NeuroImage 243 (2021) 118510

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Brain state kinematics and the trajectory of task performance improvement

Eli J. Müller[a,*], Brandon Munn[a], Holger Mohr[c,d], Hannes Ruge[c], James M. Shine[a,b]

[a] Brain and Mind Centre, The University of Sydney, Sydney, NSW, Australia
[b] Complex Systems Research Group, The University of Sydney, Sydney, NSW, Australia
[c] Department of Psychology, Technische Universität Dresden, Dresden, Germany
[d] Neuroimaging Center, Technische Universität Dresden, Dresden, Germany

A B S T R A C T

Dimensionality reduction techniques offer a unique perspective on brain state dynamics, in which systems-level activity can be tracked through the engagement of a small number of component trajectories. Used in combination with neuroimaging data collected during the performance of cognitive tasks, these approaches can expose the otherwise latent dimensions upon which the brain reconfigures in order to facilitate cognitive performance. Here, we utilized Principal Component Analysis to transform parcellated BOLD timeseries from an fMRI dataset in which 70 human subjects performed an instruction based visuomotor learning task into orthogonal low-dimensional components. We then used Linear Discriminant Analysis to maximise the mean differences between the low-dimensional signatures of fast-and-slow reaction times and early-and-late learners, while also conserving variance present within these groups. The resultant basis set allowed us to describe meaningful differences between these groups and, importantly, to detail the patterns of brain activity which underpin these differences. Our results demonstrate non-linear interactions between three key brain activation maps with convergent trajectories observed at higher task repetitions consistent with optimization. Furthermore, we show subjects with the greatest reaction time improvements have delayed recruitment of left dorsal and lateral prefrontal cortex, as well as deactivation in parts of the occipital lobe and motor cortex, and that the slowest performers have weaker recruitment of somatosensory association cortex and left ventral visual stream, as well as weaker deactivation in the dorsal lateral prefrontal cortex. Overall our results highlight the utility of a kinematic description of brain states, whereby reformatting data into low-dimensional trajectories sensitive to the subtleties of a task can capture non-linear trends in a tractable manner and permit hypothesis generation at the level of brain states.

1. Introduction

An organism's behaviour can be viewed as a low-dimensional distillation of its underlying brain activity, since even at higher levels of abstraction, other behaviours are always possible whenever one is manifest. These behaviours result from interdependencies between neurons, where neural activity is correlated across multiple spatial and temporal scales – much less than the dimensional upper bound of a purely uncorrelated brain (Shine, 2020). Indeed, recent progress in neuroscience has exploited the existence of low-dimensional brain activity (Cunningham and Yu, 2014a), and connectivity (Avena-Koenigsberger et al., 2018; Bullmore and Sporns, 2012), resolving persistent large-scale patterns indicative of underlying principles of organisation (Margulies et al., 2016).

In the context of learning, the brain is observed to transition between two partially overlapping phases: the so called *exploration* phase, in which brain states are more varied as an organism samples the unknown parameters of a task; and the *exploitation* phase, whereby the set of task parameters are known and a period of brain state refinement and optimisation occurs (Cohen et al., 2007). Despite many studies demonstrating the heuristic value of this *exploration-exploitation* model, quantitative methods sensitive to the subtleties of brain configurations specifically during phases of task exploitation that are still able to capture the associated convergence of brain states (Sadtler et al., 2014) have not been well explored.

There have been a number of studies that have investigated aspects contributing to the *exploration-exploitation* paradigm at a systems level. For instance, it has been shown that (over-)learning in a simple motor task during an extended training period can induce a shift in modular architecture that functionally partitions the somatomotor and visual systems (Bassett et al., 2015), with task performance improvements involving the liberation of cognitive control hubs – specifically frontal and cingulate cortices (Bassett et al., 2015). More generally, the flexible recruitment of fronto-parietal hubs demonstrate their central role in adaptive implementation of task demands (Cole et al., 2013). Further studies have demonstrated short-term automatization in instruction-based visuomotor learning tasks is associated with rapidly changing brain dynamics on multiple levels, including decreased activation of the fronto-parietal network, increased cortico-striatal integration (Ruge et al., 2019), and a reconfiguration of large-scale cortical networks between integrated and segregated processes (Mohr et al., 2016, 2018; Sheffield et al., 2021). It remains to be determined precisely how these brain state features dynamically evolve within the context of a task, and what their dynamics can tell us about task performance.

Dimensionality reduction techniques offer a unique perspective on brain state dynamics, and expose an axis for constructing a reduced

* Corresponding author.
E-mail address: eli.muller@sydney.edu.au (E.J. Müller).

https://doi.org/10.1016/j.neuroimage.2021.118510.
Received 10 July 2021; Accepted 23 August 2021
Available online 26 August 2021.

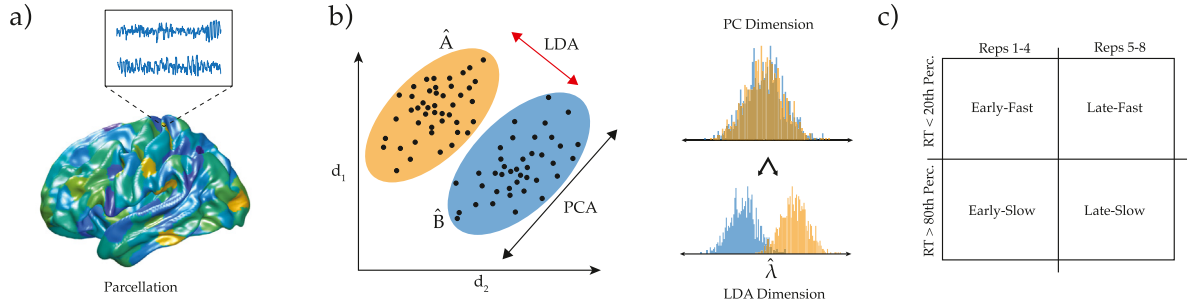

Fig. 1. Linear Discriminant Analysis. (a) Extracting parcellation timeseries. b) Comparison of Linear Discriminant Analysis and Principal Component Analysis. In Linear Discriminant Analysis, data are first divided into desired classes, and then LDA produces eigenvectors, weighting the input dimensions, which maximize the distance between the class mean while conserving variance within each class. Principal Component Analysis produces eigenvectors, weighting the input dimensions, which maximally capture covariance in the data. c) The 4 class divisions used in the LDA applied to the present dataset.

