Adaptive engagement of cognitive control in context-dependent decision-making

Michael L. Waskom[1], Michael C. Frank[1], Anthony D. Wagner[1,2]

Affiliations

1. Department of Psychology, Stanford University, Stanford, California, USA
2. Neurosciences Program, Stanford University, Stanford, California, USA

Correspondence should be addressed to mwaskom@stanford.edu

Abstract

Many decisions require a context-dependent mapping from sensory evidence to action. The capacity for flexible information processing of this sort is thought to depend on a cognitive control system in frontoparietal cortex, but the costs and limitations of control entail that its engagement should be minimized. Here, we show that humans reduce demands on control by exploiting statistical structure in their environment. Using a context-dependent perceptual discrimination task and model-based analyses of behavioral and neuroimaging data, we found that predictions about task context facilitated decision-making and that a quantitative measure of context prediction error accounted for graded engagement of the frontoparietal control network. Within this network, multivariate analyses further showed that context prediction error enhanced the representation of task context. These results indicate that decision-making is adaptively tuned by experience to minimize costs while maintaining flexibility.

Introduction

Humans inhabit complex and dynamic environments. When weighing evidence to select actions, we often must selectively attend to features that are relevant to our current goals while ignoring others that might be important in a different context. An understanding of human intelligence must account for how the brain operates in these circumstances. Simple and routine decisions can be made relatively automatically as stimuli become paired with responses through associative learning (Shiffrin and Schneider 1977; Law and Gold 2009). One approach to context-dependent behavior would likewise draw on fixed associations between all possible contextual cues, stimuli, and responses. This approach scales poorly, however, and it cannot account for the substantial flexibility of human decision-making (Botvinick et al. 2009; Duncan 2013). Flexible behavior is therefore attributed, in part, to top-down processes that dynamically align sensorimotor transformations with the rules and goals that frame a decision (Cohen et al. 1990; Miller and Cohen 2001; Botvinick and Cohen 2014). These processes are collectively termed cognitive control.

Cognitive control is strongly associated with a network of regions in lateral prefrontal and parietal association cortex (Goldman-Rakic 1987; Miller and Cohen 2001; Koechlin et al. 2003; Wagner et al. 2004; Badre 2008; Cole et al. 2013). Neurons in this frontoparietal network have diverse response properties that reflect both high-level stimulus information and abstract task states (Buschman et al. 2012; Stokes et al. 2013). When control is engaged, the pattern of activity across these neurons is thought to encode a distributed representation of the current task context (Mante et al. 2013; Rigotti et al. 2013). This signal temporarily strengthens contextually-relevant stimulus representations and governs how sensory evidence is selected for decision-making (Cohen et al. 1990; Egner and Hirsch 2005; Waskom et al. 2014). The control system is highly flexible and can in principle enable the broad diversity of human cognition (Duncan 2013). Controlled processing is also subjectively effortful, however, and it is treated as costly: humans strategically avoid situations that require control, even when doing so forgoes potential rewards (Kahneman 1973; Kool et al. 2010; McGuire and Botvinick 2010). Moreover, control is subject to capacity limitations and easily disrupted by concurrent task demands (Shiffrin and Schneider 1977; Duncan 1980; Marois and Ivanoff 2005). Normatively, these constraints entail that the engagement of control should be minimized (Shenhav et al. 2013), but it is unclear how this is achieved.

We propose that context-dependent behavior can be facilitated through a predictive model that learns over abstract representations of task context. This proposal extends theories of how statistical regularities shape sensorimotor processing (Ouden et al. 2012; Summerfield and Lange 2014) to emphasize the consequences of learning over higher-order representations. Our account provides a straightforward explanation for the recruitment of cognitive control, which should be engaged when expectations about a decision's context are violated. Violation of expectations can be quantified in terms of prediction error; a learning algorithm that minimizes prediction error will therefore also minimize the use of control. Specifying a model of this algorithm is not straightforward, however, as predictions about task context are not overtly expressed and relate only indirectly to observable behavior. The approach we take here employs a Bayesian ideal observer model to furnish an estimate of environmental information that could be exploited by such a learner. Using this model, we derived a trial-by-trial estimate of context prediction error, which we then related to behavioral and neuroimaging measurements.

We found that decision-making was sensitive to environmental statistics and that context prediction error accounted for graded engagement of the frontoparietal control network during decision-making. Within the control network, prediction error further modulated the distributed representation of task context, which was strongest on trials with the largest prediction error. Together, these findings provide support for a formal account of how adaptive learning can regulate the engagement of control processes and thus contribute to intelligent behavior.

Materials and methods

Subjects

Twenty healthy right-handed members of the Stanford community participated in the experiment after giving informed written consent in accord with the Stanford University Institutional Review Board. Of this group, one was excluded early in the training session for poor color vision, one was excluded after the training session for performing at chance, and three ended the scanning session early due to fatigue or illness. The analyses reported here comprise data from the 15 subjects who completed the scanning session (19–29 years old; 8 females). All subjects had normal or corrected-to-normal visual acuity and normal color vision. Subjects received monetary compensation for their time ($10/hour for behavioral sessions and $20/hour for scanning). For further motivation, we told subjects before the scanning session that they could receive an additional monetary bonus that would be based on their performance. This bonus ($10) was awarded to all subjects who completed the experiment.

Stimuli and experiment design

Subjects viewed a bivalent random dot stimulus that contained both color and motion information and were cued to make either a motion or a color discrimination on each trial. The experimental stimulus consisted of a $7^\circ \times 7^\circ$ field of 0.07° square dots centered on a 0.2° circular fixation point. Subjects were instructed to maintain fixation on this point, and fixation was monitored using an eye-tracking camera. The background was drawn in gray at

33% luminance, and the dots were colored either red (RGB 0.93226, 0.53991, 0.26735) or green (RGB 0, 0.74055, 0.22775). These colors were chosen in $\mathrm{H_{CL}}$ space to provide maximal and equal chroma at 67% luminance. The stimulus array was divided into three groups of 50 dots, and these groups were each drawn on separate screen refreshes. On every refresh, a proportion of the dots (corresponding to the motion coherence value) was displaced either upwards or downwards, and the rest were redrawn in a random position. If displacing a dot moved it out of the field of view, it was redrawn on the opposite side as if it had wrapped around. This procedure resulted in 5°/s motion at the specified coherence. For each subset of dots, a proportion (corresponding to the color coherence) were drawn at the target color, and the color for the remaining dots was chosen randomly.

Surrounding the dot field was a 0.5° wide frame that cued the context for each trial. The frame was constructed using sine-wave gratings to produce four different patterns, with two distinct patterns each indicating a motion trial and two distinct patterns each indicating a color trial. The assignment of frame pattern to context was randomized across subjects, and all subjects reported that they could discriminate the patterns in their visual periphery.

During the scanning session, this stimulus was presented on a 1920 × 1080 pixel LCD display with a 60 Hz refresh rate, which was placed at the back of the scanner bore and viewed through a mirror attached to the head coil. Subjects indicated their decision using a magnet-compatibile button-box held in their right hand. The two possible responses in each context (i.e., up vs. down and red vs. green) were mapped to buttons under the index and middle fingers. Stimulus presentation and response collection were controlled using PsychoPy.

The main experiment performed in the scanner consisted of 900 individual trials separated into 12 scanning runs and presented with a rapid event-related design. Trials were evenly divided into one of three types. During *early-cue* trials, the stimulus frame appeared 1 s before stimulus onset to indicate the context for the upcoming trial. During *concurrent-cue* trials, the frame and dots appeared at the same time. Dot presentation lasted for 2 s in both cases, and behavioral responses were collected until the offset of the dots and frame. During *cue-only* trials, the frame appeared on the screen for 1 s but then offset without a presentation of dots. In all cases, the fixation point turned from black to white 0.5 s before the frame onset to alert the subject to the impending trial onset, and it returned to black at the offset of the frame to signal the end of the trial. The inter-trial-interval (ITI) was sampled from a geometric distribution between 2 and 10 seconds with $p = \frac{1}{3}$, providing a mean ITI of 4 s.

The experiment was divided into 10 epochs across which we parametrically manipulated the relative frequency of motion and color trials. These epochs did not correspond to breaks between scanner runs, and the transitions between them were not otherwise indicated to subjects. Each epoch was constructed with either $\frac{1}{5}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, or $\frac{4}{5}$ color trials and a complementary proportion of motion trials. These conditions were divided across one long (120 trials) and one short (60 trials) epochs and ordered such that the two epochs for each condition occurred in different halves of the experiment. Within each epoch, trials were randomized subject to several constraints. Each cue pattern appeared equally often within the context that it cued. The two feature values on each dimension also appeared equally often within an epoch, and congruent trials (i.e. trials where the motion and color evidence specified the same response) were held at a constant rate (47%). Context switches were controlled so that the proportion of switch trials for a given context was 1 - the context frequency within that epoch. The first-order transition matrix for the three trial types (i.e. early-cue, concurrent-cue, or cue-only) was also balanced. The same experimental design was used for all subjects.

Subjects were trained on the task during a separate session 1-3 days before they were scanned. Training proceeded in three stages. Subjects first performed the task with the stimuli at full coherence in a block design (8 trials per block, all with the same context and cue) and received feedback on their responses. This stage terminated when they had performed one block for each cue with perfect accuracy. Subjects then performed a staircase task that calibrated subject-specific coherence levels independently for the motion and color evidence. This stage was also blocked (6 trials per block) and contained feedback. The calibration proceeded for 60 blocks using a 1-up-4-down staircase procedure. We set the coherence for the scanning session by averaging the coherence for the last 10 blocks of the staircase stage. Finally, subjects performed 200 trials of the task in an unblocked fashion. This practice was close to the scanner version of the task, but it lacked cue-only trials and the context was chosen randomly from a balanced distribution on each trial. On the day of the scanning session, subjects performed an additional block of these practice trials outside the scanner (minmum 150 trials, with feedback on the first 50 trials) and a final round of practice during the structural scans (100 trials). In total, subjects spent 90 minutes undergoing behavioral training outside the scanner, and the scanning session itself lasted 120 minutes.

Bayesian ideal observer model

We implemented a Bayesian ideal observer to measure the evolving state of the task environment (Fig. 2 and Supplementary Fig. 1). This model was adapted from one previously described by Behrens and colleagues in the context of reward learning (Behrens et al. 2007). We used it to predict motion and color trial probabilities over the course of the context-dependent decision-making experiment. The ideal observer approach is advantageous for this purpose because predictions about task context do not directly guide subjects' choices or other overt reports of the their expectations. Other related approaches, such as reinforcement learning models, would involve free parameters that must be fit to behavior; the Bayesian approach, in contrast, allows us to infer those parameters directly from the experimental design. It is important to emphasize, though, that in using a Bayesian model, we neither assume nor conclude that human learning is optimal or based only on the information provided to the model (Griffiths et al. 2012; Frank 2013). Instead, the model formalizes the hypothesis that the statistics of motion and color trials were relevant for decision-making, which allows us to explore how those statistics influenced the parameterization of cognitive processes that subjects used to perform the task.

Information about what is likely to happen on a given trial can be gained by considering the contexts that have been observed on previous trials. More formally, on trial i the subject has observed a sequence of contexts $c_1, c_2, ..., c_{i-1}$, which we define as 1 for color decisions and 0 for motion decisions. In this model, the context for each decision is chosen as the result of a Bernoulli trial controlled by parameter b_i, a Beta-distributed variable with continuous support between 0 and 1:

$$p(c_i = 1|b_i) = b_i.$$

The b_i parameter reflects the environmental bias towards color decisions and is what allows for the prediction of likely task demands. In a stationary environment, the bias can simply be approximated by the proportion of trials that have required a color decision. Our model assumes, however, that the bias can change over time. It thus infers the bias on each trial such that the estimate of b_i is informed by the estimate on the previous trial, b_{i-1}, and a hyperparameter s. The s parameter serves as an estimate of environmental stability and controls how much the estimate b_{i-1} is updated by the data c_i on the current trial. This term serves an analogous function to the "learning rate" in reinforcement learning models: when the environment is stable, there is less information in the outcomes of individual trials, whereas the estimate of b_i should heavily weight recent data if the environment is expected to change often.

To relate these variables, we can restate the Beta distribution in terms of its mean and variance,

$$p(b_i|b_{i-1}, S) = \mathrm{Beta}(b_i, S),$$

where $S = \exp s$ so that integrals are evaluated in log space. With these definitions, the posterior probability for the model parameters can be written as

$$p(b_i, s|c_{\leq i}) \propto p(c_i|b_i) \int p(b_i|b_{i-1}, s) p(b_{i-1}) p(s) \,\mathrm{d}b_{i-1}.$$

We fit this model by using grid sampling to approximate the integrals. The model was initialized with a uniform distribution over b_1 and s and then fit to the event schedules for the final round of practice trials and the main experiment (exclusion of the practice trials produced highly similar model predictions). This yielded the joint distributions on b_i and s for all i. We then marginalized over s to obtain a univariate distribution on b_i for each trial and computed the posterior mean, $\hat{b}_i$, by taking a weighted sum over the 99 points in the grid,

$$\hat{b}_i = \sum_{j=1}^{99} b_{ij} p(b_{ij}).$$

This value represents the ideal observer's estimate of the probability that trial i will require a color decision.

