bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Localizing Brain Function Based on Full Multivariate Activity Patterns: The Case of Visual Perception and Emotion Decoding

* Isaac David †[1], Fernando Barrios ‡[1]

[1]Instituto de Neurobiología, Universidad Nacional Autónoma de México, Querétaro, México

2020-04-05

Abstract

It's now common to approach questions about information representation in the brain using multivariate statistics and machine learning methods. What is less recognized is that, in the process, the ability to perform data-driven discovery and functional localization has diminished. This is because multivariate pattern analysis (MVPA) studies tend to restrict themselves to regions of interest and severely-filtered data, and sound parameter mapping inference is lacking. Here, reproducible evidence is presented that a high-dimensional, brain-wide multivariate linear method can better detect and characterize the occurrence of visual and socio-affective states in a task-oriented functional magnetic resonance imaging (fMRI) experiment; in comparison to the classical localizationist correlation analysis. Classification models for a group of human participants and existing rigorous cluster inference methods are used to construct group anatomical-statistical parametric maps, which correspond to the most likely neural correlates of each psychological state. This led to the discovery of a multidimensional pattern of brain activity which reliably encodes for the perception of happiness in the visual cortex, cerebellum and some limbic areas. We failed to find similar evidence for sadness and anger. Anatomical consistency of discriminating features across subjects and contrasts despite of the high number of dimensions, as well as agreement with the wider literature, suggest MVPA is a viable tool for full-brain functional neuroanatomical mapping and not just prediction of psychological states. The present work paves the way for future functional brain imaging studies to provide a complementary picture of brain functions (such as emotion), according to their macroscale dynamics.

Keywords: MVPA, multivariate, brain-wide, pattern classification, functional localization, face perception, emotion, affective neuroscience, fMRI, task-oriented

*We are extremely grateful to Luis Concha, Eduardo Garza-Villarreal, Sarael Alcauter, Zeus Gracia-Tabuenca and Victor Olalde-Mathieu for their invaluable methodological guidance. To Eric Pasaye and UNAM's MR unit personnel, to Victor Olalde-Mathieu, Ana Y. Martínez and Lluviana Rodríguez-Vidal for their help with data acquisition. We would also like to thank Leopoldo González-Santos, Aliza Brzezinski, Diego Ramírez and Enrique Chiu-Han for their general advice and technical assistance.

†isacdaavid@isacdaavid.info

‡fbarrios@unam.mx

1 Introduction

Mapping of segregated brain functions is far from a settled methodology. Take for instance the very choice of statistical model in task-oriented functional neuroimaging for modalities like fMRI and PET: while the venerable *mass-univariate analysis* fits separate models to encode each brain time series based on experimental variables; *multivariate pattern analysis* (MVPA) reverses this — particularly multivariate pattern learning — fitting one model to decode experimental conditions out of the joint activity of several brain signals. The former is excellent at uncovering simple correlations between loci and functions; whereas the latter provides increased sensitivity due to emergent informational dependencies, at the expense of computational complexity.

For this reason and because having more dimensions than samples leads to overfitting — popular wisdom goes — multivariate searches must be restricted to regions of interests (ROI)[1–3] or moving searchlights,[4,5] or otherwise greatly reduced. Furthermore, the ability to map model parameters onto anatomy may not always be available, depending on algorithm transparency and nonlinearities.[6] Indeed, rigorous group inference of sensitivity clusters (akin to activation map thresholding) is seldom performed in MVPA studies; casting doubt on whether it will ever become a true match to classical functional brain mapping.

We don't call into question the strengths of each approach. However, how big of a difference do they make *in practice*? We suspect the limits of MVPA usage have been overestimated or don't apply as much anymore. With the right choice of learning algorithm and implementation, a neuroscientist may want to explore information-bearing patterns in the brain as a whole, analogously to mass-univariate analysis, before (or with-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

out) narrowing down the search to some subsystem or subjecting the data to the caveats of some dimensionality reduction method.[7] Other reasons for doing this include avoiding the statistical pitfalls of selection bias.[8] Also, sometimes there's no reason at all to prefer one spatial scale *a priori*;[9] as with pioneering studies of high-level, poorly-understood cognitive functions and personalized imaging of plastic ad-hoc skills.

As difficult as finding a near-optimal predictive model might be, amid all the noise voxels relative to the task;[10] it's hard to tell beforehand whether *zero* evidence of interesting long-range spatial patterns could be harnessed from brain-wide activity. In fact, the most recent advances in statistical learning tell the opposite story: against all expectations, deep architectures have come to grips with notoriously difficult problems by embracing their extremely high-dimensional nature.[11] Meanwhile, stringent observation of cross-validation and ROC-curve standards already account for exaggerated (i.e. overfitted) findings. Finally, this would complete the full spectrum of available complementary analytical perspectives purported since the introduction of MVPA. Just as allowing for multivariate dependencies might give a completely different picture to the traditional methodology,[12–14] so could do multivariate data drawn from a different spatio-temporal scale (e.g. neuronal ensembles vs large-scale networks). Yet the merits of such straightforward model have never been fully put to test, to the best of our knowledge.

Just one year after introducing the general linear model (GLM) for per-voxel analysis,[15] Friston et al. applied multivariate analysis of covariance on whole-brain data, reduced to a space of 35 canonical-variate eigenimages.[16] Importantly, this study not only established that multidimensional distributions of brain volumes could be distinguished under slight differences of cognitive-motor task conditions, but also that their hemodynamic transients were very heterogeneous despite of indistinguishable stationary statistical momenta. In general, detachment of focal neural response from experimental condition is one of the main motivations behind multivariate analysis.[12,17] Ever since, the existing whole-brain experiments either keep constructing a low-dimensional state-space from brain atlases and parcellations,[18] univariate voxel selection[10,19] and multivariate methods like principal components analysis and similar ones;[7,20–25] or they avoid reduction altogether but fall short of translating machine learning models into statistically-sound parametric maps for morpho-physiological insight.[6,10,26–28]

Algo./Ref.	Modality	Emotions	ROI	Accuracy
SVM[29]	visual	fear	many	78%> 50%
SVM[30]	auditory	happiness, anger, sadness, relief	auditory cortex	33% > 20%
SMLR[31]	visual	happiness, anger, sadness, disgust, surprise, fear	superior temporal sulcus, frontal operculum	22% > 14%
RSA[32]	auditory, visual (faces and body language)	happiness, anger, sadness, disgust, fear	brain (searchlight)	N.A.
SVM[33]	auditory	happiness, anger, sadness, surprise	brain (searchlight)	28% > 20%
SVM[34]	reading, visual (faces)	21 appraisals	theory of mind network, brain (searchlight)	~8% > 5% (ROIs)
MGPC[35]	visual (faces)	happiness, anger, fear	many	~32% > 25%
SVM[36]	visual (faces)	happiness	(*canis familiaris*) right temporal cortex, caudate. Brain (searchlight)	~65% > 50% (ROIs)

Table 1: Survey of experimental studies regarding emotion perception which employed MVPA. Column descriptions: Algo./Ref.: reference in bibliography and algorithm it uses. Modality: stimuli modality. Emotions: emotions under investigation (usually supplemented with an extra neutral category). ROI: region of interest. Accuracy: average classification accuracy, compared to theoretical random accuracy given the number of emotion categories.

For instance, the latter study by Raizada et al. actually was conducted in the same spirit as ours, and provided promising results keeping track of behaviorally-separable groups according to how they perceived phonemes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Statistical-anatomical maps were derived from voxelwise testing of GLM-based classifiers, but correction for multiple comparisons was absent.[28] Others simply don't get as far as performing functional localization.