system from high-dimensional multivariate brain datasets that is often more amenable to subsequent analyses (Cunningham and Yu, 2014a). In addition, these reduction techniques can reveal persistent structures (in space, time, or both), and thus devalue independent features unique to a subsystem, such as local noise. More broadly, low-dimensional descriptions of neural activity have elucidated unique brain states across cognitive tasks (Shine et al., 2019a) that can also provide a useful basis for capturing changes in brain states within a task. In particular, these low dimensional descriptions can disassociate performance errors in a novel cognitive task as a deviation from a task-specific manifold (Shine et al., 2019b).

In this study, we leveraged dimensionality reduction techniques to quantify the convergent trajectories of brain states during the exploitation phase of a simple instruction-based visuomotor learning task. Specifically, we utilize Principal Component Analysis (PCA) to transform BOLD (blood oxygen level dependant) data into orthogonal components that maximised the major axes of variance across the dataset (Fig. 1a). We then used Linear Discriminant Analysis (LDA; Fig. 1b) to identify linear combinations of the low-dimensional components that maximised the mean differences between fast-and-slow performance and early-and-late learners (Fig. 1c), while conserving variance present within these groups. Our results were qualitatively similar for moderate changes to these boundaries. The new bases that result from LDA provided a means to maximally describe the differences between these groups and, importantly, contrast patterns of brain activity which underpin these differences. Overall, our analysis highlights the utility of a kinematic description of brain states, whereby reformatting data into low-dimensional trajectories sensitive to the subtleties of a task can capture non-linear trends in a tractable manner and permit hypothesis generation at the level of brain states.

2. Materials and methods

2.1. Learning task

In the instruction-based learning task, subjects were asked to practice stimulus-response associations between four symbols and two responses – left/right-hand index finger button presses (Mohr et al., 2016). Before practice, S–R associations were explicitly instructed. To this end, an instruction screen was presented for 10 s showing four symbols simultaneously, two on the right side and two on the left side, with side indicating the required response. The instruction screen was followed by a sequence of single trials (practice phase). In each trial, one of the four symbols was presented and subjects had to respond as instructed (i.e., with a left or right button press). Symbols were presented in randomized order. Feedback was given after a response or after maximally 1.5 s. In case of a correct response, the symbol was highlighted in a colour for 0.5 s, whereas after an erroneous response (or after 1.5 s elapsed without response), the symbol was highlighted in grey. In case of an erroneous response (or miss), the trial was repeated, and both trials were excluded from analysis. Trials were jittered with randomized 0.8 or 3.5 s inter-trial intervals. From each symbol, eight correctly performed trials were collected, that is, 32 trials per stimulus set. The whole procedure (that is, instruction screen and practice phase) was repeated 20 times, each time using a novel set of symbols. Note that the data from this task were previously reported in Mohr et al., 2016.

2.2. fMRI scanning

Functional and structural images were acquired on the same Siemens 3 T Trio Scanner equipped with a 16-channel circularly polarized head coil. A gradient echo planar sequence with repetition time (TR) = 2 s, echo time (TE) = 30 ms and flip = 80° was used for functional imaging. Volumes consisted of 26 slices with an in-plane resolution of 4 × 4 mm and a thickness of 5 mm. Presentation 12.0 (Neurobehavioral Systems) software was used to run the experiment. Structural images were also obtained but were only used for neuroradiological assessment in the current study (Mohr et al., 2016). Each subjects fMRI scan was recorded in a single ~30 min session.

2.3. Pre-processing

Pre-processing of functional data was performed with SPM8 running in Matlab 7.12. Pre-processing consisted of slice-time correction, rigid body movement correction (three translation and three rotation parameters), normalization of the mean functional image to the SPM MNI EPI template (resampling to 3 × 3 × 3 mm resolution) and smoothing with a Gaussian kernel, full width at half maximum = 8 mm.

2.4. Brain parcellation

Following pre-processing, the mean time series was extracted from 375 pre-defined regions of interest (ROIs) as defined in (Shine et al., 2016). To ensure whole-brain coverage, these ROIs included 333 cortical parcels (161 and 162 regions from the left and right hemispheres, respectively) from the Gordon atlas (Gordon et al., 2016),14 subcortical regions from Harvard-Oxford subcortical atlas (bilateral thalamus, caudate, putamen, ventral striatum, globus pallidus, amygdala, and hippocampus), and 28 cerebellar regions from the SUIT atlas (Diedrichsen et al., 2009). A total of 354 ROIs were used with 21 regions excluded from all analysis due to misalignment with some subject's brain masks. These have the following indices in the parcel scheme mentioned above: [115, 123, 124, 125, 134, 278, 280, 285, 286, 287, 288, 303, 356, 358, 360, 363, 364, 367, 368, 369, 375].

2.5. Design-matrix modelling

We extracted each trial response by modelling the haemodynamic response function (HRF) at each trial onset and averaging the follow-

ing 3rd and 4th timepoints of the BOLD signal corresponding to a peak in the HRF (TR of 2 s). We noted that due to the temporal resolution of the scanner and temporal proximity of trials, the BOLD response of some trials may overlap. A sophisticated model pipeline could be used to disambiguated effects unique to each trial, however, they may also introduce spurious regression artefacts. For the present analysis we accept this limitation of the data and note its role in potentially weakening inter-trial differences during analysis.