We next used a combination of $\hat{b}_i$ and the observed context, c_i to compute a context prediction error:

$$CPE_i = \begin{cases} 1 - \hat{b}_i & \text{if } c_i = 1 \\ \hat{b}_i & \text{if } c_i = 0. \end{cases} .$$

The resulting timeseries of CPE values formed the main computational explanatory variable in our behavioral and imaging analyses. Note that CPE is derived from the model, but it is not explicitly computed in the process of model-fitting, as is the case with the prediction error term in reinforcement learning models. Our goal in applying CPE to the analysis of neural data was not to identify regions that perform these computations, but rather to determine whether learning about the structure of the environment influenced control-related processing.

It is important to note that our model differs from the original implementation (Behrens et al. 2007) in how the environmental stability is estimated. In the original model, stability was assumed to change over time, so its estimate was based only on recent trials. In contrast, our model assumes a single rate of change in b_i, so while the estimate of s is updated on each trial, its value reflects all available data. Although these two models make different predictions about the trial-by-trial estimate of s, within our experimental design they produce highly similar estimates for $\hat{b}_i$ (r = 0.99), and so we use the model presented here for reasons of parsimony. Nevertheless, the full hierarchical model would allow the basic ideas explored here to be straightforwardly extended to environments with fluctuating stability, emphasizing a key virtue of the Bayesian approach.

In contrast to the abrupt changes in context frequency in our experimental design, the estimate of $\hat{b}_i$ in our model undergoes incremental adjustments and thus is arguably not optimal. An alternate Bayesian model that does explicitly model change-points produces highly similar values for $\hat{b}_i$ (r = 0.95), however (Wilson et al. 2010). Moreover, it is not clear whether our subjects construed the learning problem as one of identifying discrete change points, and that construal may not be generally optimal for predicting task context. In exploratory analyses, we determined that the small differences in predictions about $\hat{b}_i$ from these alternate models did not account for additional variance in the neural data. It is important to emphasize that, while we focus on the most parsimonious model in our main analysis, our goal is not to provide unique support for one particular formulation (see also Wilson and Niv 2015).

Data analysis

Data were analyzed with a mixture of published software and custom code written in Python. All custom code will be made available at https://github.com/WagnerLabPapers. Imaging data were processed with a workflow of FSL (Jenkinson et al. 2012), FreeSurfer (Fischl 2012), and ANTs (Avants et al. 2008) tools implemented in Nipype (Gorgolewski et al. 2011). The Python code used a number of libraries including numpy, scipy, matplotlib, pandas, and IPython. Cortical surface visualizations were created using PySurfer. The R package lme4 was used for mixed-effects modeling of the behavioral data and summary statistics derived from the fMRI data. These models were fit with maximal random-effects structures, and *p* values for the fixed-effects parameters were obtained using likelihood ratio tests (Barr et al. 2013). Where error bars are plotted for within-subject measures, we first removed between-subjects intercept variance and then performed a multilevel bootstrap on the effect of interest by resampling at both the subject and trial level. As our experimental design truncated long reaction times, we report behavioral regression models fit to untransformed RT data for ease of interpretation; log transforming the RT values did not change any of our main conclusions.

fMRI acquisition

Brain imaging was performed on a 3T GE Discovery MR750 system (GE Medical Systems) using a 32-channel transmit-receive head coil (Nova Medical). Functional images were obtained using a T2*-weighted two-dimensional echo planar imaging sequence with a relatively high spatial resolution (TR = 2 s, TE = 30 ms, flip angle = 77°, 33 slices, 96 x 96 matrix, 2 x 2 x 2.3 mm voxels, axial oblique interleaved acquisition). Slices were acquired aligned with and dorsal to the superior temporal sulcus to provide coverage of frontal and parietal cortex. Additionally, a whole-brain high-resolution T_1-weighted SPGR volume was acquired for cortical surface modeling, across-run alignment, and normalization to a common group space.

fMRI preprocessing

Each timeseries was first realigned to its middle volume using normalized correlation optimization and cubic spline interpolation. To correct for differences in slice acquisition times, data were temporally resampled to the TR midpoint using sinc interpolation. At this point, functional data used in the volume-based group analysis were smoothed with a 4 mm FWHM gaussian kernel using an algorithm that limits smoothing within voxels of similar intensity. This step was omitted for data used in surface-based, ROI, and multivariate analyses. Finally, the timeseries data were high-pass filtered by fitting and removing gaussian-weighted running lines with an effective cycle cutoff of 128 seconds. Images with artifacts were then automatically identified using the following rules: (1) frames on which total displacement relative to the previous frame exceeded 0.5 mm; (2) frames where the median intensity across the whole brain deviated from the run median by greater than 4.5 median absolute deviations (MADs); (3) frames where any median slice intensity, after subtracting the whole-brain median, deviated from the run median by greater than 10 MADs.

Separately, the T1-weighted anatomical volume was processed using Freesurfer to segment the grey-white matter boundary and construct tessellated meshes representing the cortical surface (Dale et al. 1999). A linear transformation (6 degrees of freedom) between the native functional space and the native anatomical space was estimated for each run using boundary-based registration (Greve and Fischl 2009). The anatomical volume was further normalized to FSL's nonlinear MNI152 template using symmetric diffeomorphic registration (Avants et al. 2008). The linear registration from functional to anatomical space and the nonlinear registration from anatomical space to the common template were then combined for volume-based normalization after first-level model fitting.

fMRI modeling

We fit three different models of the task for the voxelwise analyses. The primary model had the following regressors: (1) task events for all trials, (2) task events for error trials, (3) a parametric predictor for all stimulus trials with amplitude modulated by RT, (4) an interaction between the RT regressor and the error trial regressor, (5) a parametric regressor for all trials with amplitude modulated by CPE, (6) an interaction between the CPE regressor and error trials. These regressors were constructed as boxcar functions onsetting with the cue and lasting for the early cue duration and mean response time, where applicable. The height of the parametric regressors was determined by centering and z-scoring the parametric variable across all runs. These regressors were then convolved with a canonical difference-of-gammas model of the hemodynamic response function. Additional covariates included the temporal derivatives of these regressors, realignment parameters to control for residual motion confounds, and indicator vectors to control for artifactual effects identified during preprocessing.

A second model augmented the primary model with a regressor for all task events corresponding to trials where the task context was different from the context on the preceding trial. Design matrix creation was otherwise identical to the primary model.

A third model was similar to the primary model but split the task events into cue period (for early-cue and cue-only trials) and stimulus period (for early-cue and concurrent-cue trials). In this model, the single parametric RT regressor modulated only the stimulus events, but there were separate parametric CPE regressors for cue and stimulus periods. Interactions with the error regressor were also omitted. Design matrix creation was otherwise identical to the primary model.

First-level timeseries models were estimated in native run space using gaussian least squares with local autocorrelation correction (Woolrich et al. 2001). The resulting images of parameter estimates and standard error were then registered into a common space. This was accomplished either by combining the affine functional-to-anatomical and nonlinear anatomical-to-template transformations (for the volume-based analysis) or by combining each run's functional-to-anatomical registration with the inverted registration for the first run (for the surface-based and ROI analyses). These images were then entered into a second-level fixed effects model to combine statistical estimates across runs.

The primary mass-univariate group-level results were obtained after volume-based normalization. Statistical modeling was performed with Bayesian estimation of a linear mixed effects model to produce z statistic maps for a one-sample group mean test of the fixed-effects parameters (Woolrich et al. 2004). To correct for multiple comparisons, these maps were thresholded at $z = 2.3$ and cluster corrected using Gaussian random fields theory to control the family-wise error rate at 0.05. For visualization and better comparison to the cortical atlas of functional networks, we also fit group models in Freesurfer's average surface space. This was accomplished by sampling the statistical data from the unsmoothed, native space fixed effects models at the midpoint between the white and pial cortical surfaces, normalizing with a curvature-driven spherical transformation (Fischl et al. 1999), and smoothing along the surface manifold by averaging the values at neighboring vertices to an effective kernel size of 6 mm FWHM. Unlike the volume model, the surface model was estimated using Ordinary Least Squares, thresholded at $p < 0.01$, and cluster corrected using a Monte Carlo simulation of cluster sizes under the null hypothesis.

ROI analysis

We further characterized univariate activation in cortical regions of interest (ROIs). To avoid circularity and facilitate comparisons across experiments, we defined the ROIs using a population atlas of task-independent cortical networks (Yeo et al. 2011; Waskom et al. 2014). These regions are defined on the Freesurfer average cortical surface mesh. To obtain ROI masks in native functional space, we first reverse-normalized the ROI labels using the spherical registration parameters (Fischl et al. 1999) and then transformed the vertex coordinates into the space of the first run using the inverse of the funcational-to-anatomical registration. We then labeled all voxels intersecting the midpoint between the gray-white and gray-pial boundaries. This produced ROIs that respected the underlying two-dimensional topology of the cortical surface and minimized the contributions of voxels lying outside of gray matter.

We used the resulting masks to characterize regional activation with two different approaches. To better understand the patterns of results in the whole-brain univariate models, we extracted the mean fixed effects parameter estimates from models fit to unsmoothed data in native space. We then performed random effects analyses on the resulting summary values. To visualize the evoked responses in these regions, we also estimated finite impulse response (FIR) models on mean timeseries data within the regions. This allowed us to characterize the hemodynamic response to different task events without assuming a parameterized shape. These models were fit to residual timeseries after removing confounds (motion regressors and artifact indicator vectors, as in the main univariate model). Because our experimental design was sampled at 1 s resolution, we upsampled the timeseries data using cubic spline interpolation before fitting these models.

Dimensionality reduction

To characterize the expression of information about task context in frontoparietal cortex, we adopted a dimensionality reduction approach (Mante et al. 2013; Cunningham and Yu 2014). This approach can be motivated by thinking about the pattern of activation across voxels as a point in a high-dimensional space, where each dimension corresponds to the activation of an individual voxel. Our analysis attempts to find a single linear axis within this space along which the patterns of activation for motion and color trials are separated. Having identified this axis, we can then project the patterns measured in different experimental conditions onto it, which pools across voxels to produce a scalar estimate of the context representation's strength on those trials. By comparing the magnitude of the projection in different conditions, we can ask how different experimental variables influence the representation of task context.

Practically, this analysis involved fitting a series of univariate models to the timeseries data from each voxel in our ROIs and then operating on vectors of the resulting regression coefficients. As in the FIR analysis, we used the residual timeseries after removing motion confound effects. We then fit a model with main effects of task, errors, RT, and an interaction between RT and errors. This model was structured identically to the main univariate model used in the whole-brain analysis, except it did not include a parametric effect of CPE. All subsequent operations were performed on the z-scored residuals from this model, which subtracted out the average task-evoked response and controlled for confounding effects of RT and error rate. A subsequent control analysis additionally included a regressor for context switches in this model, which allowed us to determine whether the effects persisted after removing variance associated with switch trials.

We estimated each voxel's context tuning by fitting a single regressor that contained values of 1 for each motion trial and -1 for each color trial before convolution with a canonical HRF model. We then used the vector of tuning coefficients, after scaling the vector to a unit norm, as an estimate of the linear axis corresponding to the context representation in the population response space. We further used a cell means model to estimate the average response in 8 independent conditions defined by the interaction between task context (motion or color) and quartiles of CPE and then projected these response vectors onto the context axis. To control for differences in mean activation across the CPE variable, we normalized each projection using the norm of the population response vector. We therefore obtained a scalar value for each condition that quantified the average strength of the context representation on those trials. We calculated the difference between these values for the motion and color responses within each CPE bin and used a mixed effects model to test the linear slope of the difference values with respect to CPE. The logic of this analysis is similar to testing an interaction between context tuning and CPE in each voxel. Unlike a univariate analysis, though, the dimensionality reduction makes no assumptions about the spatial organization of voxels and can pool information from voxels that show weak but consistent effects. For maximal power, we estimated both models on the full dataset. Although this overfits the average separation along the context axis, we were interested in how that separation changed as a function of CPE, which is an independent analysis.

We further investigated the spatial organization of the context tuning coefficients to determine whether voxels with similar tuning preferences were clustered together. To measure clustering, we compared the coefficient value in each voxel to the mean value of the coefficients in adjacent voxels. We then squared the differences and took the sum across each ROI, which provides an estimate for the amount of clustering. The logic of this analysis is that, in clustered data, it is possible to predict the value in a given voxel from the values of its neighbors. Therefore, a lower sum squared difference indicates higher amounts of clustering. Although the dimensionality reduction analysis was performed on data that had not been spatially smoothed, some amount of smoothness is nevertheless expected due to influences such as interpolation during realignment, the point spread function of the hemodynamic response, and contributions from large veins. We therefore obtained a null distribution for the clustering measure by shuffling the context labels in the experimental design and refitting the tuning coefficients in 100 separate iterations. We z-scored the vector of coefficients within each ROI before computing the clustering measures for the actual and resampled values so that they would be comparable in magnitude. We then computed the percentile of the observed clustering in the shuffled null distribution to obtain a p value for each subject and ROI.

