Decoding behavioral responses from fMRI without learning behavioral responses from fMRI

Poster No:

2307

Submission Type:

Abstract Submission

Authors:

Joram Soch[1,2], John-Dylan Haynes[1,3,4,5,6,7,8,9]

Institutions:

[1]Bernstein Center for Computational Neuroscience, Berlin, Germany, [2]German Center for Neurodegenerative Diseases, Göttingen, Germany, [3]Berlin Center for Advanced Neuroimaging, Berlin, Germany, [4]Berlin School of Mind and Brain, Berlin, Germany, [5]Clinic for Neurology, Charité-Universitätsmedizin Berlin, Berlin, Germany, [6]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany, [7]EXC NeuroCure, Charité-Universitätsmedizin Berlin, Berlin, Germany, [8]EXC Science of Intelligence, Technische Universität Berlin, Berlin, Germany, [9]CRC Volition and Cognitive Control, Technische Universität Dresden, Berlin, Germany

First Author:

Joram Soch
Bernstein Center for Computational Neuroscience|German Center for Neurodegenerative Diseases
Berlin, Germany|Göttingen, Germany

Co-Author:

John-Dylan Haynes
Bernstein Center for Computational Neuroscience|Berlin Center for Advanced Neuroimaging|Berlin School of Mind and Brain|Clinic for Neurology, Charité-Universitätsmedizin Berlin|Department of Psychology, Humboldt-Universität zu Berlin|EXC NeuroCure, Charité-Universitätsmedizin Berlin|EXC Science of Intelligence, Technische Universität Berlin|CRC Volition and Cognitive Control, Technische Universität Dresden
Berlin, Germany|Berlin, Germany|Berlin, Germany|Berlin, Germany|Berlin, Germany|Berlin, Germany|Berlin, Germany|Berlin, Germany
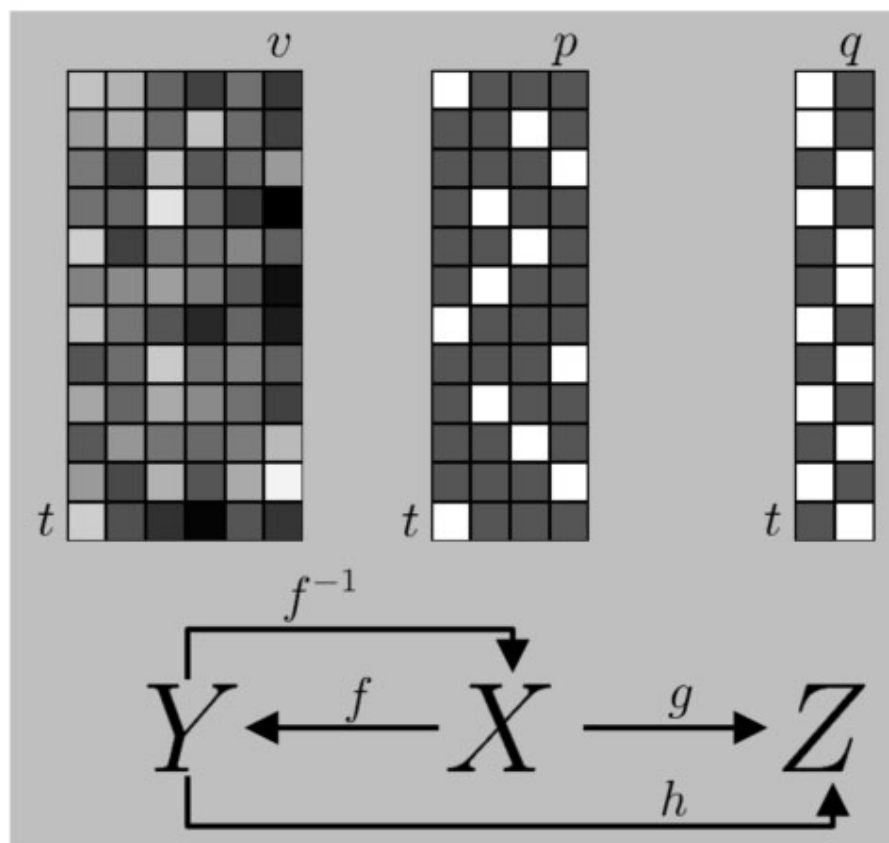
Introduction:

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), measured signals **Y** (i.e. BOLD signals in several voxels) and behavioral data **Z** (e.g. button presses, stimulus ratings). In multivariate pattern analysis (MVPA) of fMRI data [1,2], behavioral data are typically decoded by training an algorithm to predict 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 responses. In fact, behavioral data can also be 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:

Our main focus was on comparing three decoding algorithms: (1) First, behavioral data are directly predicted from the experimental design (*psychobehavioral model*, PBM), $Z = g(X)$, e.g. using a logistic regression model predicting discrete button presses from continuous stimulus variables. (2) For CRD, behavioral responses are directly predicted from measured signals, $Z = h(Y)$, here using inverse transformed encoding models (ITEM) [3] for searchlight-based trial-wise decoding. (3) For NBD, the experimental design is decoded from measured signals, $X = f^{-1}(Y)$, again using trial-wise ITEM, and the estimated behavioral model is used to indirectly predict behavioral responses from fMRI signals: $Z = g(f^{-1}(Y))$ (see Figure 1).