2.6. Principal component analysis

The complete pre-processed and parcellated data from all 70 subjects were concatenated into a single regional timeseries and spatial PCA was performed on the resultant matrix using MATLAB's *pca* function. The top 35 components (selection criteria for this number is detailed below) explained ~84% of the variance in the data.

2.7. Criteria for separating data into classes

In the following section, we perform a Linear Discriminant Analysis that requires specification of unique classes of interest within a dataset. To this end, the complete dataset used in this study consisted of 70 subjects recorded across 20 learning blocks, each comprising 8 repetitions of four stimuli . This results in ~60 trials (80 minus incorrect trials) for each repetition (8) per subject (70). The data were then split into four groups independent of subject: Early-fast (repetitions 1–4, reaction time < 20th percentile); Early-slow (repetitions 1–4, reaction time > 80th percentile); Late-fast (repetitions 5–8, reaction time < 20th percentile); and Late-slow (repetitions 4–8, reaction time > 80th percentile). There were ~3500 data points in each group. The probability distribution of each subjects trials across the four groups was 24.8 ± 0.01, 23.9 ± 0.01, 26.2 ± 0.02 and 25.2 ± 0.02 with $\chi^2 = 17.9$ ($p = 1$). The repetition divisions were chosen so as to maximise the amount of data retained and ensure early and late repetitions did not overlap, and the reaction time divisions were chosen so as to highlight atypical performance, i.e., sufficient distance above and below the mean. Supp. Fig. 2 shows the results that follow are consistent with significant changes to these data divisions.

Splitting the data in half and performing all analysis on both sets produced qualitatively similar results.

2.8. Linear discriminant analysis

We utilized Linear Discriminant Analysis (LDA)(Fisher, 1936) to separate the task data into 4 distinct classes: early-fast, early-slow, late-fast, and late slow. This method provides a linear transformation into an orthogonal basis (independent eigenvectors) such that the difference between the mean of each class is maximised. In order to protect against overfitting to artefacts or noise, we performed LDA on the PC coefficient timeseries by projecting the data onto the top 35 principal components. Since the higher order PCs contain weaker contributions to overall covariance, they are most susceptible to spatially localized noise and are more unique to a given dataset. Thus, we include the minimum amount of the variance modes that give the most robust eigenmaps following LDA. The criteria for selecting the top 35 PCs used in the final LDA algorithm (as detailed Supp. Fig. 1a) consisted of a balance between an elbow in cumulative explained variance, minimising PC dimensionality, and the stability in the resulting LDA eigenvectors. This PCA-LDA pipeline has been shown more successful in classifying fMRI response to naturalistic movie stimuli (Mandelkow et al., 2016) when compared to solely using LDA without PCA. The LDA algorithm is defined as follows.

In order to separate the PC coefficient data $x \in X$ into 4 classes, we specify each class as a matrix ω_i with rows corresponding to N_i observations and columns to the included number of variables to separate – i.e., 35 PC coefficients. The *within-class* scatter matrix is first defined as the sum of each classes covariance matrix

$$S_W = \sum_{i=1}^{4} S_i \tag{1}$$

where

$$S_i = \sum_{x \in w_i} (x - \mu_i)(x - \mu_i)^T \tag{2}$$

are the covariance matrices and

$$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x \tag{3}$$

are the class means. The *between-class* scatter matrix is then defined as the sum of the covariance matrix of each classes mean to the global mean

$$S_B = \sum_{i=1}^{4} N_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{4}$$

where

$$\mu = \frac{1}{N} \sum_{x} x \tag{5}$$

is the mean across all four classes.

To find the eigenvectors $\bar{w}$ that maximally differentiate the four classes, we solve the eigenvalue problem $S_w^{-1} S_B \bar{w} = \lambda \bar{w}$ using MATLAB's *eig* function. To ensure each eigenvector captures a unique aspect of the class differences (i.e., are independent), we orthonormalize the resulting eigenvectors using the Modified Gram-Schmidt algorithm (Pursell and Trimble, 2018).

The result of the LDA, performed on the PC coefficient timeseries of the 4 classes (Early-Fast, Early-Slow, Late-Fast, and Late-Slow), is a set of 3 dominant eigenvectors ($\lambda_1, \lambda_2, \lambda_3$) which map the 35 PCs to a corresponding eigenvalue. The eigenvectors define a subspace that maximizes the difference between the 4 classes while conserving the variance within each class. By taking these 3 eigenvectors and projecting the complete ROI data (including those data not in the 4 classes used during LDA), first into its principal components, and then into the LDA eigenvectors via the inner product, the data is effectively reduced from the original 354 dimensions to 3 dimensions. Additionally, since the PCs form an orthonormal basis, i.e., their inverse is equal to their transpose, the LDA eigenvectors can be back-projected (via the dot product with the transpose of the PC matrix) to give an ROI loading.

In contrast to traditional GLM modelling, where reaction time is typically used as a regressor to generate a static spatial map (e.g., (Yarkoni et al., 2009)), LDA is designed to identify a linear relationship that maximizes group average differences whilst conserving variability within each specified class of a dataset. Projecting the data into these resulting eigenbases may elucidate non-linear trends, and indeed the trajectory analysis presented below demonstrates an inherently non-linear relationship between BOLD activations across the task data (Fig. 2d). For later comparison with the results of LDA we calculate GLM beta maps for reaction time and repetition regressors and these are presented in Supp. Fig. 4.

The LDA eigenvectors used here are orthogonalized and as such provide unique (independent) information about group differences in the data. Whilst it can be useful to look at the loading on only a single eigenvector, the superposition of these vectors fully define the group differences and can be projected back to the original ROI space producing interpretable regional maps similar to traditional statistical parametric maps.