Results

Fifteen human subjects performed a context-dependent perceptual discrimination task (Fig. 1A). On each trial, they were cued to report either the direction of coherent motion or the dominant color of a bivalent random dot stimulus. We will use the term “context” to refer to the relevant stimulus dimension (motion or color). The strength of the motion and color evidence was calibrated for each subject before scanning and fixed during the main experiment (motion coherence: 0.25 ± 0.10; color coherence: 0.18 ± 0.06). We attempted to find coherence values that were attentionally demanding without leading to an excess of error trials. This procedure balanced accuracy in each context (motion: 0.87 ± 0.08; color: 0.89 ± 0.05; $p = 0.44$), although reaction times (RTs) during scanning were faster for motion decisions ($\beta = 0.072$ s; $\chi^2_1 = 7.1$, $p = 0.008$).

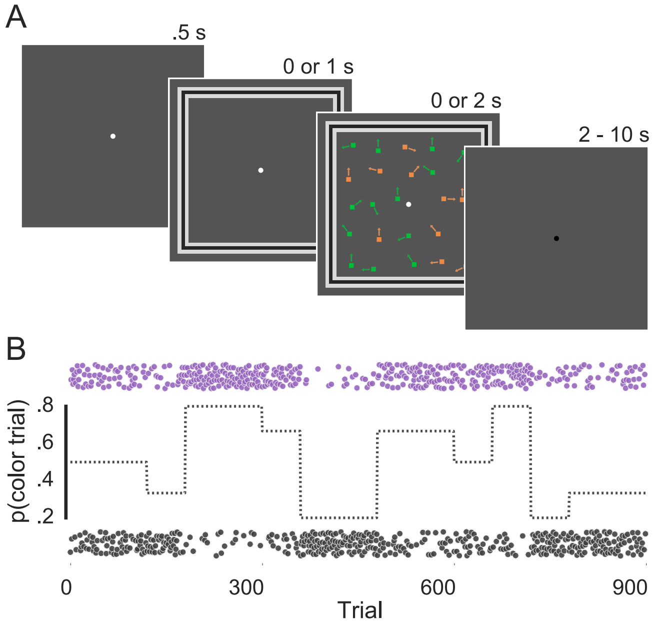

Figure 1: Experimental design for the context-dependent perceptual discrimination task. (A) Subjects were cued to make a discrimination along either the motion or color dimension of a bivalent random dot stimulus. The context for each decision was indicated by the pattern of the frame surrounding the dot field. On one third of trials, frame onset preceded the dot stimulus, and on a separate third of trials, the frame was presented without a dot stimulus. These events were concurrent on the remainder of trials. Trials were delivered in a fast event-related design with temporal jitter between task events. (B) The experiment was structured into epochs with different relative frequencies of motion and color trials. The dashed line shows the generating frequency for each epoch, and the points show the actual distribution of motion and color trials over the course of the experiment.

The experiment was structured into epochs over which we parametrically manipulated the relative frequency of motion and color trials (Fig. 1B). These epochs lasted for either 60 or 120 trials each, and the transitions between them were not overtly indicated to the subjects. It was thus usually the case that the environment favored one of the two contexts, which the subjects could in principle exploit to reduce demands on cognitive control. Because the environmental structure was not directly apparent, however, it had to be inferred from the observed sequence of trials. We used a Bayesian ideal observer (Behrens et al. 2007) to determine what information would be available from this sequence. Our model describes the context for trial i as a Bernoulli random variable with bias b_i, a Beta random variable that we defined as the probability of a color decision. The estimate of b_i was updated from trial to trial based on the observed context and a hyperparameter, s (Fig. 2A). This hyperparameter, which provides an estimate of environmental stability and controls the learning rate, was also inferred from the observed sequence of contexts.

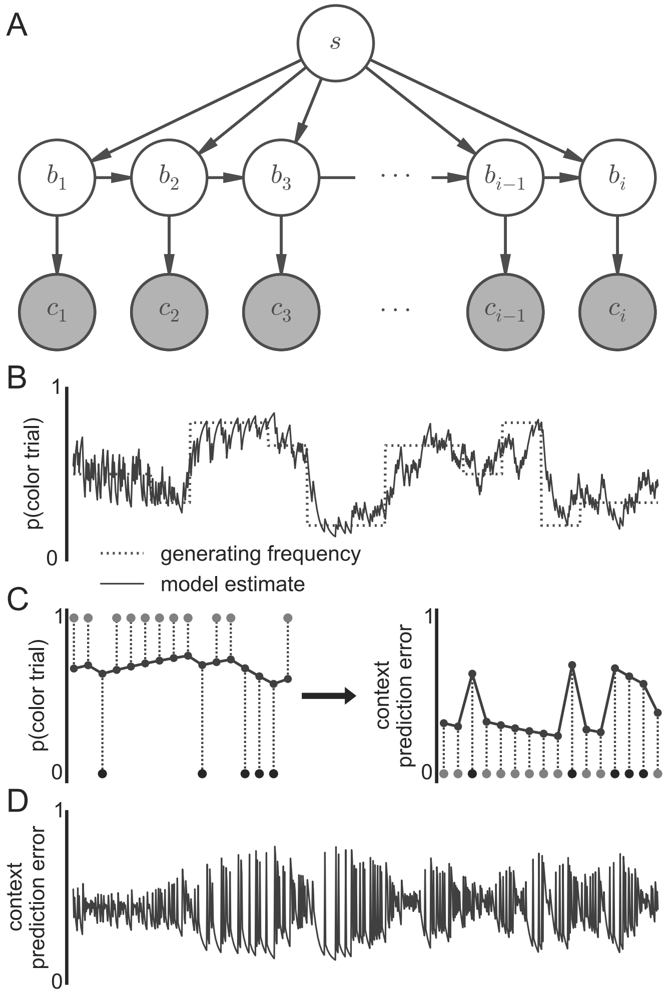

Figure 2: An ideal observer model to predict decision context and derivation of context prediction error. (A) Graphical representation of the model parameters. The estimate of environmental bias (b_i) was updated using Bayes rule from trial to trial as a function of the observed context (c_i) and an estimate of the environmental stability (s). (B) Collapsing the resulting marginal probability distributions to the posterior mean on each b_i node produced a timeseries of $\hat{b_i}$ values, which provide an optimal estimate of the probability that each trial will require a color decision. (C) Using the observed context for each trial, we transformed $\hat{b_i}$ to a measurement of prediction error expressing how unlikely or surprising the context for trial i was, given the history of trials up to that point. (D) This procedure produced a timeseries with a continuous measure of the context prediction error on each trial, which we used as an explanatory variable for behavioral and functional imaging analyses.

We used Bayes' rule to invert this model, which produced a posterior distribution on each model parameter (Supplementary Fig. 1). This procedure used only the experimental design and did not require us to fit free parameters using the subjects' responses. We then collapsed the posterior distribution on each b_i parameter to the posterior mean, $\hat{b}_i$, which provides an optimal estimate of the probability that trial i would be a color trial given the decisions that were required on trials up to that point (Fig. 2B). Finally, we used the $\hat{b}_i$ timeseries and the observed context on each trial to define a measure of context prediction error (CPE) for each trial (Fig. 2C-D). CPE is defined as $1 - \hat{b}_i$ on color trials and $\hat{b}_i$ on motion trials, which expresses how surprising the context for each decision would be to a subject who had learned the probabilities reflected by $\hat{b}_i$. Larger values of CPE indicate more surprise; note that 0.5 is an important value corresponding to the CPE on each trial in the absence of a predictive model. Our central hypothesis was that, if decision processes adapt to the environmental statistics measured by our model, CPE would account for the engagement of cognitive control on each trial.

Expectations about decision context influence reaction time

Learned expectations about the contextual relevance of the two stimulus dimensions facilitated decision-making. We determined this by relating RTs for correct decisions to CPE and other experimental variables with mixed effects regression. Visualizing the data suggested a nonlinear relationship, such that the effect was stronger at relatively low levels of CPE. We linearized this relationship by fitting all behavioral models with log(CPE) as an explanatory variable. Decisions with larger values of CPE, or more surprise about the context on that trial, took longer to successfully complete ($\chi^2_1 = 21.9$, $p < 0.001$; Fig. 3A). This indicates that decision-making was sensitive to the environmental statistics. As a more direct evaluation of our model, we recomputed CPE using the generating context frequency and added it as a covariate. The effect of model-derived CPE remained a significant predictor of RT in this regression ($\chi^2_1 = 20.72$; $p < 0.001$), indicating that the Bayesian model could capture the dynamics of learning latent environmental statistics from noisy observations.

We can gain insight into how the perceptual evidence was transformed into a decision by considering congruent and incongruent trials separately. On incongruent trials, the evidence on the two stimulus dimensions afforded different motor responses, which would yield response conflict if both streams of evidence were accumulated simultaneously. We did not observe a strong relationship between congruency and RT, however ($\beta = -0.014$ s, $\chi^2_1 = 2.77$, $p = 0.096$; Fig. 3B). This suggests that evidence was integrated predominantly along the relevant dimension during the stimulus period. Moreover, there was no indication that congruency interacted with CPE ($\chi^2_1 = 0.3$, $p = 0.56$; Fig. 3B). Although decisions were slower when the context was unexpected, this did not appear to reflect interference from a prepotent representation of the evidence along the irrelevant but expected dimension. Instead, it suggests the existence of a mechanism for selecting the relevant dimension early in the decision process.

Our experimental design included a manipulation of the cue onset time, which allowed us to test whether CPE influenced this early dimension selection mechanism. On half of the trials that required a decision, the context cue was presented 1 s before stimulus onset; these two events were concurrent on the other half of trials (Fig. 1A). Decisions informed by an early cue were faster, indicating that cues could be processed independently of the

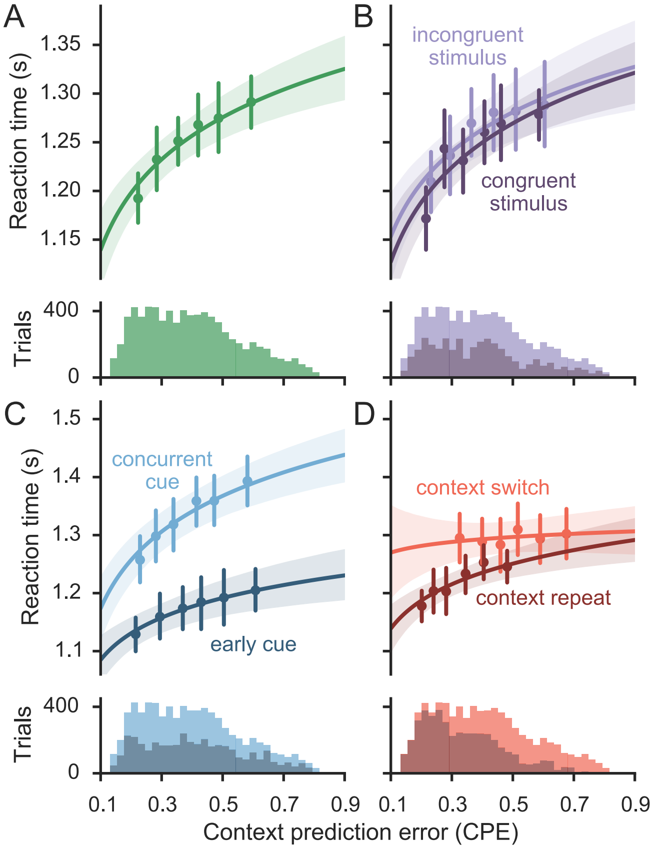

Figure 3: Expectations about task context facilitate decision-making (A) Response time (RT) on correct decisions increased as a function of log(CPE). (B) RTs were similar on trials with congruent and incongruent stimuli, and there was no evidence that the relationship between RT and log(CPE) changed as a function of stimulus congruency. (C) Presenting the context cue early both decreased RTs on average and moderated the relationship between RT and log(CPE). (D) The relationship between RT and log(CPE) remained significant when restricting analysis to trials where the context repeated from the previous trial, but context switches moderated the effect of log(CPE). In all panels, error bars show within-subject 95% confidence intervals computed by multilevel bootstrap resampling after removing betwen-subject variance in median RT.

dot stimulus (β = 0.16 s, χ^2_1 = 22.7, $p < 0.001$; Fig. 3C). The timing of cue presentation also moderated the relationship between CPE and RT (χ^2_1 = 8.8, p = 0.003; Fig. 3C). With an early cue, RTs were less strongly dependent on contextual predictions built up from recent experience, although the effect of CPE was still significant when restricting analyses to these trials (χ^2_1 = 16.9, $p < 0.001$; Fig. 3C).

Context prediction error is naturally related to changes in the decision context. Although we held the probability of a context switch constant throughout the experiment, switch trials had a higher average CPE than those where the previous context repeated. Importantly, switching did not explain away the effect of CPE, which remained significant when the analysis was restricted to context repeat trials (χ^2_1 = 18.2, $p < 0.001$; Fig. 3D). The CPE effect appeared to interact with context switches, however (χ^2_1 = 4.36, p = 0.037), such that it was absent on switch trials (χ^2_1 = 0.71, p = 0.4). This result suggests that context switches require a qualitatively different control mechanism that is independent of CPE. Alternatively, it may indicate that the switch cost, or the additional time needed to implement a switch, itself changes as a function of CPE. Our experimental design cannot distinguish these possibilities.