Here we tested the competitiveness of pattern classification analysis (i.e., supervised pattern learning) as a methodology for anatomical localization of the correlates of cognitive functions in brain-wide, high-dimensional fMRI data. To this end, we employed rather standard and easily-interpretable support vector machine (SVM) classifiers on a sample of 16 human participants who performed a visual perception task. The task poses variable levels of decoding difficulty: from simple visual stimulation and face perception to the more ethereal perception of three basic emotions. We evaluate whether SVM can learn to predict task state above empirically-estimated chance performance, and if so, whether individual models converge on what the most relevant neural correlates of each cognitive ability are.

We expect both univariate and multivariate analyses to reveal well-known early visual cortex areas in contrasts intended to capture the effect of visual stimulation, and components of the so-called face processing network in the ventral stream during face perception.[37,38] If successful, this would provide greater confidence when exploring the correlates of the more poorly-understood emotional functions.

Although over a hundred years of affective neurobiology research have been fruitful in identifying the anatomical components of the emotional central nervous system; ample disagreement still exists on the physiological characterization of particular emotional experiences, even among metanalytical reviews.[39–45] It's not clear how the distributed activity of many limbic and other mid-line structures, from the posterior perivermian cerebellum to the medial prefrontal cortex (among others), gives rise to such behaviorally and evolutionarily relevant phenomena as sadness, rage or positive hedonic valence. This realization has in turn prompted the advent of more sensitive multivariate methods in emotion research[46–54] and, closely related, emotional perception research (see table 1); whereby multivariate activity localized in ROIs and searchlights has been found to outperform its univariate counterpart distinguishing among affective states.

2 Materials and Methods

2.1 Sample

Data came from a cross-sectional group of 16 volunteers from both sexes (8 female, 8 male), and an average age of 25 years, recruited at UNAM campus Juriquilla from October 2019 to June 2020. Participants were briefly interviewed to exclude those previously diagnosed with neurological or psychiatric conditions. With the exception of one male subject, all of them reported having right-handed phenotype. Prior to the study, subjects formally consented to participating after being informed of its aims, risks and procedures — in accordance with the 1964 Declaration of Helsinki — and were compensated with their brain scans and free diagnostics by a radiologist.

$n = 16$		
Age (years)	$\bar{x}$	$S.D.$
	25	3.01
Sex	Female	Male
	8	8
Education level	Undergraduate	Postgraduate
(obtained or in progress)	8	8

Table 2: Demographic features of the 16 successfully included participants.

2.2 Image acquisition

Parameter	EPI BOLD	T1w FSPGR
Slice orientation	Axial	Axial or sagital
Slices	35	176
Field of view	64×64	256×256
Voxel size	$(4\ mm)^3$	$(1\ mm)^3$
Flip angle	$\pi/2$	$3\pi/45$
TR (ms)	2000	8.18
TE (ms)	30	3.19
TInv (ms)		450

Table 3: Sequence parameters used for the MRI protocol. Abbreviations: EPI: echo planar imaging, BOLD: blood-oxygen-level dependent, T1w: T1-weighted, FSPGR: fast spoiled-gradient echo (GE's nomenclature), TInv: inversion time parameter for FSPGR T1w imaging.

Images were obtained from a 3-Tesla General Electric Discovery MR750 scanner at the MR Unit at UNAM's Institute of Neurobiology, during a single session per participant. The protocol included 5 echo-planar imaging (EPI) blood-oxygen level-dependent (BOLD) sequences for fMRI, 185 volumes each. A T1-weighted scan of head

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

anatomy was also acquired. Sequence parameters are described in Table 3. Electromagnetic responses were recorded using a head-mounted 32-channel coil.

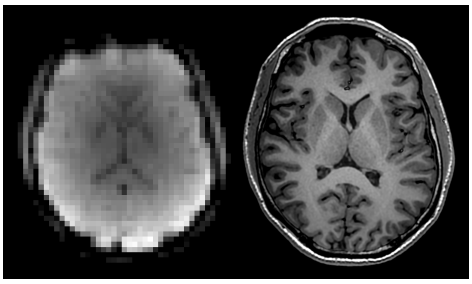

Figure 1: Raw samples of both image modalities for a single subject in our dataset (in the same order as in Table 3).

2.3 Stimuli and task

Each of the 5 fMRI sequences was temporally coupled to a psychological block-based task implemented in PsychoPy 3.0.1.[55] All 5 tasks were identical, save for the pseudo-random order in which their 30 s blocks were administered. A total of 6 block classes were used: happy faces, sad faces, angry faces, neutral faces, pseudo (scrambled) faces and low-stimulation. Neutral/inexpressive faces might provide an extra control when contrasting among emotions. Otherwise, one might risk mistakenly concluding from classification analysis that n emotions are identified, when in fact only $n-1$ are, in addition to something else that is neither the $n-1$ emotions nor the remaining one. Pseudo-faces and dim blocks were introduced so as to buttress and diagnose the analysis pipeline, by way of more trivial contrasts (like pseudo-faces vs low-stimulation and faces vs pseudo-faces).

Each block in turn comprises 10 randomly-presented images belonging to that class, each one shown for about 3 seconds and without possibility of reinstantiation during the same block. Each block occurs twice per sequence, yielding a total of 12 of them (360 s = 6 min). After their presentation, participants had to wait for 10 seconds before concluding the sequence, in order to capture the hemodynamic response (HR) elicited by the last stimuli. A selection of 10 grayscale photographs per category of frontal human faces (male and female) served as stimuli. These were chosen from the classical "Pictures of Facial Affect" database.[56] As for the low-stimulation (a.k.a. "dim") blocks, a small but visible fixation cross was made fluctuate from quadrant to quadrant at random every 3 seconds. The whole task is summarized in Figure 2.

Additionally, behavioral responses were recorded throughout the task in order to measure performance and thus evaluate the suitability of physiological data for further analysis. Participants were instructed at the beginning of every sequence to indicate whether faces belonged to a man or a woman as soon as they were perceived. The response was submitted with the press of a button — one at each hand. Analogously, for scrambled and dim blocks (when no faces should have been perceived), the instruction was to simply report image change, alternating between buttons. In this fashion, motor activity remained rather homogeneous for all blocks, minimizing a possible confounding effect when contrasting among faces and pseudo-faces. Even though the whole task was explicitly orthogonal to thinking about emotions, one cannot rule out the possibility that such linguistic-conceptual process spontaneously appeared in the participants' train of thought as percepts were experienced. Statistical analysis of behavioral data was conducted using the R programming language.

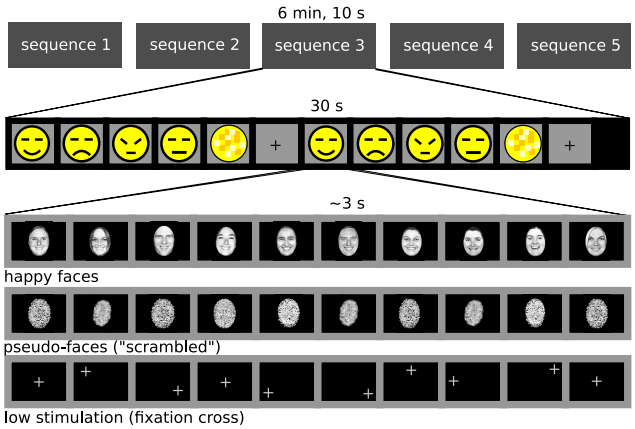

Figure 2: Block-paradigm design of the psychological experiment. The horizontal axis corresponds to the passing of time. Rectangles represent stimulation units (sequences, blocks or individual stimuli).