2.9. Subject performance groupings

Importantly, in this study Linear Discriminant Analysis is run on four groupings of the data irrespective of the subject the data was recorded from, i.e., a given subject may have data in all four groups. This ensures

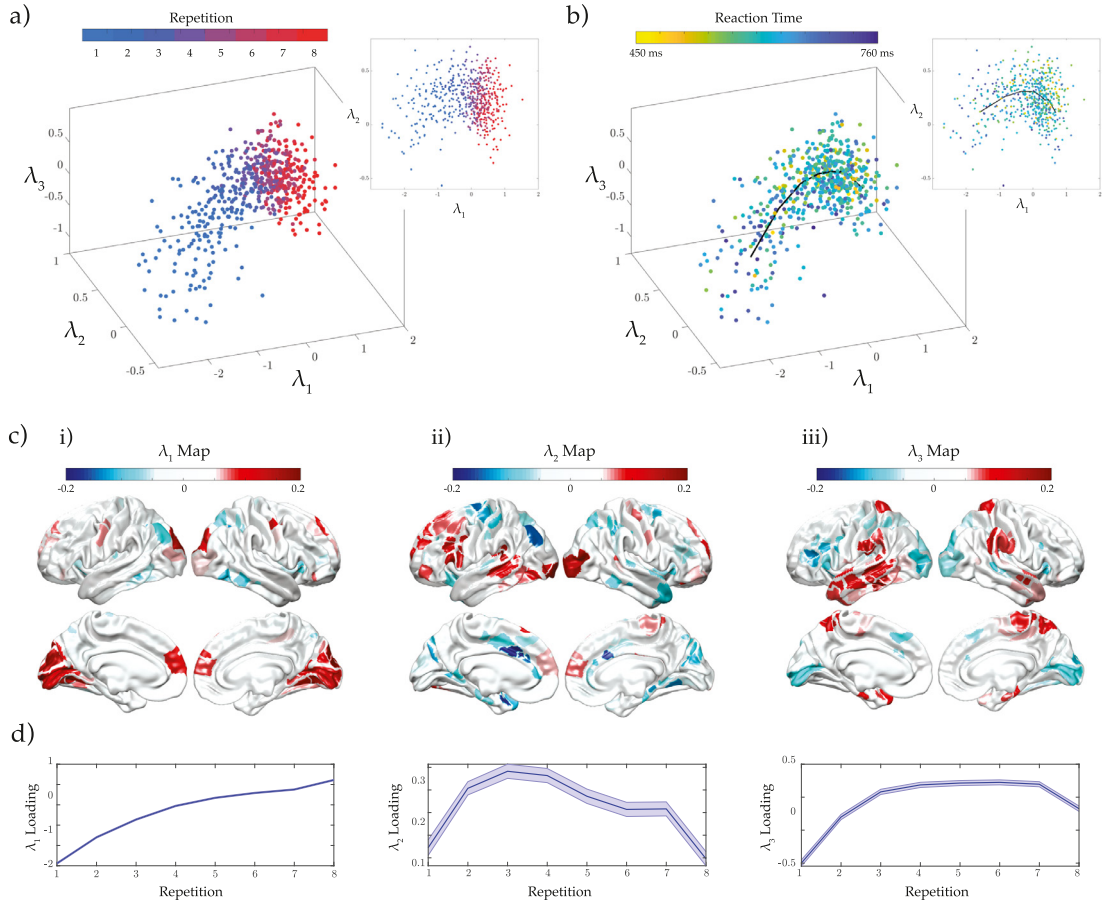

Fig. 2. LDA results. a) Trial-averaged subject loadings for each repetition. Colormap indicates the repetition number. b) Trial-averaged subject loadings for each repetition. Colormap indicates trial average reaction time. The black line shows the repetition trajectory averaged across subjects. c) Unthresholded spatial maps of the 3 eigenvectors i)-iii) resulting from LDA. d) Repetition trajectory averaged across subjects for each of the 3 eigenmaps. Shaded areas show standard error of the mean.

the resulting eigenmaps separate population-level BOLD activations patterns. In order to investigate whether these maps can provide insight into specific-subject trajectories across repetitions, each subject's average reaction time across repetitions 1–2 was subtracted from their average over repetitions 4–6 as a measure of performance improvement. Then, the average reaction time is computed over all repetitions as a measure of general performance. Subject trajectories were then sorted into 3 groupings: Slow - performance improvement $\langle$ 33rd percentile and average RT $\rangle$ 67th percentile; Fast - performance improvement $\langle$ 33rd percentile and average RT $<$ 33rd percentile; and Improvers - performance improvement $\rangle$ 67th percentile.

2.10. Visualisation on surface

We define a task manifold that aids visual comparison of the 3-dimensional trajectories across the task. In order to generate this manifold a general linear model (GLM) was fit to projected data for each repetition (1–8) using reaction time as a regressor. This produced a 3-dimensional vector, per repetition, within the embedding space that mapped to performance. These 8 vectors were then normalized by their magnitude to produce 8 direction vectors in the embedding space that corresponded to activity correlated with decreasing reaction time. That is, each of the 8 repetitions are given a unique direction in state space which corresponded with reduced reaction time. The origin of each direction vector is set to the population average loading for the corresponding repetition, and then an arbitrary extrapolation of ± 0.2 is made. Finally, a task manifold is produced by interpolating these 8 lines in order to form a surface.

3. Results

3.1. Characterisation of task performance

The error rate for subjects selecting the appropriate response (left/right hand) following the instruction phase was always ~10% (and typically much smaller, see Supp. Fig. 6b) after the first repetition of each unique set of instructions. This suggests that, for this particular task, a truncated *exploration* phase occurs wherein subjects are able to successfully complete the task very early on within the course of each trial. For this reason, the remaining repetitions of each instruction were considered to form part of an *exploitation* phase, during which the brain is likely optimising its configuration resulting in successful response. The trial average for each of the 8 repetitions per subject was used as a measure of task performance. The overall mean RT for the learning task was 580 ± 140 ms. The distribution of RTs within each repetition was similar, with an observed decrease in the coefficient of variation from early to late repetitions, see Supp. Fig. 6.