All these models are estimated in a cross-validated fashion, using leave-one-out cross-validation over fMRI recording sessions. Because behavioral responses and experimental conditions are best decoded from different parts of the brain, the approaches are made comparable by selecting the most informative searchlight based on within-sample decoding accuracy and reporting results as out-of-sample decoding accuracies.

Figure 1. *Theory behind neurobehavioral decoding.* The analyses operate on the trial-by-voxel matrix *Y* (e.g. fMRI signal amplitudes in a searchlight), the trial-by-condition matrix *X* (i.e. experimental design) and the trial-by-response matrix *Z* (i.e. behavioral responses). The theory assumes a neurophysiological model *f* (from X to Y, inverse *f⁻¹*), a psychobehavioral model *g* (from X to Z) as well as a mapping for conventional response decoding *h* (from Y to Z).

(https://files.aievolution.com/prd/hbm2101/abstracts/abs_2192/Figure_1.png)

Results:

We analyzed the entire data set (N = 108 subjects) from the *Neuroimaging Analysis and Replication and Prediction Study* (NARPS) [4,5]. In this experiment, subjects were offered a mixed gamble in each trial, with certain amounts of money to win or lose (= experimental design X), and then indicated favorability of the bet (= behavioral responses Z) using a four-point Likert scale [6]. In our investigations, we tested several analysis strategies (Analysis 1: discrete X & Z; Analysis 2: continuous X, discrete Z; Analysis 3: discrete X, continuous Z; Analysis 4: continuous X & Z) and observed that (i) the PBM clearly outperforms CRD and NBD; (ii) CRD and NBD both allow for above-chance decoding of behavioral responses; and (iii) CRD mildly outperforms NBD, except when decoding discrete Z from continuous X (see Figure 2). These results were largely unaffected by searchlight radius (r = 3, 6, 9 mm) and experimental group (a between-subject factor).
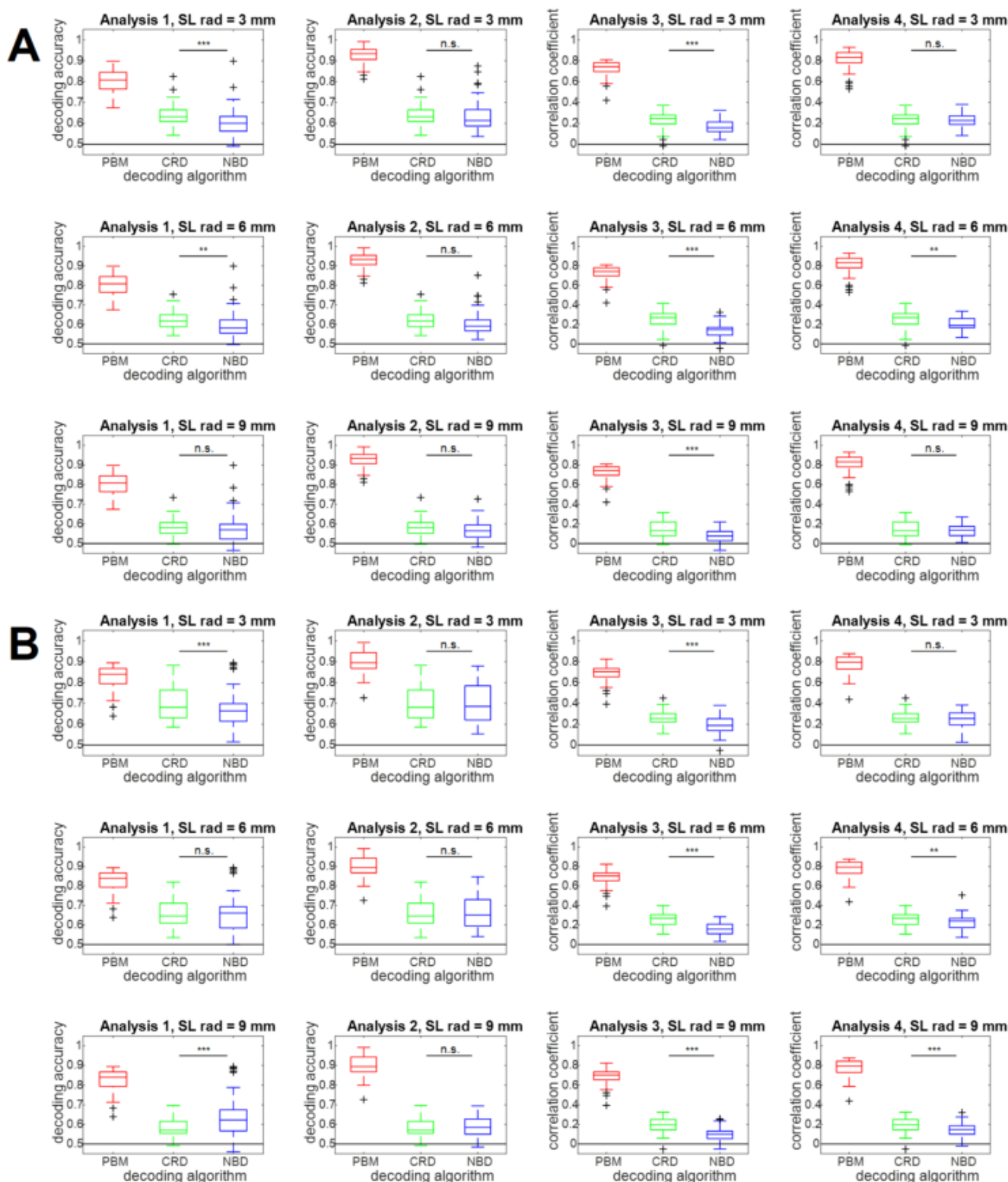
Figure 2. *Decoding accuracies as a function of decoding method.* Performances of three decoding algorithms (red: PBM; green: CRD; blue: NBD) are visualized using box plots for four analysis strategies and three searchlight radii, separately for **(A)** equal range and **(B)** equal indifference (*N* = 54 each) condition. CRD and NBD are tested against each other using a two-tailed paired t-test. Abbreviations: PBM, CRD, NBD: see text; n.s. = not significant; *p<0.05, **p<0.01, ***p<0.001.