2.4 Analysis methods

2.4.1 Data, source code and reproducibility

All the source code necessary for reproducing, analyzing or adapting the present study is made available as free software, both for the behavioral task described in the previous section (https://github.com/isacdaavid/emotional-faces-psychopy-task) and the neuroimaging analysis (https://github.com/isacdaavid/np-mvpa)

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

as described next. Original data and final group activation maps in standard space may be downloaded respectively from OpenNeuro (https://doi.org/10.18112/openneuro.ds003548.v1.0.0) and Neurovault (https://identifiers.org/neurovault.collection:9492).

2.4.2 Image preprocessing

Functional and T1w scans were converted from the DICOM format[57] to NIfTI-1[58] and structured in a file tree according to the BIDS 1.4.0 standard[59] using the Dcm2Bids 2.1.4[60] tool, which in turn was configured to use the dcm2niix 1.0.20170411 converter[61] and anonymize the faces of participants with pydeface 2.0.0.[62] The sanity of the resulting database was further checked with BIDS Validator 1.5.4. T1w images were submitted to the Volbrain tissue segmentation and volumetry Web service,[63] whose resulting brain and gray/white matter masks we used for deskulling the field-bias-corrected T1w images and, later on, for selection of fMRI voxels. fMRI sequences were concatenated by temporal order in one long sequence per subject, then the result underwent the following preprocessing pipeline due to the FSL 6.0 utilities:[64] high-pass frequency filter (>50 s) and interpolation for slice-time correction (interleaved acquisition),[65] affine movement correction and coregistration[66,67] with the respective T1w anatomical reference and the standard MNI-152 T1w template[68,69] at 1 mm of resolution. After registration, the corresponding resulting matrices were applied to the Volbrain masks, so as to transform them to the low-resolution subject-space of fMRI images. Gray-matter time series were extracted afterwards (about 10000 depending on subject), and linear trends were subtracted by preserving residuals from a simple linear regression performed on each of the 5 sequences the long concatenated time series is composed of. Finally, and seeking not to bias classification models in any dimension, the composite time series at each voxel is normalized to z-scores; pushing the covariance matrix of the multivariate data to resemble an identity matrix and thus decorrelating phase space.

2.4.3 Univariate analysis

In the vein of assessing the feasibility and performance of the brain-wide multivariate approach against the golden standard in functional brain mapping, we investigated the same 2-way contrasts included in multivariate analysis using FSL 6.0 in a classical mass-univariate analysis. Said contrasts are grouped into *visual stimulation* ("dim vs scrambled", "dim vs neutral", "dim vs angry", "dim vs sad", "dim vs happy"), *face perception* ("scrambled vs neutral", "scrambled vs happy", "scrambled vs sad" y "scrambled vs angry") and *emotion perception* ("angry vs happy", "sad vs happy", "sad vs angry", "angry vs neutral", "happy vs neutral", "sad vs neutral"). Preprocessed data (up until subtracting linear trends and normalizing) were spatially-smoothed with a Gaussian convolution kernel of 5 mm FWHM. Each of the 6 block classes described for the task was considered as a column-vector regressor in the design matrix $\mathbf{X}$, after convolving them with a zero-lag, double-gamma hemodynamic response curve. $\mathbf{X}$ was augmented with the time derivatives of each convolved regressor, but no motion covariates were added. General linear models are fitted afterwards. GLM is a matrix-form extension to multiple linear regression, which models each physiological time series (column $\mathbf{y}$) as a linear combination of $\mathbf{X}$ plus some gaussian error $\mathbf{E}$. The model reads:[13]

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Theta} + \mathbf{E}; \quad \mathbf{E} \sim \mathcal{N}(0, \boldsymbol{\Sigma}) \tag{1}$$

Assuming trial independence and homocedasticity, maximum likelihood estimation or ordinary least-squares estimation may be followed to obtain the so-called normal equation, which optimizes parameters $\boldsymbol{\Theta}$ according to:

$$\hat{\boldsymbol{\Theta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{2}$$

After estimation, contrasts of parameter estimates are subject to the same procedure as the multivariate model parameters for purposes of group-level inference. This is described in more detail in the *Group inference* subsection.

2.4.4 MVPA

Once preprocessed, multivariate decoding of fMRI patterns was conducted using the pyMVPA 2.5.0 Python library.[70] We trained a linear support vector machine (SVM) classifier per subject and block contrast combination using all available brain volumes (120 volumes per class for the training phase, 30 for testing). In ad-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

dition to the contrasts described in the previous section, emotion-related ones were augmented with the $\binom{4}{3}$ possible 3-way classification problems and the single 4-way contrast. The supervised SVM algorithm learns a hyperplane for binary classification in high-dimensional phase space.[71,72] Given a vector $\mathbf{w}$ orthogonal to the hyperplane, the SVM decision rule is equivalent to the sign of the projection of unseen data vectors $\mathbf{y}_i$ on $\mathbf{w}$, adding or subtracting the necessary constant b so as to make the result exactly 0 at the hyperplane:

$$sgn(x_i) = \mathbf{w} \cdot \mathbf{y}_i + b \tag{3}$$

Out of all possible hyperplanes, SVM's key insight is to estimate the one that maximizes separation margin to the most difficult training data: the support vectors right above opposite margin lines. Since margin width can be calculated from pairs of positive-class and negative-class support vectors according to:

$$margin = (\mathbf{y}_+ - \mathbf{y}_-) \cdot \frac{\mathbf{w}}{||\mathbf{w}||} = \frac{(\mathbf{w} \cdot \mathbf{y}_+ - \mathbf{w} \cdot \mathbf{y}_-)}{||\mathbf{w}||}; \tag{4}$$

by constraining the decision rule to satisfy $|\mathbf{w} \cdot \mathbf{y}_i + b| \geq 1$ or similar criteria and substituting on equation (4), one can show that maximizing the margin — and therefore obtaining an optimal model — is equivalent to a quadratic programming problem with $||\mathbf{w}||$ as the cost function to be minimized (mathematical details are discussed by Mahmoudi et al.[13]):

$$max\ margin = max\ \frac{2}{||\mathbf{w}||} \tag{5}$$

SVM models were cross-validated using each of the 5 sequences as a fold, and their mean classification accuracy (hits / misses) served as a summary test statistic. This was compared against an empirical null model in a non-parametric rank-based test, by estimating the probability distribution of average classification accuracy given H_0 via Monte-Carlo simulations: surrogate data is computed by randomly shuffling the class labels of the training data partition 5000 times, and 5-fold cross-validation is again conducted for each permutation. Then, a p-value for that particular subject and class combination can be calculated as the proportion of random results equal or greater to the original classification accuracy.

We also explored the effect of different delays between stimulus onset and volume labeling (from 0 s to 10 s, every 2 s), instead of assuming a single optimal HR peak;[14] although, based on common practice, a delay of 4 s was fixed a priori for all group-level statistical inference, avoiding inflation of type-I error.

2.4.5 Group inference

For every contrast, per-subject permutation tests are brought together to be assigned an average p-value, and to estimate effect size on classification accuracy (signal-to-noise ratio) compared to the null distribution according to Cohen's D statistic.

Contrasts for which sample evidence of successful multivariate decoding above chance levels was observed were further selected for discovery of anatomical clusters. Failing contrasts are also inspected, but only to serve as qualitative reference for the true merits of the good ones; we warn against doing anatomical inference in real applications in the absence of model evidence. Our conservative expectation is that, while parameter inspection should only be justified for successful classifiers; the features driving such models still may or may not display coherence among different participants, placing an extra statistical safeguard before drawing conclusions.