3.2. Linear discriminant analysis

Following the projection of the data onto its first 35 PCs, and using the criteria detailed in Section 2.5, we split the data into 4 distinct classes (as shown in Fig. 1c): Early-Fast (repetitions 1–4; reaction time $<$ 20th percentile), Early-Slow (repetitions 1–4; reaction time $>$ 80th

percentile), Late-Fast (repetitions 5–8; reaction time < 20th percentile), Late-Slow (repetitions 4–8; reaction time > 80th percentile). As detailed in the Methods (Section 2.8), the result of the LDA is a set of 3 dominant eigenvectors ($\lambda_1, \lambda_2, \lambda_3$) which map the 354 ROIs to a corresponding set of 3 eigenvalues. The robustness of these eigenmaps to changes in the arbitrary class divisions used here are detailed in Supp. Fig. 2. Notably, the maps are consistent for significant changes to these divisions with λ_3 shown to be the most sensitive. Furthermore, a split-half comparison was performed by running LDA on two random halves of the data for 100 iterations, and showed a mean correlation between each subsets eigenmaps of $\lambda_1 : r = 0.96,\ p = 10^{-128}$, $\lambda_2 : r = 0.92,\ p = 5 \times 10^{-97}$, and $\lambda_3 : r = 0.27,\ p = 0.03$.

Using this approach, the LDA embedding for each subject was averaged across trials and blocks such that each subject had 8 data points (i.e., one for each repetition). Fig. 2a shows the distribution of these repetition averages with a clear trajectory from early-to-late and this corresponds predominantly to loading onto the first LDA map - λ_1(as shown in Fig. 2d). Fig. 2b shows that RT improvement does follow this same embedding trajectory. Overall, the trajectories show a convergence at middle to late repetitions (reps 3–7), with a general deceleration to low velocity in all 3 dimensions.

3.3. LDA eigenmaps

Fig. 2c shows the surface projections that correspond to the 3 resulting eigenvectors ($\lambda_1, \lambda_2, \lambda_3$) from the LDA. The first eigenmap – λ_1 (Fig. c-i) – had the highest discriminatory power between the 4 classes, with a dominant positive weighting for medial visual cortex, and dorsal and medial prefrontal cortex. The second eigenmap – λ_2 (Fig. c-ii) – had broad positive weightings in the left-hemisphere dorsal prefrontal and premotor cortices, in addition to positive weightings in early ventral visual stream and a contrasting negative weighting for dorsal visual stream. The third eigenmap – λ_3 (Fig. c-iii) – had a positive weighting for left-hemisphere ventral visual stream, in addition to negative weightings in the dorsal lateral prefrontal cortex and medial visual cortex. The λ_3 map also demonstrated high positive weightings for bi-hemispheric somatosensory association cortex.

3.4. LDA eigenmap loadings

The implications of these 3 eigenmaps are best observed by considering their population average loadings as a function of repetition number (Fig. d)). λ_1 loading had the highest amplitude across repetitions, suggesting task duration coincided with activation within the medial visual and medial prefrontal cortices. The λ_2 eigenmap demonstrated an inverted-U relationship with repetition count, and hence highlights regional activations that are unique for intermediate repetitions. λ_3 shows a decelerating trajectory that converges for late repetitions. Importantly, since each eigenvector contributes a unique (independent) piece of the trajectory, the complete activation map results from their superposition. This means that loading onto λ_3 suppresses the medial visual activation seen in the λ_1 loading. Additionally, both λ_2 and λ_3 positively superimpose to give strong ventral stream activation in the left hemisphere, and the high prefrontal cortical activation observed in λ_2 is constrained more dorsally due to the prefrontal deactivation observed in λ_3. Finally, a bi-hemispheric activation of the somatosensory association cortex, from the λ_3 map, plays a key role in task performance with a convergence to high loadings across the middle-to-late repetitions.

3.5. RT model

For each repetition, a general linear model (GLM) was fit using reaction time as a regressor to produce a direction vector within the embedding space mapping performance. These RT direction vectors are unique to each repetition and can be combined, in conjunction with the early-late trajectory averaged across subjects, to define a task manifold shown in Fig. 3a. The results of Fig. 3a distils those of Fig. 2a and 2b and clearly demonstrate a non-linear shift in brain activity from early repetitions to late repetitions in addition to a reconfiguration of brain activity underpinning reaction time, indicated by the folding and twisting of the task manifold.

3.6. Subject performance

In order to investigate differences between subject-specific trajectories, each subjects' average reaction time across repetitions 1–2 was subtracted from their average over repetitions 4–6 and taken as a measure of performance improvement. Then, an additional subject average reaction time was computed over all repetitions as a measure of general performance. Subject trajectories were then sorted into 3 groupings: Slow - performance improvement ⟨ 33rd percentile and average RT ⟩ 67th percentile, Fast - performance improvement ⟨ 33rd percentile and average RT < 33rd percentile, and Improvers performance improvement ⟩ 67th percentile.

The average trajectory of the 3 groups is shown in Fig 3b. All 3 groups have a similar progression in λ_1, which agrees with its correlation to repetition count (Fig. 2a)). λ_2 shows that the overall Fast and Slow reactors recruited dorsal lateral prefrontal cortex and the ventral visual stream (Fig. 2c ii) in a manner similar for early repetitions, whereas the improvers had a relatively delayed activation of these regions and a compensatory high activation late. The statistical significance of this difference is shown by the non-overlapping standard errors of each groups mean trajectory. λ_3 loadings clearly distinguish the Slow trajectories from the group with an up to 50% lower loading during middle repetitions. Importantly, this means that Slow performers had weaker bi-hemispheric activation of the somatosensory association cortex known to play a key role in object recognition (Reed et al., 2004). Together, these results suggest that activation of the λ_2 map is a measure of performance stability since both fast and slow subjects had similar trajectories in this dimension, and that activation of the λ_3 map is indicative of performance optimization since both fast subjects and those showing the greatest improvement have qualitatively similar and convergent trajectories.

