regression and decoding

analyses (h, f^{-1}) were performed with trial-wise logistic regression. Because behavioral responses and experimental conditions are best decoded from different parts of the brain, the approaches had to be made comparable. For each approach, the most informative searchlight was selected based on within-sample decoding accuracy and results are reported as out-of-sample decoding accuracies (see Figure 1B).

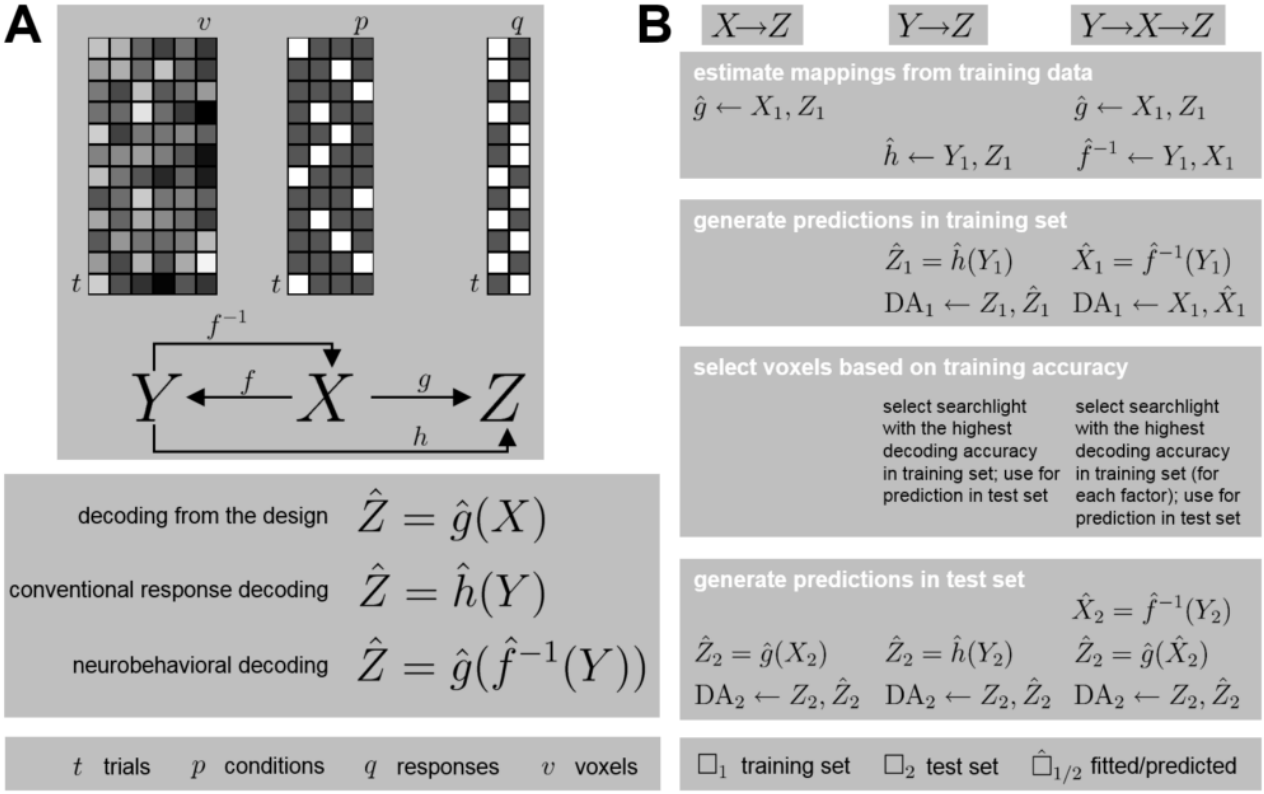

Figure 1. *Neurobehavioral decoding: theory and analysis logic.* **(A)** The analyses operate on the trial-by-voxel matrix Y (fMRI signal amplitudes in searchlight), the trial-by-condition matrix X (experimental design) and the trial-by-response matrix Z (behavioral responses) and assume a neurophysiological model f (from X to Y) as well as a psychobehavioral model g (from X to Z). **(B)** The performances of decoding from the design (left), conventional response decoding (CRD, middle) and neurobehavioral decoding (NBD, right) are evaluated via the cross-validated decoding accuracy (DA). For each decoding analysis (searchlight radius $r = 6\,\mathrm{mm}$), the searchlight with the highest training set DA is selected. Within this searchlight, the test set DA is reported.