To this end, we employed yet another non-parametric test with 5000 sign permutations using FSL 4.0's *randomise*[73] with the Threshold-Free Cluster Enhancement,[74] operating on the group of SVM weight vectors in a 2-tail test (after L2-normalization, i.e. vectors are unitary, transformation to the standard 1 mm MNI-152 space and spatial smoothing with a 5 mm FWHM gaussian kernel). In the case of univariate analysis, input data to TFCE were the group of GLM contrasts of parameters (both positive and negative effects, also transformed to the 1 mm MNI-152 space). Given a parameter map $h(v)$, the TFCE statistic at some voxel v is defined as the integral (in the Lebesgue sense) of cluster size $s(v, h)$ times the cluster-defining "height" h:

$$TFCE(v) = \int_{h=0}^{h(v)} s(v, h) h\ dh \tag{6}$$

Equation (6) is often modified for fMRI and EEG data, where the default is to favor h, squaring it, while taking the square root of s. Since all possible cluster-forming

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

thresholds are considered, TFCE is regarded as a more principled (as well as more powerful and specific) alternative to other nonparametric cluster-informed inference methods, while still providing strong family-wise error (FWE) control;[75] as expected of permutation-based approaches. We set a cutoff significance value of 0.01 in the corrected *p_value* brain maps. Under TFCE, surviving voxels are interpreted as belonging to some signal-containing cluster, but no guarantee exists as to where the exact clusters lie (although adjacency of many such surviving voxels may make this visually obvious).

3 Results

3.1 Behavior

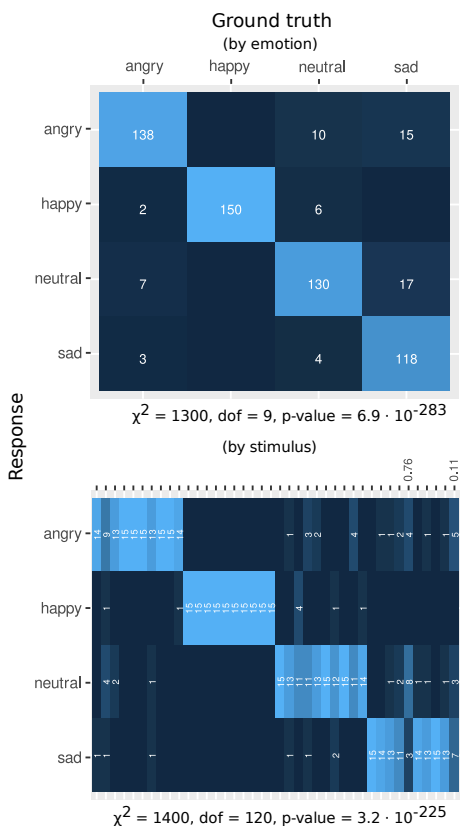

Figure 3: Confusion matrices with the group joint frequency of responses obtained during preparatory picture validation. For datasets true to their purpose, a strong diagonal should be observed, indicating agreement between subjective perception and preset categories. This is quantified with Pearson's χ^2 tests, whose results are displayed below each contingency table. **Top**: when grouping stimuli by emotion. **Bottom**: grouping only by picture, to detail the fine-grained structure of errors, Holm-corrected p-values are shown for the only two stimuli with $p > 0.05$, according to one-tail binomial tests under the hypothesis that the correct category is only assigned to 1/4th of all Bernoulli trials, presupposed to be statistically independent.

Prior to entering the MRI machine, 15 of the 16 participants were asked to answer a brief randomized stimuli categorization task using the same faces they would later experiment inside the scanner. Faces could be assigned to one of four classes with a computer mouse: angry, happy, neutral or sad. A Pearson's χ^2 test for association strength between intended emotion and subjective interpretation assigned a probability of $6.9 \cdot 10^{-283}$ to the possibility that participants were categorizing stimuli at random. A similar test was performed separating by stimulus as opposed to emotion class; nonetheless, the p-value remained very low at $3.2 \cdot 10^{-225}$. Both results are shown in Figure 3. Despite variability recognizing among different basic emotions, our success rates turn out to be similar to those reported in independent validations of other datasets.[76,77] Similarly, per-participant χ^2 tests (with Bonferroni correction for FWE) revealed that even the worst-performing participant had a probability of less than $5 \cdot 10^{-8}$ of being involved in guesswork.

With regard to instantaneous responses during the task, we ran binomial tests to quantify success probability detecting face gender and image change, assuming statistical independence and a chance level of 50%. Figure 4 shows the aggregate of hits through time. Errors, in red, are comparatively low. A probability of $1.95 \cdot 10^{-60}$ (Holm-corrected) of finding such hits/misses ratio by chance was found for the worst participant, and the probability for the worst block type (including pseudo-faces and dim-stimulation) is even smaller.

Participant's reaction times (RTs) were analyzed as well, as a measure of attention to the task. Each curve in Figure 4 corresponds to the RTs of some subject. The superimposed dotted black line projects the relevant part of a general linear mixed-effects model (GLMM). GLMM is a generalization of GLM regression which uses two — as opposed to one — design matrices to account for random effects. This is specially suitable to hierarchical factorial designs; since the variance of measurements at some time t could come from intrinsic differences among participants, whose personal variance is captured by the random effects. The model was fitted using the participant factor as a random effect, and block lapse and block class as fixed effects. The aim is to find the effect of time upon RTs, because big changes would signify disengagement from the task. On the contrary, we observed a

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

negligible, downward slope (0.11 ms faster RTs every 30 s block).

Moreover, a post-hoc Tukey test for a one-way ANOVA of RTs was inspected, using block types as factor levels. The only emotion to elicit considerably different reaction times was anger (angry vs neutral p=.014, angry vs happy p=.03). No significant difference was found between reacting to pseudo-faces vs to fixation crosses. However, we measured extremely large differences between reacting to any type of visuofacial stimulus and any type of non-visuofacial stimuli, for which not only stimulus complexity is lower, but task complexity is also lower (telling gender vs noticing any change at all).

All these lines of behavioral evidence converge towards the conclusion that participants understood the task and that stimuli were correctly observed in general. Accordingly, no participant or block type was discarded for analysis of the fMRI data after this screening.

3.2 Visual stimulation and face perception

All contrasts meant to distinguish between high and low visual stimulation and between face and pseudo-face perception presented strong evidence of successful decoding using the multivariate model, both on an individual and on a group-level basis. In the case of visual stimulation, all p-values on classification accuracy per contrast (both individual and average) were found to be lesser than $2{\cdot}10^{-4}$: the smallest result that could have been obtained with 5000 permutations. In other words, no classification accuracy greater to the models' was ever found by chance. Moreover, we found extremely huge group effects; always greater than Cohen's $D = 6.5$ (dim vs neutral) and as big as $D = 7.6$ (dim vs happy). Similarly, face perception compared to a pseudoface perception baseline always resulted in a $p < 2 \cdot 10^{-4}$; both individually and as group means. Cohen's statistic ranged from $D = 5.2$ (scrambled vs neutral) to $D = 6.8$ (scrambled vs angry), also to great effect. Figure 5 displays hypothesis tests for a couple of experimental contrasts, for the sake of illustrating the nature of results.

Group analysis of model parameters is presented in figure 6. FWE-corrected "$1-p$" anatomical maps for the 5 visual-related contrasts are averaged and thresholded at $1-p > .99$, in yellow. This is repeated for the 4 face-perception-related contrasts, shown in cyan.