(https://files.aievolution.com/prd/hbm2101/abstracts/abs_2192/Figure_2.png)

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 (favorability), but two design dimensions (gain and loss) 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 investigate the performance of NBD when the mapping from experimental conditions to behavioral responses is more deterministic [7,8] or completely random [9,10].

Emotion, Motivation and Social Neuroscience:

Reward and Punishment

Higher Cognitive Functions:

Decision Making

Modeling and Analysis Methods:

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

Keywords:

Cognition
Data analysis
FUNCTIONAL MRI
Machine Learning
Multivariate
Statistical Methods
Other - MVPA

[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
Computational modeling
Other, Please specify -   multivariate pattern analysis

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

Provide references using author date format

[1] Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. Nature Reviews Neuroscience. 7(7):523-534. doi:10.1038/nrn1931
[2] Haynes JD. 2015. A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. Neuron. 87(2):257-270. doi:10.1016/j.neuron.2015.05.025
[3] Soch J, Allefeld C, Haynes JD. 2020. Inverse transformed encoding models – a solution to the problem of correlated trial-by-trial parameter estimates in fMRI decoding. NeuroImage. 209:116449. doi:10.1016/j.neuroimage.2019.116449
[4] Botvinik-Nezer R, Iwanir R, Holzmeister F, et al. 2019. fMRI data of mixed gambles from the Neuroimaging Analysis Replication and Prediction Study. Sci Data. 6(1):106. doi:10.1038/s41597-019-0113-7
[5] Botvinik-Nezer R, Holzmeister F, Camerer CF, et al. 2020. Variability in the analysis of a single neuroimaging dataset by many teams. Nature. 582(7810):84-88. doi:10.1038/s41586-020-2314-9
[6] Likert R. 1932. A technique for the measurement of attitudes. Archives of Psychology. 22 140:55-55.
[7] Reverberi C, Görgen K, Haynes JD. 2012. Compositionality of Rule Representations in Human Prefrontal Cortex. Cereb Cortex. 22(6):1237-1246. doi:10.1093/cercor/bhr200
[8] Hebart MN, Donner TH, Haynes JD. 2012. Human visual and parietal cortex encode visual choices independent of motor plans. NeuroImage. 63(3):1393-1403. doi:10.1016/j.neuroimage.2012.08.027
[9] Soon CS, Brass M, Heinze HJ, Haynes JD. 2008. Unconscious determinants of free decisions in the human brain. Nature Neuroscience. 11(5):543-545. doi:10.1038/nn.2112
[10] Schultze-Kraft M, Birman D, Rusconi M, et al. 2016. The point of no return in vetoing self-initiated movements. PNAS. 113(4):1080-1085. doi:10.1073/pnas.1513569112