3.7. General linear modelling

For comparison, we ran a more traditional approach by calculating the GLM beta maps of ROI BOLD activity using reaction time and repetition regressors, as shown in Supp. Fig.2a-b. Notably these maps demonstrated a faster reaction time corresponded to bi-hemispheric deactivation of premotor and somatomotor associations cortices, both medially and laterally, as well as dorsal lateral prefrontal cortices. And repetition corresponded to activation of the medial visual and prefrontal cortices, as well as the temporal pole. Since PCA is used before LDA in the current methodology we additionally run both GLMs (RT and repetition) using the first 35 principal components and back-projecting the resulting beta maps (defined in PC-space) into the ROI space, shown in Supp. Fig. 2c-d. The truncated repetition beta map showed a correlation with the first LDA eigenmap λ_1, of $r = 0.66,\ p = 4 \times 10^{-49}$ and the truncated RT beta map showed a correlation with the second LDA eigenmap λ_2, of $r = 0.43,\ p = 1 \times 10^{-18}$. This supports the overall utility of the LDA approach presented here, whereby it is able to both capture insights from a more traditional method, whilst elaborating on performance trajectories across the learning task to give a more nuanced description of non-linear trends in the data.

3.8. Comparison to 1-back control task

The dataset used in this study (Mohr et al., 2016) also contained a control dataset where stimulus material and temporal structure were identical to the learning task, however, the subjects were instead required to perform a 1-back working memory task. Using an identical

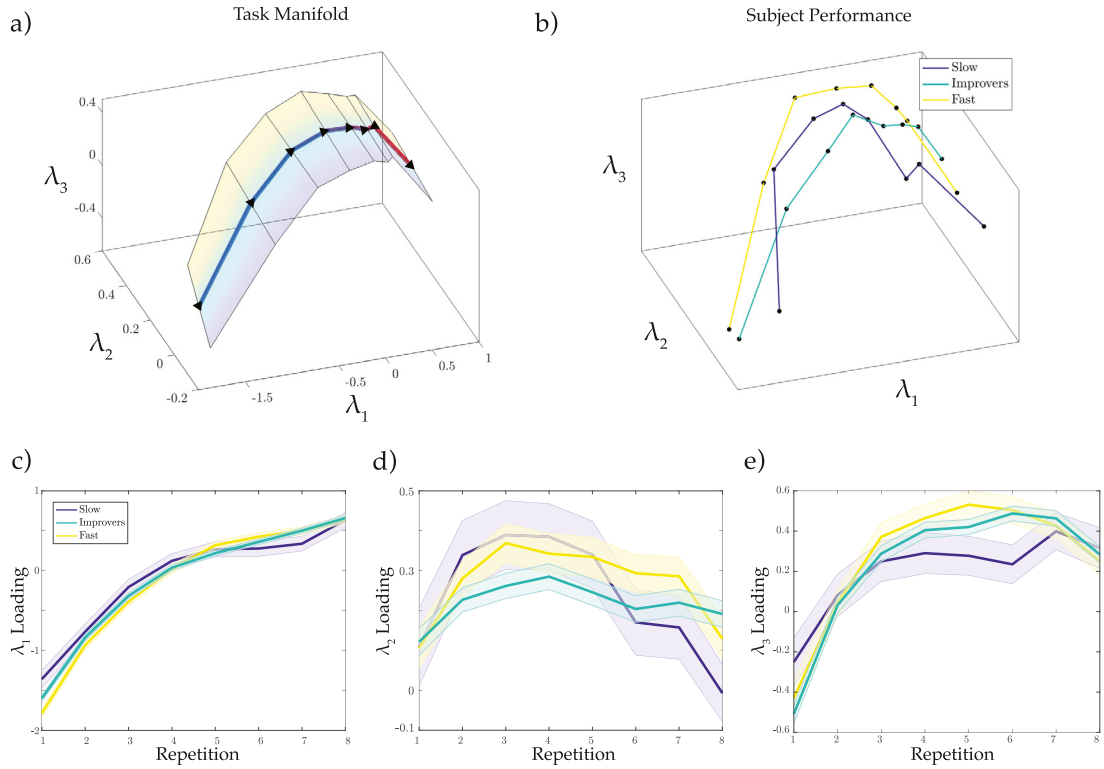

Fig. 3. Subject Performance. a) Task manifold generated by first computing the repetition trajectory as the subject-averaged loadings for each repetition (blue 1 – red 8). A general linear model is then fit to each repetition for all subject loadings using reaction time as a regressor. The resultant beta weights are normalized to define a direction vector, indicative of performance, and this is then centred on the repetition trajectory. For visualization the direction vectors are extrapolated ±0.2. b) Average repetition trajectory for 3 groupings: Slow (RT improvement ⟨ 33rd percentile and average RT ⟩ 67th percentile), Fast (RT improvement < 33rd percentile and average RT < 33rd percentile), and Improvers (RT improvement < 67th percentile). c) LDA 1 repetition average loadings for the 3 groupings. Shaded areas show standard error of the mean. d) LDA 2 repetition average loadings for the 3 groupings. e) LDA 3 repetition average loadings for the 3 groupings.

analysis pipeline to that from the learning analysis, i.e., first running PCA and performing LDA on the first 35 principal components, it can be seen in Supp. Fig. 5a that two of the resulting eigenmaps are distinct from those found for the learning group. The first eigenmap from the learning task has a strong correspondence to the second eigenmap from the control group $r = 0.53,\ p = 8\ \times 10^{-29}$ and is consistent with both maps correlating with to their respective task repetition.