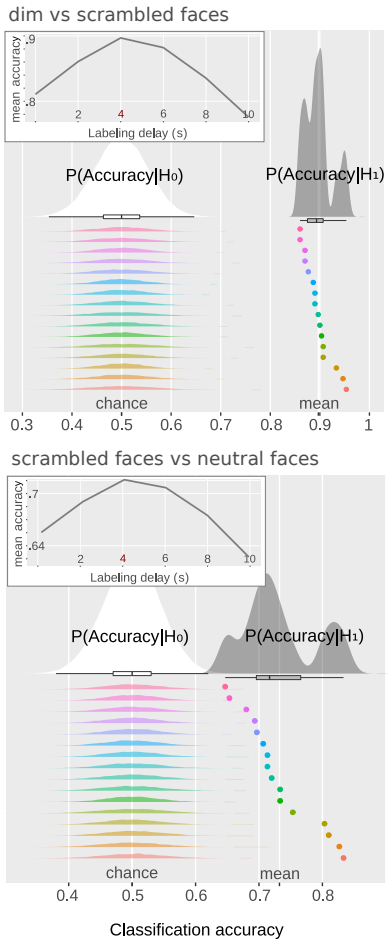

Figure 5: All 9 contrasts related to simple visual perception and face perception (only 2 shown here) are strongly dissociable, according to their brain-wide activity patterns. Embedded at the top-left corner of each subfigure: time series of group-mean classification accuracy as a function of labeling delay. Greater subfigure: hypothesis tests of classification accuracy for a preset labeling delay of 4 s (H_1: classification accuracy is greater than chance performance). The rainbow-colored dots at the bottom stand for the cross-validated classification accuracy of each participant, compared to their respective null distributions estimated from 5000 permutations of data labels.

Both univariate and multivariate approaches agree on two prominent, bilaterally-symmetric occipital clusters whose activity correlates significantly with the presentation of our visual stimuli, relative to the "dim" fixation cross: one posteromedial, encompassing part of the primary visual cortex (V1) and extrastriate areas like V2; which become less specific in the univariate model, possibly reaching parts of ventral V3 (so-called VP) and color and form-related V4. From there it crosses parenchymal boundaries to the medial posterior cerebellar lobe in its anterior portion. The second major cluster lies at the anterior medial occipital cortex, and on the univariate

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

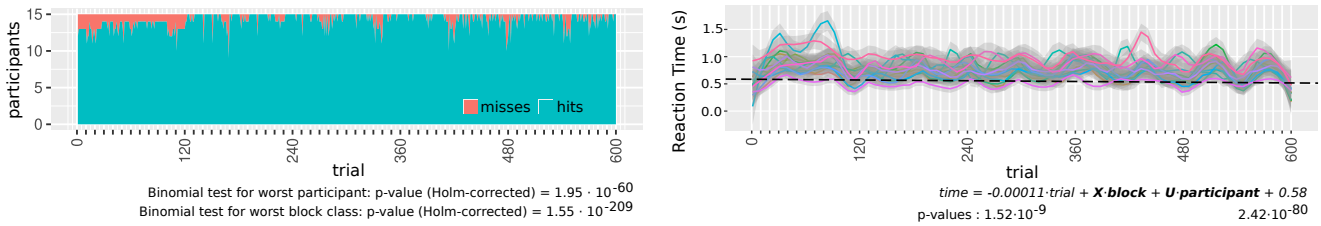

Figure 4: Instantaneous performance during the task. Each notch in the horizontal axis represents 10 consecutive trials; i.e., a 30 s block. Each sequence is delimited by the 120-trial-long marks. **Left**: ratio between hits and misses as a function of time. **Right**: per-subject reaction time, and a linear fit according to a GLMM. Each solid curve stands for the LOESS polynomial fit of a different participant and its confidence interval at 95%. The dotted line shows the almost-null linear trend from the GLMM. The actual GLMM is shown underneath, with explicit values for important parameters together with their respective p-values.

model is due to anticorrelation. It includes a calcarine component at anterior V1, part of the V2 ring and then extends more dorsally into the cuneus to motion-related V3a and, possibly, the midline section of form-related V3. Associated activity also survives well into dorsal-stream regions, perhaps because of the moving fixation cross, particularly at Brodmann area 7 of the superior parietal lobes (SPL). This is stronger for the univariate model, where noticeable parietal anticorrelations were also present at the right precuneus, inside the postcentral sulcus to be precise. Finally, both models also display a bilateral component at the posterior middle temporal gyrus — although more prominent on the left hemisphere and for the univariate model — near or at the angular gyrus and suggestive of V5/MT. On the anteroventral direction, the second cluster also extends bilaterally to the lingual gyri (LG), close to parahippocampal gyri (PG) tissue. Interestingly, a whole stripe of medial occipital cortex is spared, making the two clusters distinct, consistent with an unexpected effect of stimuli position on the activity of retinotopic cortical topology.

Face perception was related to two clusters: an anticorrelated, bilaterally-symmetrical pattern at V1 and V2 (roughly a subset of the first cluster described for visual stimulation). The second one comprises a bilaterally-symmetrical portion of the inferior occipital gyri within the lateral occipital cortex (LOC), which has been dubbed the "occipital face area" (OFA).[37,38] The multivariate approach was able to consistently localize the fusiform face area (FFA) at the right inferior temporal cortex as well, with important below-threshold evidence at its left counterpart. In comparison, univariate analysis barely detected a small cluster of correlated activity at OFA; whereas only 2/4 contrasts showed above-threshold evidence at FFA, which is why their average failed to make their way into figure 6.

3.3 Emotion perception

Figure 7 is a compilation of group tests for the remaining 11 emotion-related contrasts in the form of a Venn diagram, including block combinations which included neutral faces. Once again, the null classification accuracy model is represented by the white probability distribution and true decoding success is in gray. We observed wide variation in model success, yet this variability turned to be structured according to the emotions under probe; from $\bar{p} = 0.05$ and huge effects at $D = 3.3$ (happy vs neutral), to $\bar{p} = 0.85$ and $D = -1.3$ (anger vs sadness). It's evident from the diagram in Figure 7 that contrasts which included *happiness* in general outperformed their respective null models. On the other hand, classification models that excluded this emotion did not. The importance of happiness driving accurate prediction was further confirmed via qualitative inspection of confusion matrices.

Group-level inference of model parameters resulted in suprathreshold clusters with high colocalization — even among contrasts — with two possible anatomical spans. The two spatial configurations were contingent upon whether neutral faces had been included in the classification problem. In their absence ("happy vs angry", "happy vs sad", "happy vs angry vs sad"), detection of happiness consistently depended on activity at the occipital pole and its midline and ventral surroundings (posterior V1, posterior V2, ventral V3, V4, V8) as well as anterior V2 (both dorsal and ventral) and V3. This is shown in yellow in figure 8. When also faced with the neutral facial expression controls, SVM was forced to ex-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