Results:

We analyzed the example data set [6] of The Decoding Toolbox (TDT) [4,5]. This experiment (see Figure 2A) used five experimental dimensions (cue, stimulus color and direction, color and direction requiring left button press) and one behavioral dimension (left vs. right button press). In CRD, button presses were directly decoded from the fMRI signal. In NBD, stimulus color and direction were decoded from the fMRI signal; these were then combined with cue and response mapping to yield a reconstructed design; this was then combined with the estimated condition-to-response mapping to yield decoded responses. Overall, we obtain statistically indistinguishable performance of CRD and NBD in both subjects (see Figure 2B). This was irrespective of the searchlight size (3, 6, 9 mm) used for decoding analyses.

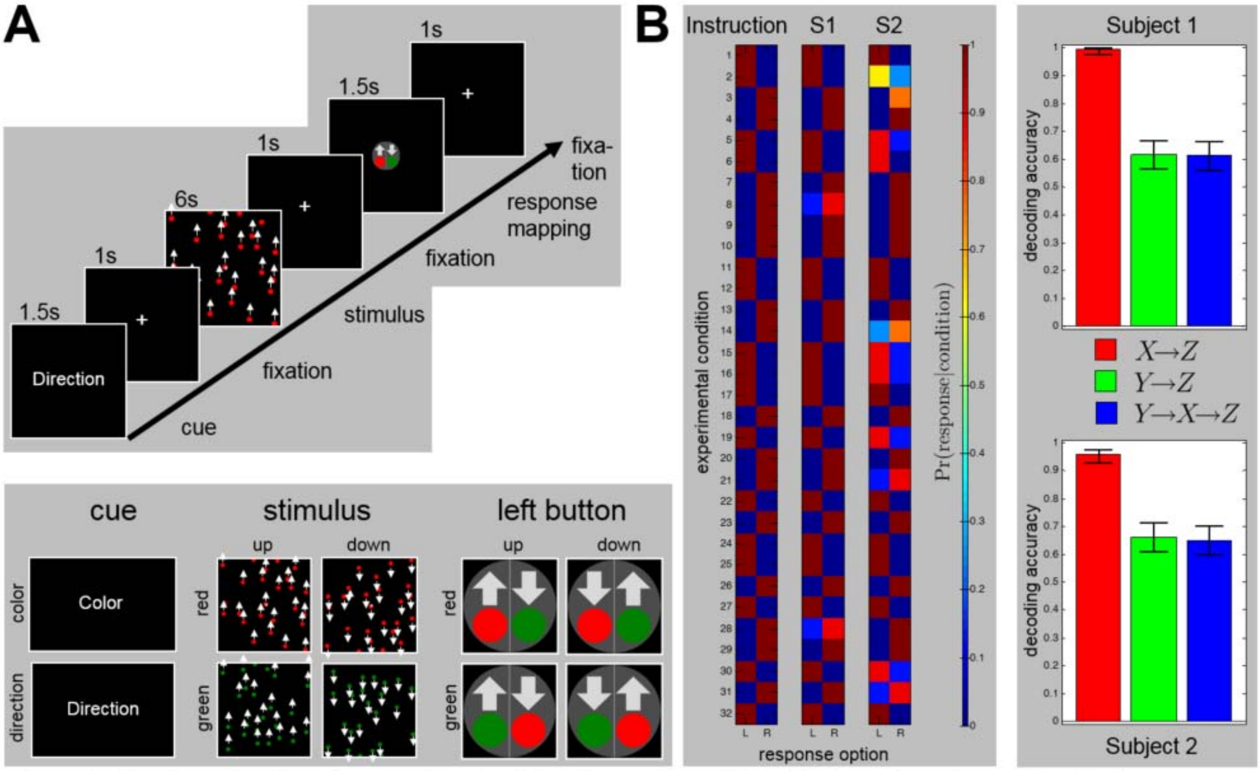

Figure 2. *Neurobehavioral decoding: experiment and decoding results.* **(A)** Two subjects performed a complex stimulus-response-mapping task that varied the cue which modality to respond to (color vs. direction), stimulus color (red vs. green), stimulus direction (up vs. down) and stimuli requiring a left button press (red vs. green, up vs. down). **(B)** For decoding from the design (red), transition probabilities were calculated and closely matched the instructed task (left). For CRD (green), responses were decoded from the response (mapping) phase. For NBD (blue), the design was decoded from the stimulus phase; cue and response mapping were supplied to the classifier. Error bars on decoding accuracies represent 90% binomial confidence intervals ($n = 256$).