4. Discussion

Our results demonstrate the utility of applying a trajectory analysis to neuroimaging data. Specifically, we highlight key differences across a spectrum of task performance within an otherwise seemingly simple neuropsychological task by leveraging a combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The embedding space that results from this technique is defined by a set of basis brain maps that detail independent components contributing to these key differences while remaining sensitive to the subtleties within it.

The task data analysed here entailed fMRI BOLD recordings captured during an instruction-based learning task with left-right button presses prescribed to each of four patterns shown in the instruction phase. By first projecting the data into its principal components, LDA was used to separate the data into 4 groups: fast-early, slow-early, fast-late, slow-late. This results in a 3 dimensional embedding space optimally sensitive to differences between these groups. Additionally, this embedding space is defined by an orthonormal basis which allows straightforward projections back to the ROI brain space for interpretability. Projecting subject data into LDA defined space and computing the population average trajectory across the eight repetitions showed a convergence towards a more stable brain state configuration. Since trial error rates were always ~10% (and much smaller for most subjects) after the first repetition, this convergence betrays a phase of optimization (or exploitation) wherein subjects have identified the task parameters and thus act to refine their behaviour and its underlying brain activity. Notably, subjects were not encouraged to perform this task as fast as possible and thus this optimization is likely motivated by energy efficiency constraints and resource liberation (Bassett et al., 2015). These results compliment those of (Mohr et al., 2016) who found this same learning dataset showed task automatization coincided with increase network integration across subjects.

Using this combined PCA/LDA approach, we were able to identify a task-specific manifold whose twisted geometry demonstrated a non-linear shift in brain configurations across the task. Our results identified a subject trajectory in these new bases that began with a large velocity early in the task, but then converged toward a more stable configuration as the subjects began to optimize their performance. Importantly, our approach allowed for the comparison of subject trajectories which fall within specific performance groups (fast, slow, and greatest improvers) and to map performance back to 'brain space' for interpretation. In other words, we could observe the reconfiguration of the brain across task performance, while also retaining sufficient variance for the analysis of individual differences in task execution.

Subject level trajectories grouped based on reaction time measures showed that slow reactors had large trajectory deviations from both fast subjects, as well as the best improving subjects which constituted a diminished activation in the left visual ventral stream and bi-hemispheric somatosensory association cortex. These slow performers also demonstrated weaker deactivation in the left dorsal lateral prefrontal cortex and anterior insula (Fig 2c). The somatosensory association cortex is known to play a key role in object recognition (Reed et al., 2004), and

its recruitment by the fastest and greatest improving subjects is consistent with its use here in performance optimization.

The frontoparietal network has been argued to play a central role in adaptive implementation of task demands (Cole et al., 2013) and rule based problem solving (Zhang et al., 2013) – this network involves the dorsal lateral prefrontal and posterior parietal cortices, amongst other regions. Our results demonstrate that these areas are differentially recruited and inhibited across an instruction-based task and highlights their importance in facilitating task performance optimization. Additionally, rule instantiation in an instruction-based visual stimulus-response task has been shown to predominantly involve activation of the visual cortex (Zhang et al., 2013) and our findings recapitulate this result whereby the first eigenmap (Fig 2c i), which differentiated early-to-late brain states in the present task, showed broad visual activation.

Learning has been shown to induce autonomy of sensorimotor areas in a simple motor skills task by liberating cognitive control areas – in particular frontal and cingulate cortices (Bassett et al., 2015). Although the task that we analysed was not explicitly designed to interrogate meta-level learning, we did observe an increased recruitment of frontal cortical regions, along with a disengagement of parts of the cingulate cortex in the left hemisphere during early-to-intermediate phases of the present task. Importantly, subjects with largest improvement in reaction time across repetitions showed diminished activation of this pattern early in the task, and a compensatory activation to levels comparable with the overall fast reactors late in the task.

A key benefit of our approach is that it provides a tractable means to view non-linear trends in multivariate measurements of brain activity by reducing its dimensionality in a targeted fashion. That is, the dimensionality reduction can be tailored to address the investigators specific scientific hypothesis. For this reason, the approach is generalizable with many applications appropriate beyond those presented in this work. In conjunction with traditional methods, dimensionally-reduced trajectory analysis can provide complementary, but unique, insights into trends inherent within multi-dimensional brain data, and could have particular utility in cognitive tasks where moment to moment changes in brain state are of paramount interest. One such application could be to apply PCA and LDA to time varying connectivity (Esfahlani et al., 2020) whereby the trajectory of connectivity changes could be explored in contrast to the trajectory of region-wise BOLD activity presented here.

There are many other dimensionality reduction techniques, such as independent component analysis, Laplacian embedding, t-distributed stochastic neighbour embedding (tSNE), general linear modelling, community detection, clustering, etc., in addition to principal component and linear discriminant analysis used in this work.

This family of methods seek to optimise some cost function across a given dataset and can provide novel insight into trends in data. However, their utility hinges upon the neuroscience question to which they are being applied (Cunningham and Yu, 2014b). In particular, dimensionality reduction techniques have successfully elucidated system level trends across a library of cognitive tasks (Shine et al., 2019a), during memory encoding (Heusser et al., 2021), and can reveal stable dynamics during behaving animals (Gallego et al., 2020; Kato et al., 2015). In the present case, PCA and LDA were chosen as they provide basis for maximally distinguishing group differences in a dataset were the groupings have previously been defined, i.e., task repetition and subjects. Furthermore, the resulting low-dimensional basis allows the process of data projection to be inverted such that manipulations made in the embedding space can be resolved in the original region-of-interest brain space.

An important limitation of this work is that LDA highlights differences between the classes specified and not the features that the classes share. This means that the method is inherently hypothesis driven, in that partitioning of the data is required, and thus is an unavoidable bias constrained by the investigators line of questioning. Other data driven methods, such as PCA, can supplant this deficit by remaining blind to a given hypothesis and instead highlight shared features of the data – i.e., principal axes of covariance.