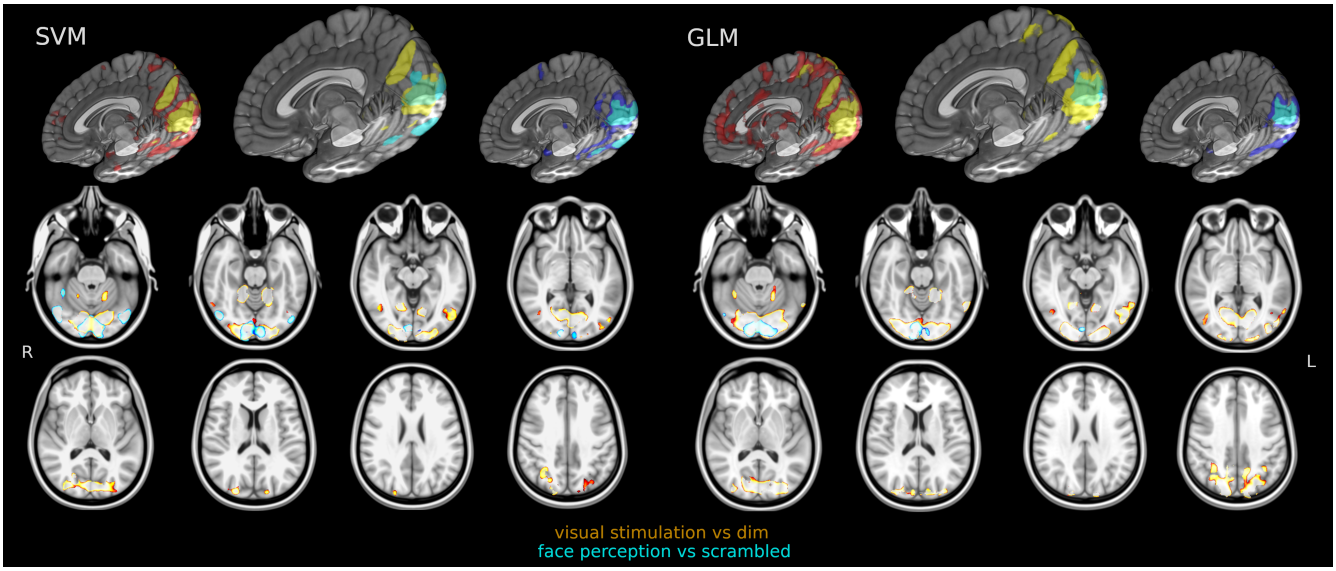

Figure 6: Average anatomical distribution of thresholded ($p_{FWE} < 0.01$) TFCE maps. Yellow: visual stimulation (average of corrected p-values for the following contrasts: "dim vs scrambled", "dim vs neutral", "dim vs angry", "dim vs sad", "dim vs happy"). Cyan: face perception (average of "scrambled vs neutral", "scrambled vs happy", "scrambled vs sad" y "scrambled vs angry"). Two views of the T1w MNI-152 template are included: 3D volume rendering of right hemisphere (on top), axial slices in radiological orientation (bottom). The smaller 3D volumes include the same maps as the main 3D rendering of each model, but separated by contrast family and with an additional lower threshold at $p_{FWE} < 0.5$ in darker colors. This is to highlight model anatomical specificity at the minimum level that could still be considered evidentiary. Full maps may be downloaded from https://identifiers.org/neurovault.collection:9492 .

tend the search to higher-order and lower-order structures: both SPLs, and anterior-LG/posterior-PG and the right LOC appear again. The same medial portions of posterior cerebellum are also included. The only new cluster with respect to the first to be described for figure 6 is the posteromedial thalamus, likely including both pulvinar bodies entirely. Although not shown in the figures, important subthreshold evidence ($p_{FWE} < .05$) exists at the left amygdala ("happy vs neutral", "happy vs angry", "happy vs sad"), at LOC/OFA (all contrasts) and the right inferior temporal sulcus in medial patches (happy vs angry vs neutral). All successful subject-level SVM models gave prominent weighting to the ventromedial prefrontal cortex (orbitofrontal cortex), however, the amount of clusters and parameter signs in that anatomical region were too heterogeneous to accumulate the evidence at the group-inference level.

As discussed in our methodological considerations, unsuccessful decoders returned considerably smaller or virtually nonexistent anatomical clusters: the left posterior thalamus ("angry vs neutral", "sad vs neutral", "angry vs sad vs neutral"), posterior V1 and V2 ("angry vs sad", "angry vs sad vs neutral") and small clusters in PG and the quadrigeminal area in the brainstem (angry vs neutral). These are rendered in the three bottom rows of figure 8. No suprathreshold cluster or voxel, correlated or anticorrelated, was found for any of the 6 emotion-related GLM contrasts.

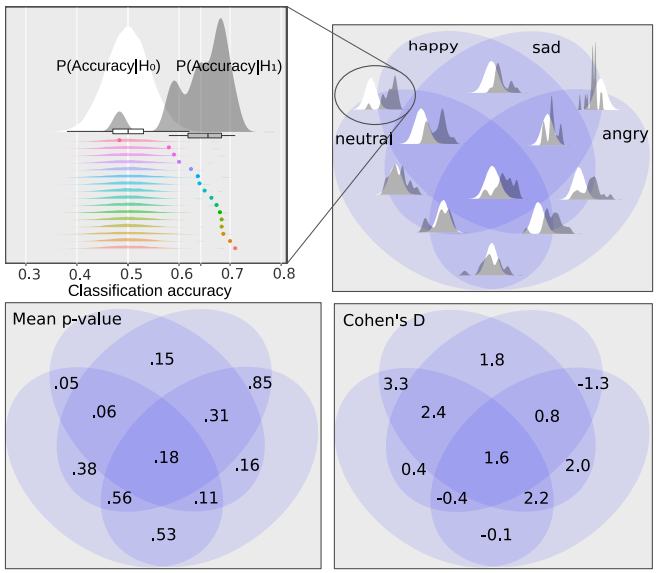

Figure 7: Group aggregates of hypothesis tests on classification accuracy for all emotion combinations, together with their associated mean p-value and effect size (Cohen's D).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 8: Anatomical distribution of thresholded ($p_{FWE} < 0.01$) TFCE maps (SVM parameter vectors) for emotion perception contrasts, one row per emotion combination. Contrasts in cyan color also included the neutral faces class. Slices are in radiological axial orientation in the standard MNI-152 space. Full maps may be downloaded from https://identifiers.org/neurovault.collection:9492 .

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4 Discussion

Results validate the feasibility of looking for multivariate correlations between functional neuroimaging and perceptual phenomena of varying complexity, and of turning learned data patterns into statistical-anatomical maps for localization of task-related brain structures. This was shown for an extrinsically high-dimensional phase space using all the available gray-matter information, which differs substantially from the routes taken by classical univariate analysis and other MVPA studies. It is remarkable that an algorithm as modest as the kernel-less SVM can characterize many psychological states of scientific interest in a purely data-driven approach, to the point of surpassing the *en masse* linear regression methodology for its brain mapping features.

For instance, the neural correlates for the simple visual stimulation subexperiment (plus alleged stimulus-motion artifacts) are largely identical according to both approaches (figure 6), not to mention consistent with the known neurophysiology of vision, pointing in the direction of methodological validity for the multivariate pipeline. However, remaining contrasts decidedly favoured the brain-wide multivariate analysis, sample size being equal. Classical analysis notoriously failed to consistently discover the right FFA for the face perception subexperiment.

As to emotion perception, it might seem tempting to disparage the multivariate approach for its major reliance on visual, rather than emotional areas other than the perivermian posterior cerebellar cortex[78] and the parahippocampal gyrus. However, the fact that *perceived* happiness reliably elicits a distributed activity fingerprint — which was invisible to GLM — still counts as a victory point for our proposal. Whether this particular "happy interlocutor state" is truly a noncollateral biological feature of social significance is hard to answer with our data. On one hand, recent experiments using electrophysiological and calcium-imaging techniques on rodents have emphasized the existence of notorious motor and arousal-related information in areas traditionally thought of as sensory.[79,80] On the other, it may be argued that areas like V1 aren't particularly concerned with constructing face or affect percepts, yet the whole of their lower-order computation may be more readily leveraged by a statistical model about facial expression; similar to how artificial vision and object recognition systems emulate cortical computations starting from nothing but raw pixels. That would certainly pose a methodological challenge to our approach; which we showed was alleviated to some extent by the diligent use of control stimuli (neutral faces).

This result is of great interest, in light of the incipient works on emotion as seen through the MVPA prism and the relative looseness with which they have been conducted. Some of the literature from table 1 also included anger and sadness-loaded stimuli, with better results than us;[30–33] yet it raises reasons to be skeptical upon closer inspection. For instance, the ROI-based, auditory study by Ethofer at al. reported average classification accuracies (n=22) of 30% and >35% for sadness and anger respectively; among 5 emotions.[30] Nonetheless, models were trained only pair-wisely: that is, contrasting target emotion against an "everything else" metaclass. No empirical null model was estimated. This one-vs-all scheme without nonparametric testing was repeated by Kotz et al. [33], yet here anger showed the poorest results. Other issues include comparing against a scarce number of permutations, for instance, in the study conducted by Said et al. [31].