Because we used two visually distinct cues for each context, we could examine the effect of changing the sensory stimulus that conveyed task demands while holding the more abstract context information constant. Although decisions were slightly faster when the cue repeated (β = 0.016 s, χ^2_1 = 3.85, p = 0.05; Supplementary Fig. 2C), the size of this effect did not account for the relationship between CPE and behavior. There also was no indication that decisions were facilitated when the relevant sensory evidence matched the evidence on the previous trial (β = -0.014 s, χ^2_1 = 1.56, p = 0.2; Supplementary Fig. 2D). Thus, our central findings are unlikely to arise from repetition priming of low-level visual information.

Although we have focused on reaction time, it can also be informative to consider response accuracy and its relationship to context prediction error. Using mixed effects logistic regression, we determined that trials with higher CPE were significantly more likely to result in an error (χ^2_1 = 9.75, p = 0.002), although the size of this effect was relatively small (2% difference in choice accuracy between the highest and lowest quintile of CPE). Moreover we found evidence that the relationship between CPE and response accuracy depended on stimulus congruency (χ^2_1 = 4.12; p = 0.04) and was not present on trials with a congruent stimulus (χ^2_1 = 0.07; p = 0.8). The effect of congruency on response accuracy also depended on the timing of the cue, such that accuracy on incongruent trials was slightly higher when the cue was presented early (χ^2_1 = 4.23; p = 0.04). These results indicate that subjects were more likely to integrate evidence along the wrong dimension of the stimulus when CPE was high, thus producing errors on incongruent trials. The lack of an effect on congruent trials, however, suggests that the evidence accumulation process itself was largely independent of CPE, further emphasizing that control was exerted prior to evidence accumulation in selecting the relevant dimension of the stimulus and that it was at this stage that CPE influenced cognitive processing.

Context prediction error drives frontoparietal control network engagement

We related neural responses to context prediction error and other aspects of the task using high-resolution fMRI. We first employed a mass-univariate approach to estimate the relationship between the model-derived measure of CPE and the activation in each voxel while controlling for the effects of RT and response errors. To understand the magnitude of this effect, we then defined independent regions of interest (ROIs) and estimated the timecourse of activation across the range of CPE sampled by our design. These analyses were performed both for all task events and for the cue and stimulus periods, which were modeled independently. Collectively, this approach allowed us to ask both (a) where in the brain and (b) at what point in the process of forming a decision did CPE best account for BOLD activation.

The task drove activations strongly throughout the brain (Supplementary Fig. 3). Task-evoked increases were strongest in the cortical dorsal attention, frontoparietal, and cingulo-opercular networks and in premotor and motor regions corresponding to the right-hand button-press response. Although we did not manipulate stimulus coherence, BOLD activation in this network of regions has been shown to reflect the strength of evidence in both the classic motion discrimination (Kayser, Buchsbaum, et al. 2010) and context-dependent (Kayser, Erickson, et al. 2010) versions of the random dot task. We also observed strong activations subcortically in the basal ganglia and thalamus. Corresponding decreases in activation during task events were observed in cortical default network regions.

Context prediction error scaled the task-evoked responses within a set of prefrontal and parietal regions (Fig. 4). These effects primarily fell laterally in the inferior frontal sulcus (IFS) and intraparietal sulcus (IPS) and medially in the posterior cingulate cortex (PCC) and superior parietal lobule (mSPL). Notably, the frontal activations extended rostrally into the frontal pole, and the parietal activations extended across the lateral bank of the IPS onto the surface of the inferior parietal lobule. The CPE effect therefore includes regions that did not exhibit a strong task-evoked response. In contrast to these widespread positive effects, no areas exhibited a negative relationship with CPE, even at a relaxed statistical threshold ($p < 0.01$, uncorrected). Importantly, the CPE effect cannot be explained away by context switching, as we observed the same pattern of results when controlling for switch events (Supplementary Fig. 4).

To complement the mass-univariate analysis, we extracted data from four bilateral ROIs defined using a population atlas of task-independent cortical networks (Fig. 5A) (Yeo et al. 2011). We focused primarily on two nodes of the frontoparietal control network (IFS and IPS) and two nodes of a medial cingulo-parietal network (mSPL and PCC).

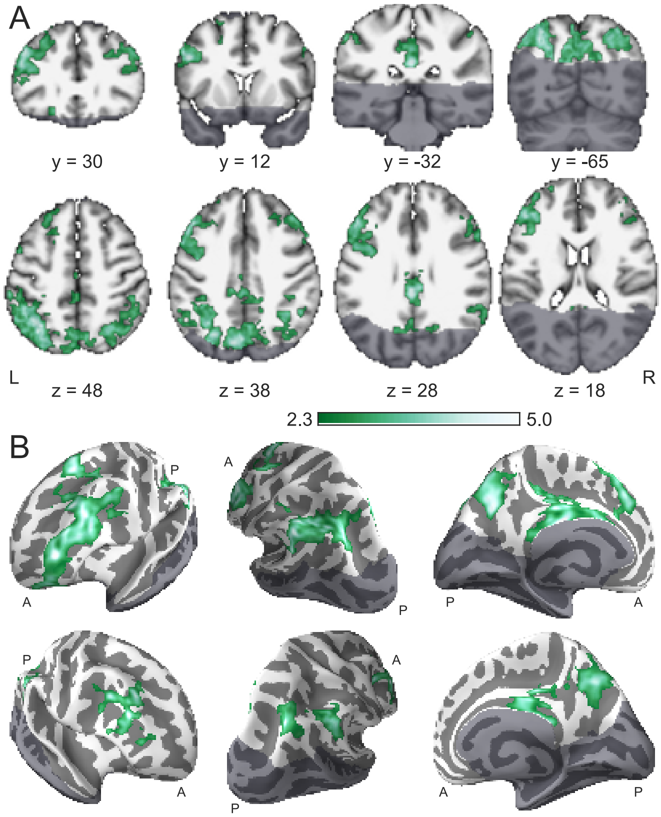

Figure 4: Areas exhibiting a significant parametric effect of context prediction error (A) Results from a volume-based mixed effects group analysis. The position of the slices are indicated in MNI coordinates. The background image is the mean normalized anatomy for subjects included in the analyses. (B) Results from a surface-based random effects group analysis plotted on the Freesurfer average surface anatomy. In both panels, statistical overlays were thresholded at $z > 2.3$ and cluster corrected to control the FWE at 0.05. The colormap shows z scores, and voxels outside of the field of view are dimmed.

Our selection of the IFS and IPS was made *a priori*, as we have previously shown that the IFS and IPS represent task context during controlled decision-making (Waskom et al. 2014). Although we lacked similar expectations about the role of the two medial areas, their definition in the atlas (Yeo et al. 2011) is strongly convergent with the pattern of activations observed in the mass-univariate analysis.

Each of these ROIs exhibited a strong effect of context prediction error when we averaged the model coefficients across their voxels (IFS: $t_{14} = 6.16$, $p < 0.001$; IPS: $t_{14} = 5.93$, $p < 0.001$; mSPL: $t_{14} = 5.47$, $p < 0.001$; PCC: $t_{14} = 5.29$, $p < 0.001$). For further insight, we extracted the average timeseries data from each region and estimated the profile of the evoked response using finite impulse response (FIR) models. To understand the magnitude of the CPE effect, we binned the CPE values by quartiles and modeled the response separately within each of the four bins. Visualizing the resulting estimates showed that CPE accounted for a substantial proportion of the variance in the evoked response (Fig. 5B). Moreover, in contrast to the behavioral results, the neural effect of CPE appeared strongest at relatively high values of the CPE distribution (Fig. 5C).

To test whether context prediction error modulated responses that were driven by the context cue and could be dissociated from stimulus-related processing, we fit a model that was structured as above but with task events split into cue and stimulus periods. We were able to estimate these events separately because we used a partial-trial design, such that some early cues were not followed by the dot stimulus. All four ROIs exhibited an evoked response when the cue was presented in isolation (IFS: $t_{14} = 5.90$, $p < 0.001$; IPS: $t_{14} = 6.25$, $p < 0.001$; mSPL: $t_{14} = 2.28$, $p = 0.039$; PCC: $t_{14} = 4.79$; $p < 0.001$; Fig. 6A), and CPE parametrically scaled the amplitude of these responses (IFS: $t_{14} = 3.01$, $p = 0.009$; IPS: $t_{14} = 3.99$, $p = 0.001$; mSPL: $t_{14} = 2.53$, $p = 0.024$; PCC: $t_{14} = 2.46$; $p = 0.028$; Fig. 6B). This result indicates that control was operating, at least partially, over the selection of relevant stimulus-response transformations in absence of decision evidence.

Our primary model included a parametric effect of RT, which we can use to identify the set of regions that are more generally engaged in decision-making. The effect of RT was widespread, but can be largely characterized in terms of the frontoparietal, cingulo-opercular, and dorsal attention networks (Supplementary Fig. 5). We used the Yeo atlas (Yeo et al. 2011) to extract parameter estimates from major nodes of these three networks and then directly tested dissociations between them (Supplementary Fig. 6). When aggregated across these ROIs, the effect of CPE was significant in both the cingulo-opercular ($t_{14} = 3.20$, $p = 0.006$) and dorsal attention ($t_{14} = 2.54$, $p = 0.024$) networks, although both effects were significantly weaker than the average effect in the frontoparietal network (cingulo-opercular: paired $t_{14} = 5.52$, $p < 0.001$; dorsal attention: paired $t_{14} = 5.51$, $p < 0.001$). Moreover, the effect of response errors was relatively stronger in the cingulo-opercular network than it was in the frontoparietal network ($t_{14} = 4.52$, $p < 0.001$). These results emphasize that, while the context-dependent discrimination task recruits large swaths of cortex, these activations can be dissociated based on intrinsic network

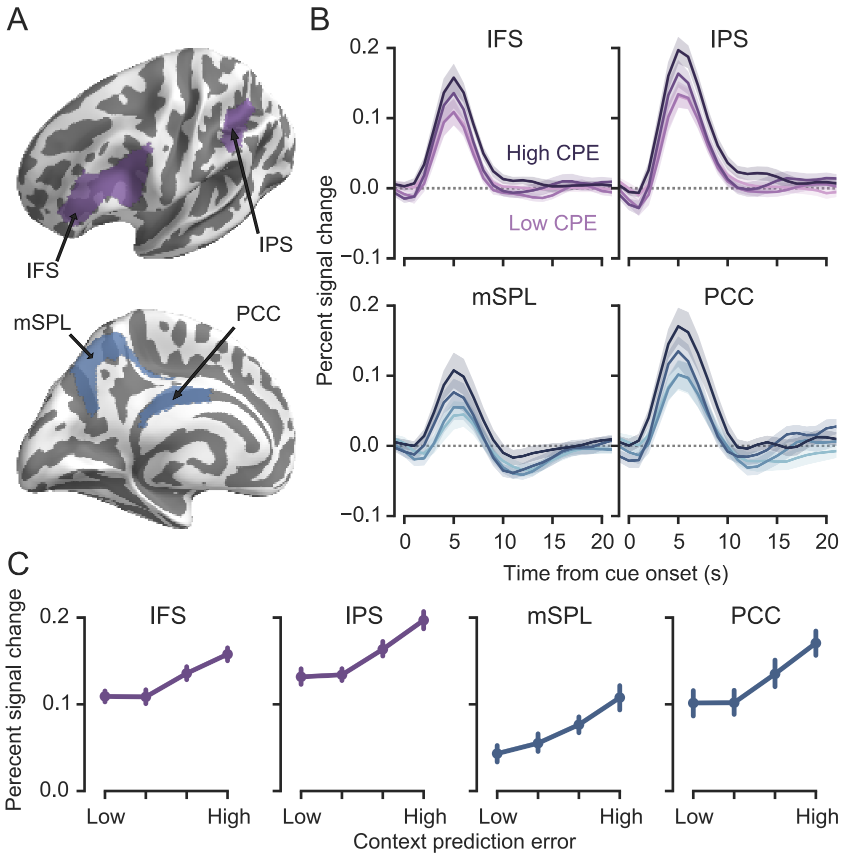

Figure 5: Context prediction error modulates the amplitude of task-evoked responses. (A) Four regions of interest shown on the Freesurfer average surface mesh. These labels were reverse-normalized to the subject-specific surfaces and used to define ROIs in gray matter voxels. The ROIs are colored by their network membership in the functional connectivity atlas. Left-hemisphere ROIs are shown in this figure, but bilateral ROIs were used for all analyses. (B) The trial-evoked response was measured by fitting an FIR model with trials binned using quartiles of the CPE distribution (error trials were fit separately and are not shown). Timeseries were upsampled to 1s (the sampling frequency of the design) using cubic spline interpolation before fitting. Error bands show between-subjects standard error. (C) Amplitude of the response at 5 s following cue or stimulus onset, corresponding to the response peak across CPE quartiles and ROIs. Points show means and error bars show within-subject 95% confidence intervals.