Conclusion

Here we demonstrate the utility of a dimensionally-reduced trajectory analysis of neuroimaging data captured during a simple instruction-based visuomotor task. This kinematic perspective reveals nonlinear changes in brain state that coincide with a convergence to stable configurations late in the task. Furthermore, the brain state kinematics can be freely transformed back into ROI space, and thus afford unique insights into how canonical control networks are differential recruited during task optimization. The generality of this approach permits utility well beyond the application presented here, and thus can complement standardized methods for analysing brain data during cognitive tasks.

Data and code availability statement

The data that supports the findings of this study are available from the corresponding author on request.

The code used for all analysis of the data in this study is available from the corresponding author on request or can be found at https://github.com/elimuller/BrainStateTraj.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118510.

References

Avena-Koenigsberger, A., Misic, B., Sporns, O., 2018. Communication dynamics in complex brain networks. Nat. Rev. Neurosci. 19, 17–33.

Bassett, D.S., Yang, M., Wymbs, N.F., Grafton, S.T., 2015. Learning-induced autonomy of sensorimotor systems. Nat. Neurosci. 18, 744–751.

Bullmore, E., Sporns, O., 2012. The economy of brain network organization. Nat. Rev. Neurosci. 13, 336–349.

Cohen, J.D., McClure, S.M., Yu, A.J, 2007. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos. Trans. R. Soc. B Biol. Sci. 362, 933–942.

Cole, M.W., Reynolds, J.R., Power, J.D., Repovs, G., Anticevic, A., Braver, T.S., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. Nat. Neurosci. 16, 1348–1355.

Cunningham, J.P., Yu, B.M, 2014a. Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. 17, 1500–1509.

Cunningham, J.P., Yu, B.M, 2014b. Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. 17, 1500–1509.

Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., Ramnani, N., 2009. A probabilistic MR Atlas of the human cerebellum. Neuroimage 46, 39–46.

Esfahlani, F.Z., Jo, Y., Faskowitz, J., Byrge, L., Kennedy, D.P., Sporns, O., Betzel, R.F., 2020. High-amplitude cofluctuations in cortical activity drive functional connectivity. Proc. Natl. Acad. Sci. U. S. A. 117, 28393–28401.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7, 179–188.

Gallego, J.A., Perich, M.G., Chowdhury, R.H., Solla, S.A., Miller, L.E., 2020. Long-term stability of cortical population dynamics underlying consistent behavior. Nat. Neurosci. 23, 260–270.

Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., Petersen, S.E., 2016. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. Cereb. Cortex 26, 288–303.

Heusser, A.C., Fitzpatrick, P.C., Manning, J.R., 2021. Geometric models reveal behavioural and neural signatures of transforming experiences into memories. Nat. Hum. Behav..

Kato, S., Kaplan, H.S., Schrödel, T., Skora, S., Lindsay, T.H., Yemini, E., Lockery, S., Zimmer, M., 2015. Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. Cell 163, 656–669.

Mandelkow, H., De Zwart, J.A., Duyn, J.H., 2016. Linear discriminant analysis achieves high classification accuracy for the BOLD fMRI response to naturalistic movie stimuli. Front. Hum. Neurosci. 10, 1–12.

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., et al., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. U. S. A. 113, 12574–12579.

Mohr, H., Wolfensteller, U., Betzel, R.F., Mišić, B., Sporns, O., Richiardi, J., Ruge, H., 2016. Integration and segregation of large-scale brain networks during short-term task automatization. Nat. Commun. 7.

Mohr, H., Wolfensteller, U., Ruge, H., 2018. NeuroImage Large-scale coupling dynamics of instructed reversal learning. Neuroimage 167, 237–246.

Pursell, L., and Trimble, S.Y. (2018). Gram-Schmidt orthogonalization by Gauss elimination. 9890.

Reed, C.L., Shoham, S., and Halgren, E. (2004). Neural substrates of tactile object recognition : an fMRI study. 246, 236–246.

Ruge, H., Schafer, T., Zwosta, K., Mohr, H., and Wolfensteller, U. (2019). Neural representation of newly instructed rule identities during early implementation trials. 1–25.

Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., Batista, A.P., 2014. Neural constraints on learning. Nature 512, 423–426.

Sheffield, J.M., Mohr, H., Ruge, H., and Barch, D.M. (2021). Disrupted salience and cingulo-opercular network connectivity during impaired rapid instructed task learning in schizophrenia.

Shine, J.M., 2020. The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics. Prog. Neurobiol., 101951.

Shine, J.M., Bissett, P.G., Bell, P.T., Koyejo, O., Balsters, J.H., Gorgolewski, K.J., Moodie, C.A., Poldrack, R.A., 2016. The dynamics of functional brain networks: integrated network states during cognitive task performance. Neuron 92, 544–554.

Shine, J.M., Breakspear, M., Bell, P.T., Ehgoetz Martens, K., Shine, R., Koyejo, O., Sporns, O., Poldrack, R.A, 2019a. Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. Nat. Neurosci. 22, 289–296.

Shine, J.M., Hearne, L.J., Breakspear, M., Poldrack, R.A., Mattingley, J.B., and Cocchi, L. (2019b). The low-dimensional neural architecture of cognitive complexity is related to activity in medial thalamic nuclei. 1–7.

Yarkoni, T., Barch, D.M., Gray, J.R., Conturo, T.E., and Braver, T.S. (2009). BOLD correlates of trial-by-trial reaction time variability in gray and white matter : a multi-study fMRI analysis. 4.

Zhang, J., Kriegeskorte, N., Carlin, J.D., and Rowe, J.B. (2013). Choosing the rules : distinct and overlapping frontoparietal representations of task rules for perceptual decisions. 33, 11852–11862.