Conclusions:

In this proof-of-concept study, we have demonstrated that behavioral responses can be decoded without training on neurophysiological data measured during behavioral responses, but rather indirectly by taking a detour via the experimental design. This is particularly interesting, because CRD is commonly seen as a sanity check, the decoding accuracy of which should not be exceeded by other analyses. It is also worth noting that in our example, just one response dimension (left vs. right), but two design dimensions (color and direction) had to be decoded. We hypothesize that decoding the design from the data acts as a feature reduction mechanism which helps NBD predicting behavior using the psychologically most meaningful factors. In the future, we want to validate this finding in a larger cohort [7,8] and extend it to continuous behavioral measures such as stimulus ratings and reaction times [9,10].

Higher Cognitive Functions:

Decision Making

Modeling and Analysis Methods:

Classification and Predictive Modeling [1]
Methods Development [2]
Multivariate Approaches

Perception, Attention and Motor Behavior:

Perception: Visual

Keywords:

Data analysis
Design and Analysis
Experimental Design
Machine Learning
Modeling
Multivariate
Perception
Statistical Methods
Vision

[1|2]Indicates the priority used for review

My abstract is being submitted as a Software Demonstration.

No

Please indicate below if your study was a "resting state" or "task-activation" study.

Task-activation

Healthy subjects only or patients (note that patient studies may also involve healthy subjects):

Healthy subjects

Was any human subjects research approved by the relevant Institutional Review Board or ethics panel? NOTE: Any human subjects studies without IRB approval will be automatically rejected.

Yes

Was any animal research approved by the relevant IACUC or other animal research panel? NOTE: Any animal studies without IACUC approval will be automatically rejected.

Not applicable

Please indicate which methods were used in your research:

Functional MRI
Behavior

For human MRI, what field strength scanner do you use?

3.0T

Which processing packages did you use for your study?

SPM
Other, Please list - ITEM Toolbox (https://github.com/JoramSoch/ITEM)

Provide references using author date format

[1] Haynes, J.-D. (2006), 'Neuroimaging: decoding mental states from brain activity in humans', Nature Reviews Neuroscience, vol. 7, iss. 7, pp. 523-534.
[2] Haynes, J.-D. (2015), 'A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives', Neuron, vol. 87, iss., 2, pp. 257-270.
[3] Görgen, K. (2018), 'The same analysis approach: Practical protection against the pitfalls of novel neuroimaging analysis methods', NeuroImage, vol. 180, pp. 19-30.
[4] Görgen, K. (2012), 'The Decoding Toolbox (TDT): A new fMRI analysis package for SPM and Matlab', OHBM 2012, Poster #378MT; URL: https://f1000research.com/posters/1092032.
[5] Hebart, M.N. (2015), 'The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data', Frontiers in Neuroinformatics, vol. 8, art. 88.
[6] Hebart, M.N. (2015), 'Example dataset for The Decoding Toolbox (TDT)'; URL: https://sites.google.com/site/tdtdecodingtoolbox/.
[7] Reverberi, C. (2012), 'Compositionality of rule representations in human prefrontal cortex', Cerebral Cortex, vol. 22, iss. 6, pp. 1237-1246.
[8] Reverberi, C. (2012), 'Distributed representations of rule identity and rule order in human frontal cortex and striatum', Journal of Neuroscience, vol. 32, iss. 48, pp. 17420-17430.
[9] Todd, M.T. (2013), 'Confounds in multivariate pattern analysis: theory and rule representation case study', NeuroImage, vol. 77, pp. 157-165.
[10] Woolgar, A. (2014), 'Coping with confounds in multivoxel pattern analysis: what should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013', NeuroImage, vol. 98, pp. 506-512.