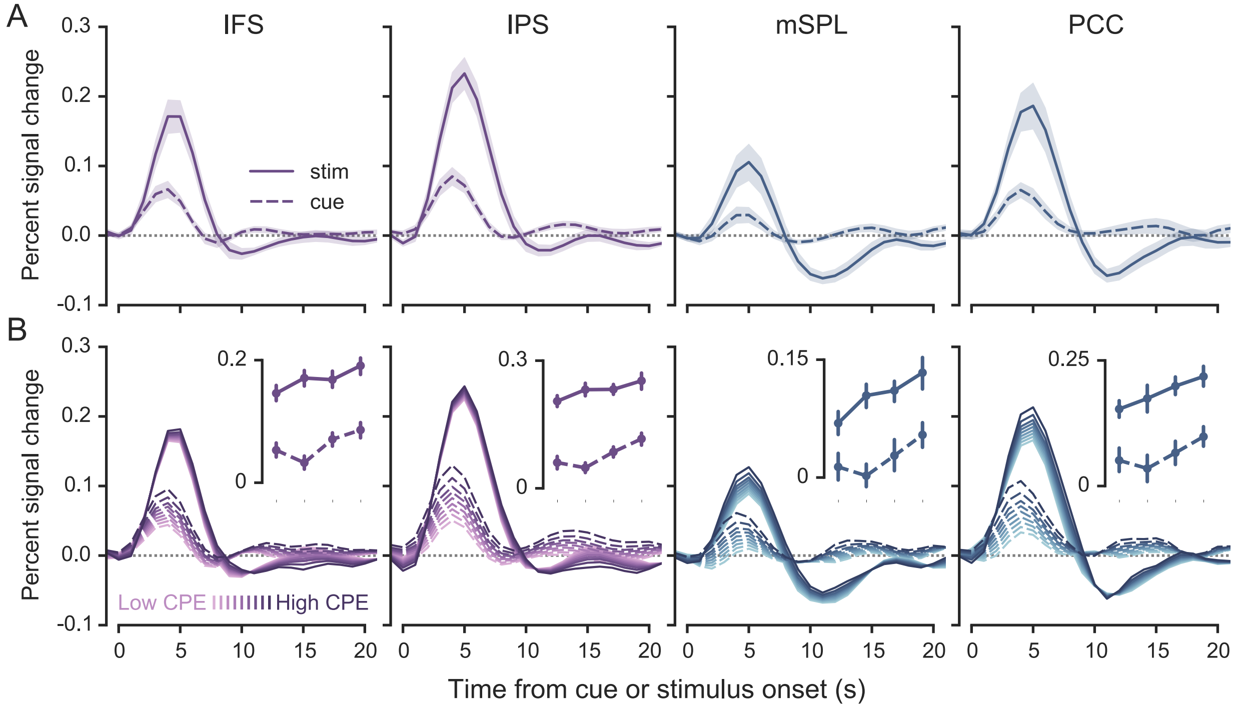

Figure 6: Context prediction error modulates the amplitude of cue-evoked responses. (A) The cue- and stimulus-evoked responses were modeled separately using an FIR model and are plotted as in Fig. 5B. The stimulus period on error trials was also included in this model but is not shown. (B) The temporal profile of the CPE effect was measured using a parametric FIR model, which estimated the size of the CPE effect at each peristmulus timepoint while controlling for RT and errors. The traces show the predicted responses from the fitted model at evenly-spaced quantiles between the 5th and 95th percentiles of the CPE distribution. The spread of the traces shows the amount of variation in evoked response amplitude that can be accounted for by CPE. Insets show the peak (4 s after cue onset and 5 s after stimulus onset) response amplitude in bins defined by quartiles of CPE with the same conventions as Fig. 5C.

membership and the relationship to task variables that place demands on different levels of control.

Context prediction error enhances frontoparietal context representations

It is thought that the distributed pattern of activity in frontoparietal control regions encodes a representation of task context, a signal that determines the modulation of goal-relevant processing in sensorimotor circuits (Miller and Cohen 2001). Neurons in these regions have heterogenous response properties, which endows the regions with high-dimensional population representations. When behavior depends on a distinction between two particular contexts, such as the motion and color rules in our task, this coding scheme allows downstream regions to determine the relevant state with a linear read-out of control network activity (Mante et al. 2013; Stokes et al. 2013). This perspective indicates that the strength of a context representation can be thought of in terms of how distinct its pattern of activity is in relationship to those for competing contexts.

Predictions about task context could, in principle, influence these representations in one of two ways. If predictions establish a strong frontoparietal representation of the expected context before task onset, then prediction errors would be associated with interference and decreased ability to discriminate opposing contexts. In contrast, strong task-level predictions might facilitate automatic selection of the relevant stream of evidence such that context representation would be less necessary for expected decisions. In this case, we would expect the strength of the representations to positively track prediction error. Either of these accounts would be consistent with our univariate results, so it is important to test the informational content of frontoparietal activation patterns directly.

We employed a dimensionality reduction analysis to measure the strength of the context representation and thus evaluate these possibilities (Fig. 7A; Mante et al. 2013; Cunningham and Yu 2014). This approach quantifies the representation of task context by projecting the pattern of activation across voxels within an ROI onto a one-dimensional axis that corresponds to the task context. We defined the context axis by measuring each voxel's context tuning during task events after removing the effects of RT and errors. To evaluate the reliability of the tuning estimates, we fit models separately using data from odd and even scanning runs and asked whether the two resulting coefficient vectors had a nonzero correlation across voxels, i.e., whether the direction vectors were aligned (Fig. 7B). We could reliably identify the context axis in the IFS (mean $r = 0.40$; $t_{14} = 6.71$; $p < 0.001$) and IPS (mean $r = 0.46$; $t_{14} = 8.32$; $p < 0.001$). There was considerable between-subjects variability in the tuning stability metric, and much of this variability was shared between the IFS and IPS ($r = 0.81$; $p < 0.001$). Context tuning also generalized across the two visual cues used for each context (IFS: mean $r = 0.25$; $p < 0.001$; IPS: mean $r = 0.29$; $p < 0.001$), indicating that it was related to the abstract context information and not just the visual patterns used to cue it. Evidence for stable tuning in the medial regions, while significant, was less strong (mSPL: mean $r = 0.23$; $t_{14} = 4.14$; $p = 0.001$; PCC: mean $r = 0.06$; $t_{14} = 2.10$ $p = 0.05$). Following both our *a priori* expectations and this result, we restricted subsequent analyses to the IFS and IPS.

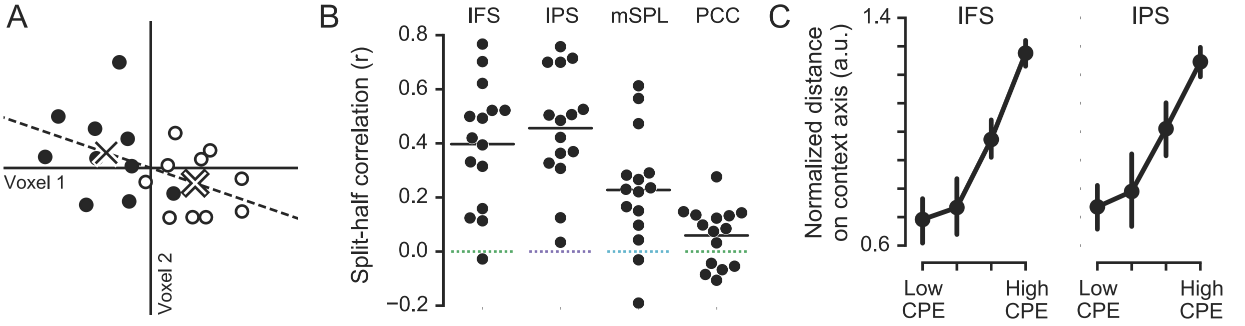

Figure 7: Representation of task context in lateral frontoparietal regions is enhanced by context prediction error. (A) A toy example illustrating the dimensionality reduction approach. This example shows responses in two voxels with different context tunings. The responses shown here have been residualized against average task-evoked responses and confounding variables, so they are centered in the response space. The dashed line indicates the context axis estimated from these responses, and the projection of the average response for each context onto this axis is marked with an X. The separation between the two Xs along the dashed line quantifies the strength of the context representation. (B) The correlation between context tuning coefficients that were estimated separately for even and odd scanner runs. Each point shows the correlation for a single subject, and the mean correlation across all subjects for each ROI is shown with a horizontal line. (C) The distance between average responses for each context separated by CPE quartiles and projected onto the context axis. Error bars show within-subject 95% confidence intervals.

These results imply that information about context is distributed across cortex at a spatial scale that can be resolved with fMRI. The principles that govern such organization are poorly understood. When plotted on the cortical surface, the tuning coefficients appeared to have a relatively fine-scaled, but not random, organization (Supplementary Fig. 7). Using a permutation test, we determined that the maps exhibited a higher amount of spatial clustering than would be expected by chance for 14/15 subjects in the IFS and 13/15 subjects in the IPS ($p < 0.05$). Clusters of voxels that were strongly tuned for each of the two contexts did not appear to be consistently located across subjects, however. These results suggest that our ability to measure context information in frontoparietal cortex depends on a spatial scale of organization that is larger than our voxels but smaller than the components of the large-scale networks that we used to define our ROIs.

To evaluate whether context prediction error influenced the representation of context in the IFS and IPS, we estimated the response on trials grouped by context and quartiles of CPE and then examined the projection of the population response vectors onto the context axis. We found that separation along the context axis increased for the higher CPE bins (IFS: $\chi^2_1 = 34.1$, $p < 0.001$; IPS: $\chi^2_1 = 25.9$, $p < 0.001$; Fig. 7C). To evaluate the robustness of this result, we performed two control analyses. The first analysis controlled for effects of context switches. Similar to what we found in the behavioral and univariate imaging analyses, this analysis showed that the influence of CPE on context separation could not be explained by switch trials (IFS: $\chi^2_1 = 42.77$, $p < 0.001$; IPS: $\chi^2_1 =$

34.56, $p < 0.001$). We also wondered whether our findings could be explained by differences in evoked response amplitude. Our method for computing the separation measure controls for the average response within each CPE bin, but our results might nevertheless be driven by higher signal-to-noise ratio on trials with larger evoked responses. Although we cannot measure the trialwise signal-to-noise ratio directly, we reasoned that the stability measure used above could serve as a proxy for it. We thus computed the split-half correlation between context tuning values estimated within each CPE bin. When including the resulting estimates as a covariate in a multiple regression model, we found that CPE remained a significant predictor of context separability (IFS: $\chi^2_1 = 18.45$, $p < 0.001$; IPS: $\chi^2_1 = 10.0$, $p = 0.002$). We conclude that the representation of task context in frontoparietal cortex is strongest on trials with the highest demands on cognitive control, as reflected in context prediction error.

Discussion

We presented subjects with a task that required context-dependent integration of noisy sensory evidence. Subjects performed the task in a stochastic environment that, at various points, favored one of two possible contexts. The context cues were deterministic, and knowledge of the latent environmental statistics was not strictly necessary to perform the task correctly. Nevertheless, subjects developed predictions about the likely context on each trial, which facilitated decision-making. Although predictions can thus promote efficient processing, in a stochastic environment, they will sometimes be incorrect. Our imaging results indicate that the engagement of frontoparietal control systems can be understood as a graded response to violated predictions about task context. Moreover, we found that prediction error modulated activity within regions that contained a distributed representation of task context and that the strength of the context representation was positively related to context prediction error.

These results extend our understanding of the neural mechanisms that implement simple decision-making. Perceptual decisions arise from a process where noisy sensory information is integrated over time until the weight of evidence in favor of one alternative reaches a threshold (Roitman and Shadlen 2002; Gold and Shadlen 2007; Kiani et al. 2008). Our results indicate that, when the relevant dimension of a visual stimulus depends on context, that dimension must first be selected before its information can bear on choice. In our task, demands on control appeared strongest at this early selection stage. The different functional forms of the relationships between CPE and either RT or BOLD activation, however, imply that our behavioral and neural measurements might reflect different influences on the selection mechanism. We suggest that CPE effects on behavior, which were strongest when the context was relatively expected, primarily reflect facilitation through automatic selection of the relevant dimension. In contrast, CPE effects on neural signals, which were strongest when the context was relatively unexpected, might more directly index how cognitive control is engaged to overcome automatic selection of the irrelevant dimension. Fully elucidating these relationships will require further experimentation.

A recent experiment in nonhuman primates used a similar task to study the effects of context on the representation of evidence and choice in caudal prefrontal cortex (Mante et al. 2013). The results were consistent with a model where a top-down signal representing task context was responsible for determining how the relevant stream of visual evidence was gated into the integration process. The source of the context representation, however, and the parameters that influence its strength remained unspecified. Our results indicate that the signal may arise from the frontoparietal control network, including IFS and IPS, and that its strength is partly a function of expectations about context. It should be noted that the "late selection" model advocated by Mante and colleagues is not in tension with our proposed "early selection" mechanism: we focus on when, in the timecourse of a decision, the relevant dimension is selected, whereas Mante and colleagues focused on where, in the hierarchy of visual processing, that selection occurs. Although the timecourse of the context signal in the Mante et al. data is consistent with temporally early selection, their experiment used a fixed duration stimulus and a block design, making it difficult to directly compare the patterns of behavioral results that support our interpretation.