In conclusion, we fail to find convincing reasons in previous works to suppose that our failure decoding anger and sadness is due to a failure on our part; other than the choice of algorithm and input data. Perhaps results for these two basic emotions could have improved, had we used a more localized search. Those affective-perceptual states might be genuinely underrepresented in the coarse fMRI data, or they may not be linearly separable or the system dynamics may not be sufficiently stationary in the relevant dimensions. As argued in the *Introduction*, this study aimed at testing the limits of linear SVM as a data-driven anatomical mapping tool, at the expense of decoding performance. In that sense, and joined by the modest sample size, being able to retrieve just *some* emotional states out of BOLD activity emanating from well-defined structures already cements the accomplishment of our goals. We think our tooling and testimony have the potential to influence a plethora of noninvasive neuronal prediction and "mind reading" studies, plus many more regarding the neurobiological segregation of cognitive, affective and behavioral features of humans and

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

other organisms.

The present work also suffers from limitations and future oportunity areas. Further analysis is required to characterize the intrinsic dimensionality of each cluster system, for instance by use of covariance-matrix decomposition methods, information-theoretic methods or topological/manifold embeddings. Similarly, system dynamics could be studied and mathematically modelled to provide further characterization and understanding of each succesfully decoded state, as well as the encompassing attractor set. It would also be interesting to extend the task to other emotions, modalities and theoretical models of emotion; for instance, in order to be able to tell whether we have a sufficient characterization of happiness (as opposed to appetitive hedonic valence generally, as posited by dimensional theories of emotion). A second strand of further studies should explore these findings using more direct causal interventions with a number of techniques, so as to assess the relevance of multivariate statistical connections.

5 Conclusions

The realization that some cognitive and affective states are not very localized, but might emerge from the *joint* activity of distributed neuronal populations has led to the popularization of machine learning tecniques in cognitive neuroscience; multivariate pattern classification in functional neuroimaging among them. Despite of it, to this date such analyses by an large have perpetuated localizationist presuppositions, by exploring one brain subregion at a time; or with no intention of deriving statistical maps to infer functional localization on populations (an important feature of the classical mass-univariate analysis).

This work explored the extent to which MVPA can overcome those limits and applied it to a number of problems, including the open problem of narrowing down the slippery substrate of emotional representation in the central nervous system. We asked ourselves whether it's possible to decode different basic emotions, the presence of face percepts and simpler visual distinctions based on the multidimensional patterns of brain activity, as measured with BOLD fMRI. If true, a classification algorithm might be able to distinguish brain images when perceiving one emotion or another, and neuroanatomical maps of the most relevant structures might be obtained from successful models.

Results of visual and face perception experiments demonstrate that commonplace functional MR imaging indeed can be analysed in this *mass and multivariate* fashion, with arguably better results than classical mass univariate analysis. Moreover, we discovered an anatomically-distributed pattern of information which apparently encodes for happiness; which the multivariate algorithm learned to identify well-beyond prediction levels by chance, whereas the univariate correlation analysis failed to replicate this. This suggests that at least certain forms of multivariate pattern classification analysis are a viable tool for mapping brain functions in a whole-brain, data-driven fashion; and not merely a tool for hard-to-interpret disease diagnosis and prediction of psychological states.

6 Declaration of Interest

The authors of these Research Objects certify that they have no affiliations with or involvement in any organization or entity with any financial, personal or professional interest in the materials presented in this submission. The authors declare that they have no known competing financial interests or personal relationships that may have influenced their research. I. David is a graduate student at the Master's program in neurobiology from the National Autonomous University of Mexico, for which he received a fellowship (CVU 891935) from CONACyT.

7 References

1. Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI)"brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage. 2003;19(2):261–70.

2. Haynes J-D, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature neuroscience. 2005;8(5):686–91.

3. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nature neuroscience. 2005;8(5):679–85.

4. Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. Proceedings of the National Academy of Sciences. 2006;103(10):3863–8.

5. Björnsdotter M, Rylander K, Wessberg J. A monte carlo method for locally multivariate brain mapping. Neuroimage. 2011;56(2):508–16.

6. Schmah T, Yourganov G, Zemel RS, Hinton GE, Small

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SL, Strother SC. Comparing classification methods for longitudinal fMRI studies. Neural computation. 2010;22(11):2729–62.

7. Kherif F, Poline J-B, Flandin G, Benali H, Simon O, Dehaene S, et al. Multivariate model specification for fMRI data. Neuroimage. 2002;16(4):1068–83.

8. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: The dangers of double dipping. Nature neuroscience. 2009;12(5):535.

9. Hoel EP, Albantakis L, Tononi G. Quantifying causal emergence shows that macro can beat micro. Proceedings of the National Academy of Sciences. 2013;110(49):19790–5.

10. De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. Neuroimage. 2008;43(1):44–58.

11. Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. Proceedings of the National Academy of Sciences. 2020;

12. Huettel SA, Song AW, McCarthy G, others. Functional magnetic resonance imaging. 2nd ed. Vol. 1. Sinauer Associates Sunderland, MA; 2009.

13. Mahmoudi A, Takerkart S, Regragui F, Boussaoud D, Brovelli A. Multivoxel pattern analysis for FMRI data: A review. Computational and mathematical methods in medicine. 2012;2012.

14. Lewis-Peacock JA, Norman KA. Multi-voxel pattern analysis of fMRI data. The cognitive neurosciences. 2013;911–20.

15. Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: A general linear approach. Human brain mapping. 1994;2(4):189–210.

16. Friston KJ, Frith CD, Frackowiak R, Turner R, others. Characterizing dynamic brain responses with fMRI: A multivariate approach. Neuroimage. 1995;2(2):166–72.

17. Diedrichsen J, Kornysheva K. Motor skill learning between selection and execution. Trends in cognitive sciences. 2015;19(4):227–33.

18. Zhang J, Zhang G, Li X, Wang P, Wang B, Liu B. Decoding sound categories based on whole-brain functional connectivity patterns. Brain imaging and behavior. 2020;14(1):100–9.

19. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. Neuroinformatics. 2014;12(2):229–44.

20. McIntosh A, Bookstein F, Haxby JV, Grady C. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage. 1996;3(3):143–57.

21. Mørch N, Hansen LK, Strother SC, Svarer C, Rottenberg DA, Lautrup B, et al. Nonlinear versus linear models in functional neuroimaging: Learning curves and generalization crossover. In: Biennial international conference on information processing in medical imaging. Springer; 1997. pp. 259–70.

22. McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, et al. Analysis of fMRI data by blind separation into independent spatial components. Human brain mapping. 1998;6(3):160–88.

23. Kjems U, Hansen LK, Anderson J, Frutiger S, Muley S, Sidtis J, et al. The quantitative evaluation of functional neuroimaging experiments: Mutual information learning curves. NeuroImage. 2002;15(4):772–86.

24. Carlson TA, Schrater P, He S. Patterns of activity in the categorical representations of objects. Journal of cognitive neuroscience. 2003;15(5):704–17.

25. LaConte S, Anderson J, Muley S, Ashe J, Frutiger S, Rehm K, et al. The evaluation of preprocessing choices in single-subject bold fMRI using npairs performance metrics. NeuroImage. 2003;18(1):10–27.

26. Mourão-Miranda J, Bokde AL, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. NeuroImage. 2005;28(4):980–95.

27. LaConte S, Strother S, Cherkassky V, Anderson J, Hu X. Support vector machines for temporal classification of block design fMRI data. NeuroImage. 2005;26(2):317–29.