The active representation of task context, which provides top-down influence over evidence integration, is a fundamental mechanism of cognitive control (Miller and Cohen 2001). A central question in research on cognitive control concerns how the engagement of this mechanism is regulated (Botvinick and Cohen 2014). In one influential class of models, control is thought to be recruited in response to the detection of processing conflict (Botvinick et al. 2001). This basic perspective has been extended to predictive accounts, which argue that control can be engaged when conflict is perceived to be likely due to trial history or learned associations with other contextual factors (King et al. 2012; Jiang et al. 2014). Models that emphasize adaptive predictions have also been proposed to account for variability in response inhibition (Brown and Braver 2005; Ide et al. 2013), and there is evidence that task switch costs are lower when switches are predictable (Monsell et al. 2003; Crump and Logan 2010; De Baene and Brass 2013). These approaches share a basic architecture in which a variable that places demands on control, such as conflict, is either detected or predicted with the goal of properly specifying the intensity of control.

The account presented here differs from these previous efforts in a subtle but fundamental way. What our model learns is not how much control is likely to be needed on each trial but rather what kind of decision is likely to be encountered. Predictions about task context thus do not directly specify control intensity. Instead, our account portrays control as emerging, at least in part, directly from a mismatch between expectations and reality. This associates control with the concept of prediction error in a way that previous models do not. It is this association that allows our account to provide a parsimonious explanation for the minimization of control engagement. To see the distinction, consider what happens when a learner that expects frequent conflict encounters an incongruent Stroop trial (Jiang et al. 2014). For this learner, prediction error would be relatively low, but demands on control would still be relatively high. While our model is therefore distinct from previous efforts, it is not in tension with them, as there are multiple influences on the demand for cognitive control, and learning can operate in parallel over these different levels of information. Our experiment, however, did not afford predictions about conflict, errors, or switches. Instead, our results demonstrate that a predictive model learning over representations of task context is sufficient to account for substantial modulation of control systems in the human brain.

In its emphasis on prediction error and the reduction of surprise, our account bears some similarity to predictive coding models of perception (Rao and Ballard 1999; Ouden et al. 2012; Summerfield and Lange 2014). These models provide a formal theory to explain why unexpected stimuli drive sensory responses more strongly than expected stimuli do (Summerfield et al. 2008; Egner et al. 2010; Kok et al. 2012) and can account for surprise over learned stimulus-stimulus or stimulus-response associations (Strange et al. 2005; Ouden et al. 2010; Iglesias et al. 2013). Our results indicate that surprise over abstract representations of task context modulates activity in more anterior regions of prefrontal cortex than has been observed in previous experiments. This is consistent with models of prefrontal organization that separate the source of control over abstract and concrete representations along a rostro-caudal axis (Koechlin and Summerfield 2007; Badre 2008). It has been argued that this functional architecture can support learning in cases where rules must be discovered using feedback (Badre et al. 2010; Donoso et al. 2014). Our results suggest that it also supports learning in environments where the rule for each decision is explicitly cued but predictable, such that learning can enable more efficient decision-making.

The results of our dimensionality reduction analysis, however, draw a contrast with previously observed perceptual expectation effects. Although early visual regions exhibit reduced activation in response to expected stimuli, there is evidence that expectation actually sharpens feature tuning in these regions (Kok et al. 2012). This distinction can likely be explained by the differences between sensory and association cortex. While neurons in early sensory cortex are tuned for specific stimulus features, neurons in frontoparietal association cortex are highly adaptive and alter their responses to reflect the information that is relevant for control in a particular context (Miller and Cohen 2001; Duncan 2013). Our results indicate that, at the population level, this adaptation is graded and that the representation of context is strongest under high demands on control. The relationship between demands on control and the strength of the frontoparietal context representation has not been systematically evaluated in previous work, but our findings are consistent with the observations that distributed information about context in frontoparietal cortex (Waskom et al. 2014), together with response amplitude (De Baene et al. 2012), gradually declines following a context switch and, reflecting learning on a different timescale, becomes substantially reduced following extensive training on a task (Woolgar et al. 2011).

Our data provide compelling evidence that the frontoparietal control network is engaged during relatively controlled decisions, but they leave open the question of what mechanisms enable relatively automatic decision-making under conditions of low surprise. It is likely that an account of these mechanisms will involve the basal ganglia. The basal ganglia system is associated with both habitual control over behavior and relatively fast incremental learning dynamics that allow habits to be adapted to the environment (Packard and Knowlton 2002; Buschman and Miller 2014). Through recurrent connections with premotor and motor cortex, the basal ganglia exerts influence over the selection of cortical representations that span from low-level motor commands to higher-level action policies (Alexander et al. 1986; Redgrave et al. 1999). It has been proposed that these selection dynamics can be, in turn, modified by input from more anterior prefrontal regions that correspond with areas strongly modulated by CPE in our task (Badre and Frank 2011; Frank and Badre 2011). Our data align with this model: the basal ganglia were robustly activated by task events but did not appear to be strongly modulated by CPE. Therefore, what differs between automatic and controlled decision-making may not be the overall engagement of the basal ganglia but whether selection on each trial is driven automatically through the strength of striatal synaptic weights, which are tuned by preceding experience, or by supervisory input from prefrontal cortex, which reacts to the particular task demands and how they align with the predictions that emerged from learning. Future work could explore these issue by extending elements of our paradigm to animal models that afford different approaches to measurement and thus may be more sensitive to such mechanisms.

In addition to lateral frontoparietal regions, we also found that responses in two medial parietal regions, PCC and mSPL, were strongly modulated by CPE. It has recently been proposed that these regions, which emerge as a coherent network from analyses of task-independent connectivity (Yeo et al. 2011), subserve recognition memory with graded responses that reflect familiarity (Gilmore et al. 2015). It is difficult to account for our effects with the model proposed by Gilmore and colleagues (2015), however. This model emphasizes enhanced responses over repeated encounters with an item, but we found that these regions responded most strongly on trials with high amounts of surprise, which is more closely related to novelty than to familiarity. Moreover, our results are broadly consistent with the observations that mSPL is engaged during policy shifts at multiple levels of processing (Chiu and Yantis 2009) and that, in macaques, the PCC represents evidence for environmental changes that signal the need for such shifts (Pearson et al. 2011). It is only recently appreciated that these regions form a network independent from nearby default mode regions, and the computations that this network performs remain poorly understood. Our results imply that a full account of its function should consider a broad range of cognitive paradigms.

The ideal observer modeling approach allowed us to identify a latent environmental variable, context prediction error, that has substantial influence on both behavioral and neural signatures of controlled decision-making. Our findings thus make progress towards an account of how control is regulated at the computational level (Marr 1982), but they leave unspecified the algorithm by which environmental statistics are learned and updated. Based on our results, it may be possible to develop a model of this algorithm that can be fit to behavior using RTs, rather than, or together with, choice data. This could potentially be accomplished within the drift-diffusion modeling framework, which can account for both choice and RT in simple decision tasks (Ratcliff and McKoon 2008). As our results suggest that RTs were the product of a multi-stage process where the relevant evidence was first selected and then integrated, however, it may be necessary to modify standard diffusion models of two-alternative decisions to account for the complexities associated with context-dependent choice.

It also remains unknown how different neural systems interact to shape expectations about task context and other abstract representations involved in control. An intriguing possibility is that the engagement of control might itself signal the magnitude of context prediction error and hence update the system that generates those predictions. This hypothesis draws on the complementary relationship between control and surprise, which should both be minimized to enable efficient and adaptive processing. Perhaps the brain exploits this symmetry, which would provide a parsimonious solution for achieving optimal behavior in dynamic environments. While speculative, this hypothesis emphasizes the fundamental importance of learning across multiple domains of cognitive processing. Understanding how such learning is achieved, and the full extent of its relationship to decision-making, will contribute to a more general understanding of cognition and may begin to explain how we successfully navigate a complex and constantly changing world.

Funding

This work was supported by the Stanford Center for Cognitive and Neurobiological Imaging and the Knut and Alice Wallenberg Foundation (Network Initiative on Culture, Brain, and Learning). M.L.W. was supported as an NSF IGERT Traineeship Recipient under Award 0801700.

Acknowledgements

We thank Serra Favila and Karen LaRocque for assistance with data collection and helpful discussions.

References

Alexander GE, DeLong MR, Strick PL. 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual Review of Neuroscience. 9:357–381.

Avants BB, Epstein CL, Grossman M, Gee JC. 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis. 12:26–41.

Badre D. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends in cognitive sciences. 12:193–200.

Badre D, Frank MJ. 2011. Mechanisms of Hierarchical Reinforcement Learning in Cortico-Striatal Circuits 2: Evidence from fMRI. Cerebral cortex (New York, NY : 1991).

Badre D, Kayser AS, D'Esposito M. 2010. Frontal cortex and the discovery of abstract action rules. Neuron. 66:315–326.

Barr DJ, Levy R, Scheepers C, Tily HJ. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 68:255–278.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. 2007. Learning the value of information in an uncertain world. Nature neuroscience. 10:1214–1221.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. Psychological Review. 108:624–652.

Botvinick MM, Cohen JD. 2014. The computational and neural basis of cognitive control: charted territory and new frontiers. Cognitive Science. 38:1249–1285.

Botvinick MM, Niv Y, Barto AC. 2009. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. Cognition. 113:262–280.

Brown JW, Braver TS. 2005. Learned predictions of error likelihood in the anterior cingulate cortex. Science (New York, NY). 307:1118–1121.

Buschman TJ, Denovellis EL, Diogo C, Bullock D, Miller EK. 2012. Synchronous Oscillatory Neural Ensembles for Rules in the Prefrontal Cortex. Neuron. 76:838–846.

Buschman TJ, Miller EK. 2014. Goal-direction and top-down control. Philosophical Transactions of the Royal Society B: Biological Sciences. 369.

Chiu Y-C, Yantis S. 2009. A domain-independent source of cognitive control for task sets: shifting spatial attention and switching categorization rules. Journal of Neuroscience. 29:3930–3938.

Cohen JD, Dunbar K, McClelland JL. 1990. On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychological Review. 97:332–361.

Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS. 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. Nature neuroscience. 16:1348–1355.

Crump MJC, Logan GD. 2010. Contextual control over task-set retrieval. Attention, perception & psychophysics. 72:2047–2053.

Cunningham JP, Yu BM. 2014. Dimensionality reduction for large-scale neural recordings. Nature neuroscience. 17:1500–1509.

Dale AM, Fischl B, Sereno MI. 1999. Cortical Surface-Based Analysis I. Segmentation and Surface Reconstruction. NeuroImage. 9:179–194.

De Baene W, Brass M. 2013. Switch probability context (in)sensitivity within the cognitive control network. NeuroImage. 77:207–214.

De Baene W, Kühn S, Brass M. 2012. Challenging a decade of brain research on task switching: Brain activation in the task-switching paradigm reflects adaptation rather than reconfiguration of task sets. Human Brain Mapping. 33:639–651.

Donoso M, Collins A, Koechlin E. 2014. Foundations of human reasoning in the prefrontal cortex. Science (New York, NY). 344:1481–1486.

Duncan J. 1980. The demonstration of capacity limitation. Cognitive psychology. 12:75–96.

Duncan J. 2013. The Structure of Cognition:Attentional Episodes in Mind and Brain. Neuron. 80:35–50.

Egner T, Hirsch J. 2005. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nature neuroscience. 8:1784–1790.

Egner T, Monti JM, Summerfield C. 2010. Expectation and surprise determine neural population responses in the ventral visual stream. Journal of Neuroscience. 30:16601–16608.

Fischl B. 2012. FreeSurfer. NeuroImage. 62:774–781.

Fischl B, Sereno MI, Tootell RB, Dale AM. 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping. 8:272–284.

Frank MC. 2013. Throwing out the Bayesian baby with the optimal bathwater: response to Endress (2013). Cognition. 128:417–423.

Frank MJ, Badre D. 2011. Mechanisms of Hierarchical Reinforcement Learning in Corticostriatal Circuits 1: Computational Analysis. Cerebral cortex (New York, NY : 1991).

Gilmore AW, Nelson SM, McDermott KB. 2015. A parietal memory network revealed by multiple MRI methods. Trends in cognitive sciences. 1–10.

Gold JI, Shadlen MN. 2007. The neural basis of decision making. Annual Review of Neuroscience. 30:535–574.

Goldman-Rakic PS. 1987. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Plum F, editor. Bethesda, MD: Handbook of physiology, Section 1: The nervous system. Vol. 5: Higher functions of the brain. p. 373–416.

Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS. 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Frontiers in neuroinformatics. 5:13.

Greve DN, Fischl B. 2009. Accurate and robust brain image alignment using boundary-based registration. NeuroImage. 48:63–72.

Griffiths TL, Chater N, Norris D, Pouget A. 2012. How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). Psychological Bulletin. 138:415–422.

Ide JS, Shenoy P, Yu AJ, Li C s R. 2013. Bayesian Prediction and Evaluation in the Anterior Cingulate Cortex. Journal of Neuroscience. 33:2039–2047.

Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, Ouden HEM den, Stephan KE. 2013. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron. 80:519–530.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. 2012. FSL. NeuroImage. 62:782–790.

Jiang J, Heller K, Egner T. 2014. Bayesian modeling of flexible cognitive control. Neuroscience and biobehavioral reviews. 46 Pt 1:30–43.

Kahneman D. 1973. Attention And Effort. Prentice Hall.

Kayser AS, Buchsbaum BR, Erickson DT, D'Esposito M. 2010. The functional anatomy of a perceptual decision in the human brain. Journal of Neurophysiology. 103:1179–1194.

Kayser AS, Erickson DT, Buchsbaum BR, D'Esposito M. 2010. Neural representations of relevant and irrelevant features in perceptual decision making. Journal of Neuroscience. 30:15778–15789.

Kiani R, Hanks TD, Shadlen MN. 2008. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. Journal of Neuroscience. 28:3017–3029.

King JA, Korb FM, Egner T. 2012. Priming of control: implicit contextual cuing of top-down attentional set. Journal of Neuroscience. 32:8192–8200.

Koechlin E, Ody C, Kouneiher F. 2003. The Architecture of Cognitive Control in the Human Prefrontal Cortex. Science (New York, NY). 302:1181–1185.

Koechlin E, Summerfield C. 2007. An information theoretical approach to prefrontal executive function. Trends in cognitive sciences. 11:229–235.

Kok P, Jehee JFM, Lange FP de. 2012. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron. 75:265–270.

Kool W, McGuire JT, Rosen ZB, Botvinick MM. 2010. Decision making and the avoidance of cognitive demand. Journal of experimental psychology General. 139:665–682.

Law C-T, Gold JI. 2009. Reinforcement learning can account for associative and perceptual learning on a visual-decision task. Nature neuroscience. 12:655–663.

Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature. 503:78–84.

Marois R, Ivanoff J. 2005. Capacity limits of information processing in the brain. Trends in cognitive sciences. 9:296–305.

Marr DC. 1982. Vision. MIT Press.

McGuire JT, Botvinick MM. 2010. Prefrontal cortex, cognitive control, and the registration of decision costs. Proceedings of the National Academy of Sciences. 107:7922–7926.

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 24:167–202.

Monsell S, Sumner P, Waters H. 2003. Task-set reconfiguration with predictable and unpredictable task switches. Memory & cognition. 31:327–342.

Ouden HEM den, Daunizeau J, Roiser J, Friston KJ, Stephan KE. 2010. Striatal prediction error modulates cortical coupling. Journal of Neuroscience. 30:3210–3219.

Ouden HEM den, Kok P, Lange FP de. 2012. How prediction errors shape perception, attention, and motivation. Frontiers in Psychology. 3:548.

Packard MG, Knowlton BJ. 2002. Learning and memory functions of the Basal Ganglia. Annual Review of Neuroscience. 25:563–593.

Pearson JM, Heilbronner SR, Barack DL, Hayden BY, Platt ML. 2011. Posterior cingulate cortex: adapting behavior to a changing world. Trends in cognitive sciences. 15:143–151.

Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience. 2:79–87.

Ratcliff R, McKoon G. 2008. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. Neural computation. 20:873–922.

Redgrave P, Prescott TJ, Gurney K. 1999. The basal ganglia: a vertebrate solution to the selection problem? Neuroscience. 89:1009–1023.

Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S. 2013. The importance of mixed selectivity in complex cognitive tasks. Nature. 497:585–590.

Roitman JD, Shadlen MN. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. Journal of Neuroscience. 22:9475–9489.

Shenhav A, Botvinick MM, Cohen JD. 2013. The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. Neuron. 79:217–240.

Shiffrin RM, Schneider W. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review. 84:127.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. 2013. Dynamic coding for cognitive control in prefrontal cortex. Neuron. 78:364–375.

Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ. 2005. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural networks : the official journal of the International Neural Network Society. 18:225–230.

Summerfield C, Lange FP de. 2014. Expectation in perceptual decision making: neural and computational mechanisms. Nature Reviews Neuroscience. 15:745–756.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. Nature neuroscience. 11:1004–1006.

Wagner AD, Bunge SA, Badre D. 2004. Cognitive control, semantic memory, and priming: Contributions from prefrontal cortex. In: Gazzaniga MS, editor. The cognitive neurosciences. Cambridge, MA: MIT Press.

Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD. 2014. Frontoparietal representations of task context support the flexible control of goal-directed cognition. Journal of Neuroscience. 34:10743–10755.

Wilson RC, Nassar MR, Gold JI. 2010. Bayesian online learning of the hazard rate in change-point problems. Neural computation. 22:2452–2476.

Wilson RC, Niv Y. 2015. Is Model Fitting Necessary for Model-Based fMRI? PLoS Computational Biology. 11:e1004237.

Woolgar A, Hampshire A, Thompson R, Duncan J. 2011. Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex. Journal of Neuroscience. 31:14592–14599.

Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM. 2004. Multilevel linear modelling for FMRI group analysis using Bayesian inference. NeuroImage. 21:1732–1747.

Woolrich MW, Ripley BD, Brady M, Smith SM. 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage. 14:1370–1386.

Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology. 106:1125–1165.

Supplemental Figures

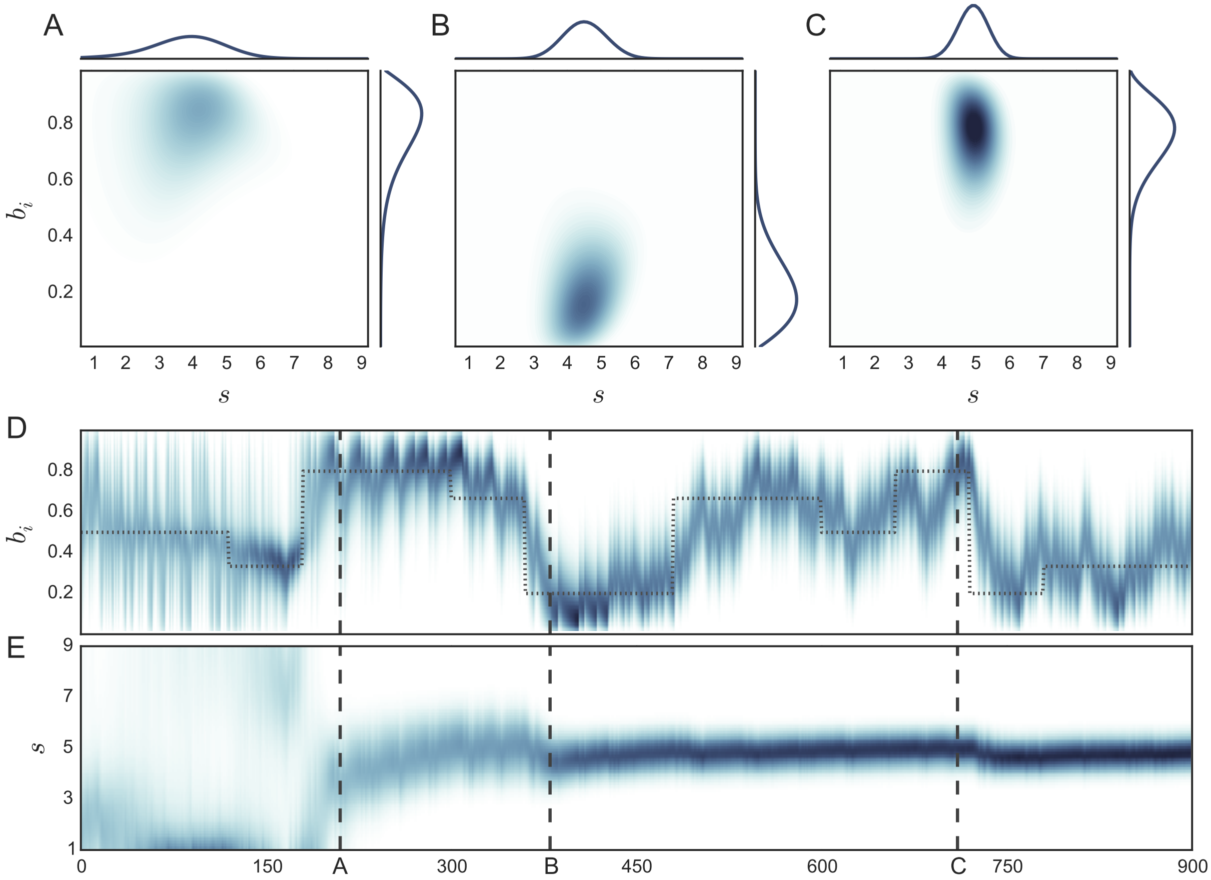

Figure 1: The ideal observer model in more detail. (a-c) The joint and marginal distributions of b_i and s on three trials in different periods of the experiment (indicated in panels *d* and *e*). Over time, the estimate of s becomes more certain, while the estimate of b_i shifts to reflect the changing environmental parameters. (D) The marginal posterior density on b_i for each trial. (E) The marginal posterior density on s for each trial.

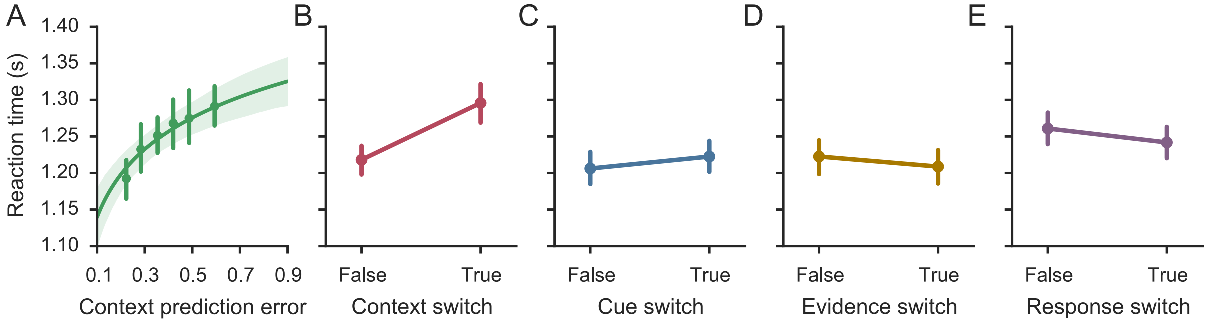

Figure 2: The effect of CPE compared with the repetitions and switches of different task variables. (A) The relationship between RT and log(CPE) replotted from Figure 4. (B) The effect of context switches, collapsing over CPE (note that context switches have higher CPE; see the histogram in Figure 4d.) (C) The effect of context cue switches while holding context constant. (D) The effect of coherent evidence switches while holding context constant. (E) The effect of response switches. Error bars on all panels show within-subject 95% confidence intervals.

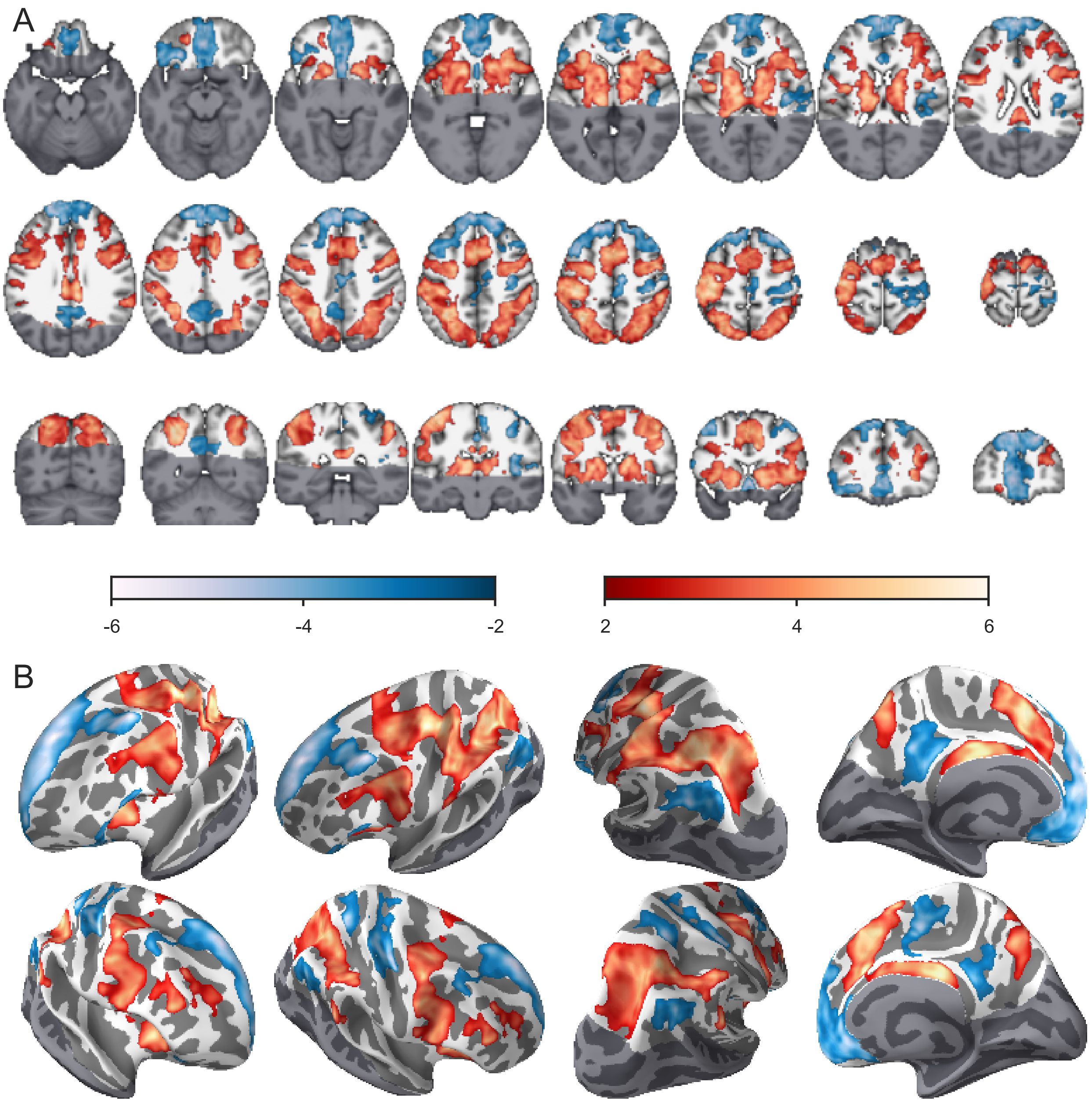

Figure 3: Mass-univariate analysis of positive and negative task-evoked responses. (A) Results from a volume-based mixed effects group analysis. The background image is the mean normalized anatomy for subjects included in the analyses. (B) Results from a surface-based random effects group analysis plotted on the Freesurfer average surface anatomy. In both panels, the statistical overlays are thresholded at $z > 2.3$ and cluster corrected to control the FWE rate at 0.05. The colormaps shows z scores, and voxels outside of the field of view are dimmed.