28. Raizada RD, Tsao F-M, Liu H-M, Holloway ID, Ansari D, Kuhl PK. Linking brain-wide multivoxel activation patterns to behaviour: Examples from language and math. NeuroImage. 2010;51(1):462–71.

29. Pessoa L, Padmala S. Decoding near-threshold perception of fear from distributed single-trial brain activation. Cerebral cortex. 2007;17(3):691–701.

30. Ethofer T, Van De Ville D, Scherer K, Vuilleumier P. Decoding of emotional information in voice-sensitive cortices. Current Biology. 2009;19(12):1028–33.

31. Said CP, Moore CD, Engell AD, Todorov A, Haxby JV. Distributed representations of dynamic facial expressions in the superior temporal sulcus. Journal of vision. 2010;10(5):11–1.

32. Peelen MV, Atkinson AP, Vuilleumier P. Supramodal representations of perceived emotions in the human brain. Journal of Neuroscience. 2010;30(30):10127–34.

33. Kotz SA, Kalberlah C, Bahlmann J, Friederici AD, Haynes J-D. Predicting vocal emotion expressions from the human brain. Human Brain Mapping. 2012;34(8):1971–81.

34. Skerry AE, Saxe R. Neural representations of emotion are organized around abstract event features. Current biology. 2015;25(15):1945–54.

35. Wegrzyn M, Riehle M, Labudda K, Woermann F, Baumgartner F, Pollmann S, et al. Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. Cortex. 2015;69:131–40.

36. Hernández-Pérez R, Concha L, Cuaya LV. Decoding human emotional faces in the dog's brain. bioRxiv. 2018;134080.

37. Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends in cognitive sciences. 2000;4(6):223–33.

38. Haist F, Anzures G. Functional development of the brain's face-processing system. Wiley Interdisciplinary Reviews: Cognitive Science. 2017;8(1-2):e1423.

39. Vytal K, Hamann S. Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. Journal of cognitive neuroscience. 2010;22(12):2864–85.

40. Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF. The brain basis of emotion: A meta-analytic review. Behavioral and brain sciences. 2012;35(3):121–43.

41. Hamann S. Mapping discrete and dimensional emotions onto

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the brain: Controversies and consensus. Trends in cognitive sciences. 2012;16(9):458–66.

42. Kragel PA, LaBar KS. Advancing emotion theory with multivariate pattern classification. Emotion Review. 2014;6(2):160–74.

43. Guillory SA, Bujarski KA. Exploring emotions using invasive methods: Review of 60 years of human intracranial electrophysiology. Social cognitive and affective neuroscience. 2014;9(12):1880–9.

44. Kragel PA, LaBar KS. Decoding the nature of emotion in the brain. Trends in cognitive sciences. 2016;20(6):444–55.

45. Celeghin A, Diano M, Bagnis A, Viola M, Tamietto M. Basic emotions in human neuroscience: Neuroimaging and beyond. Frontiers in Psychology. 2017;8:1432.

46. Rolls ET, Grabenhorst F, Franco L. Prediction of subjective affective state from brain activations. Journal of Neurophysiology. 2009;101(3):1294–308.

47. Baucom LB, Wedell DH, Wang J, Blitzer DN, Shinkareva SV. Decoding the neural representation of affective states. Neuroimage. 2012;59(1):718–27.

48. Chikazoe J, Lee DH, Kriegeskorte N, Anderson AK. Population coding of affect across stimuli, modalities and individuals. Nature neuroscience. 2014;17(8):1114.

49. Shinkareva SV, Wang J, Kim J, Facciani MJ, Baucom LB, Wedell DH. Representations of modality-specific affective processing for visual and auditory stimuli derived from functional magnetic resonance imaging data. Human brain mapping. 2014;35(7):3558–68.

50. Chang LJ, Gianaros PJ, Manuck SB, Krishnan A, Wager TD. A sensitive and specific neural signature for picture-induced negative affect. PLoS biology. 2015;13(6):e1002180.

51. Sitaram R, Lee S, Ruiz S, Rana M, Veit R, Birbaumer N. Real-time support vector classification and feedback of multiple emotional brain states. Neuroimage. 2011;56(2):753–65.

52. Kassam KS, Markey AR, Cherkassky VL, Loewenstein G, Just MA. Identifying emotions on the basis of neural activation. PloS one. 2013;8(6):e66032.

53. Saarimäki H, Gotsopoulos A, Jääskeläinen IP, Lampinen J, Vuilleumier P, Hari R, et al. Discrete neural signatures of basic emotions. Cerebral cortex. 2015;26(6):2563–73.

54. Kragel PA, LaBar KS. Multivariate neural biomarkers of emotional states are categorically distinct. Social cognitive and affective neuroscience. 2015;10(11):1437–48.

55. Peirce JW. PsychoPy—psychophysics software in python. Journal of neuroscience methods. 2007;162(1-2):8–13.

56. Ekman P. Pictures of facial affect. Consulting Psychologists Press. 1976;

57. Mustra M, Delac K, Grgic M. Overview of the DICOM standard. In: 2008 50th International Symposium ELMAR. IEEE; 2008. pp. 39–44.

58. Cox R, Ashburner J, Breman H, Fissell K, Haselgrove C, Holmes C, et al. A (sort of) new image data format standard: NIfTI-1: WE 150. Neuroimage. 2004;22.

59. Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific data. 2016;3(1):1–9.

60. Bedetti C, Carlin J, Joseph M, Isla, Stojic H, Kastman E. Cbedetti/dcm2bids 2.1.4 [Internet]. Zenodo; 2019. Available from: https://doi.org/10.5281/zenodo.2628725

61. Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. Journal of neuroscience methods. 2016;264:47–56.

62. Gulban OF, Nielson D, Poldrack R, lee, Gorgolewski C, Vanessasaurus, et al. Poldracklab/pydeface: V2.0.0 [Internet]. Zenodo; 2019. Available from: https://doi.org/10.5281/zenodo.3524401

63. Manjón JV, Coupé P. VolBrain: An online MRI brain volumetry system. Frontiers in neuroinformatics. 2016;10:30.

64. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage. 2012;62(2):782–90.

65. Woolrich MW, Ripley BD, Brady M, Smith SM. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage. 2001;14(6):1370–86.

66. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Medical image analysis. 2001;5(2):143–56.

67. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 2002;17(2):825–41.

68. Fonov VS, Evans AC, McKinstry RC, Almli C, Collins D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage. 2009;(47):S102.

69. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, et al. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage. 2011;54(1):313–27.

70. Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S. PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics. 2009;7(1):37–53.

71. Vapnik V, Chervonenkis A. Theory of pattern recognition. Nauka, Moscow; 1974.

72. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on computational learning theory. 1992. pp. 144–52.

73. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. Neuroimage. 2014;92:381–97.

74. Smith SM, Nichols TE. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage. 2009;44(1):83–98.

75. Roiser J, Linden D, Gorno-Tempinin M, Moran R, Dickerson B, Grafton S. Minimum statistical standards for submissions to neuroimage: Clinical. NeuroImage: Clinical. 2016;12:1045.

76. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: Judgments from untrained research participants. Psychiatry research. 2009;168(3):242–9.

77. Conley MI, Dellarco DV, Rubien-Thomas E, Cohen AO, Cervera A, Tottenham N, et al. The racially diverse affective expression (RADIATE) face stimulus set. Psychiatry research. 2018;270:1059–67.

78. Schmahmann JD, Sherman JC. Cerebellar cognitive affective syndrome. International review of neurobiology. 1997;41:433–

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.04.438425; this version posted April 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

40.

79. Vinck M, Batista-Brito R, Knoblich U, Cardin JA. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron. 2015;86(3):740–54.

80. Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. Spontaneous behaviors drive multidimensional, brainwide activity. Science. 2019;364(6437):eaav7893.