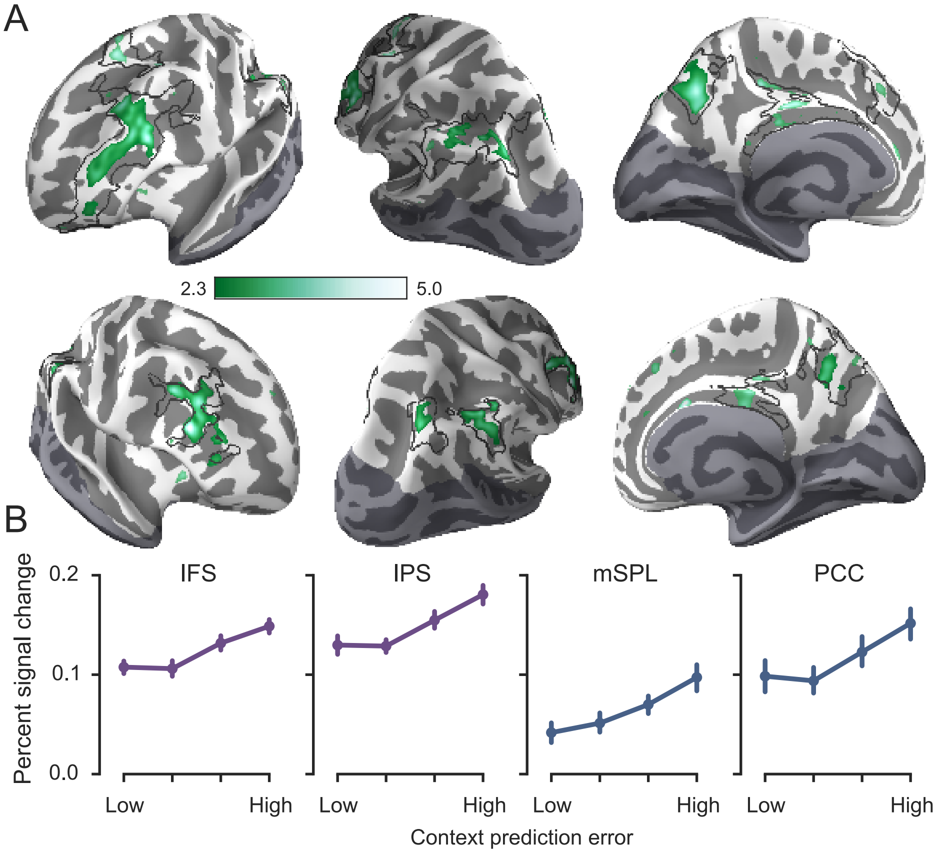

Figure 4: Context switching does not explain away the neural effect of CPE. (A) Mass-univariate analysis of CPE controlling for switches. The statistical overlay is shown in full opacity for clusters that are significant at a corrected level of $p < 0.05$ and at 50% opacity for clusters that are significant at an uncorrected level of $p < 0.005$. Black outlines show the extent of the CPE effect in the main model. Figure conventions are otherwise as in Figure 4. (B) Peak amplitude of the evoked response within bins defined by quartiles of CPE estimated while controlling for task switches. Figure conventions are as in Figure 5C.

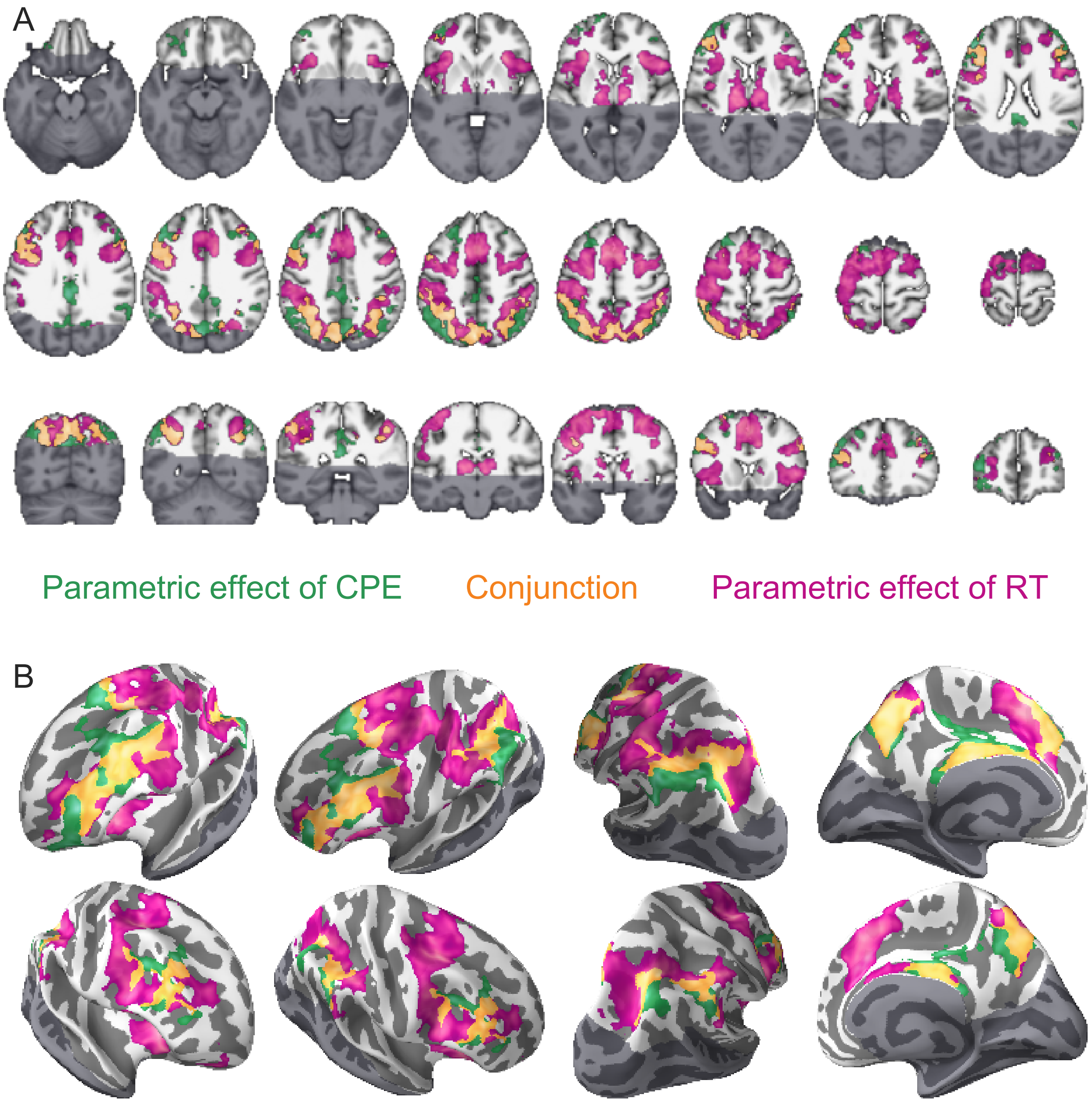

Figure 5: The conjunction between parametric effects of CPE and parametric effects of RT. (A) Results from a volume-based mixed effects group analysis. The background image is the mean normalized anatomy for subjects included in the analyses. (B) Results from a surface-based random effects group analysis plotted on the Freesurfer average surface anatomy. The overlays in both panels show z scores from the group analyses of the CPE and RT parameters in green and magenta, respectively. Each overlay was thresholded at $z > 2.3$ and cluster corrected to control the FWE rate at 0.05. Regions of overlap between the resulting maps, i.e. voxels where the formal conjunction test is significant, are shown in yellow. Voxels outside the field of view are dimmed.

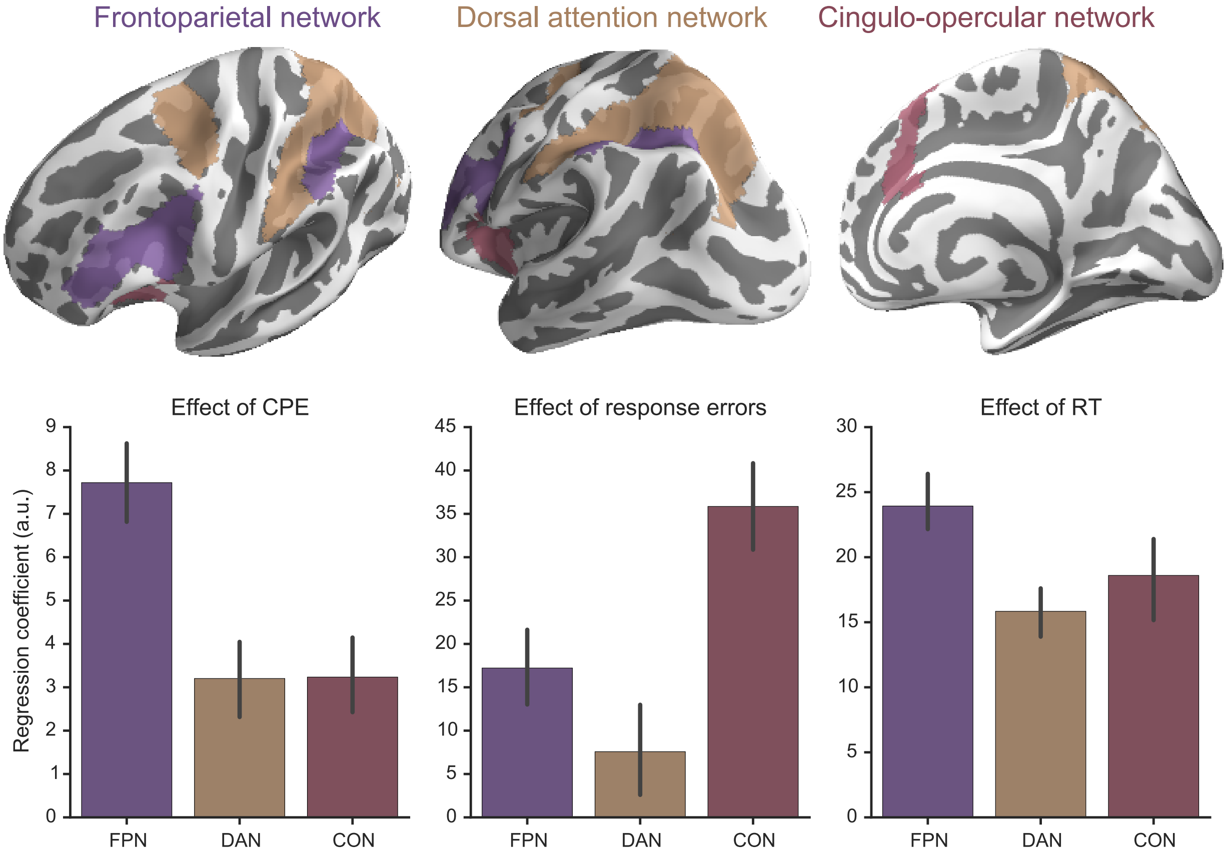

Figure 6: Comparison of task variables in three large-scale association cortex networks. The networks are shown on the Freesurfer average inflated surface. Bar plots show the average regression coefficients for three effects of interest after extraction from the unsmoothed, native space model. All networks were bilateral with clear left and right node definitions. Error bars show within-subject 95% confidence intervals.

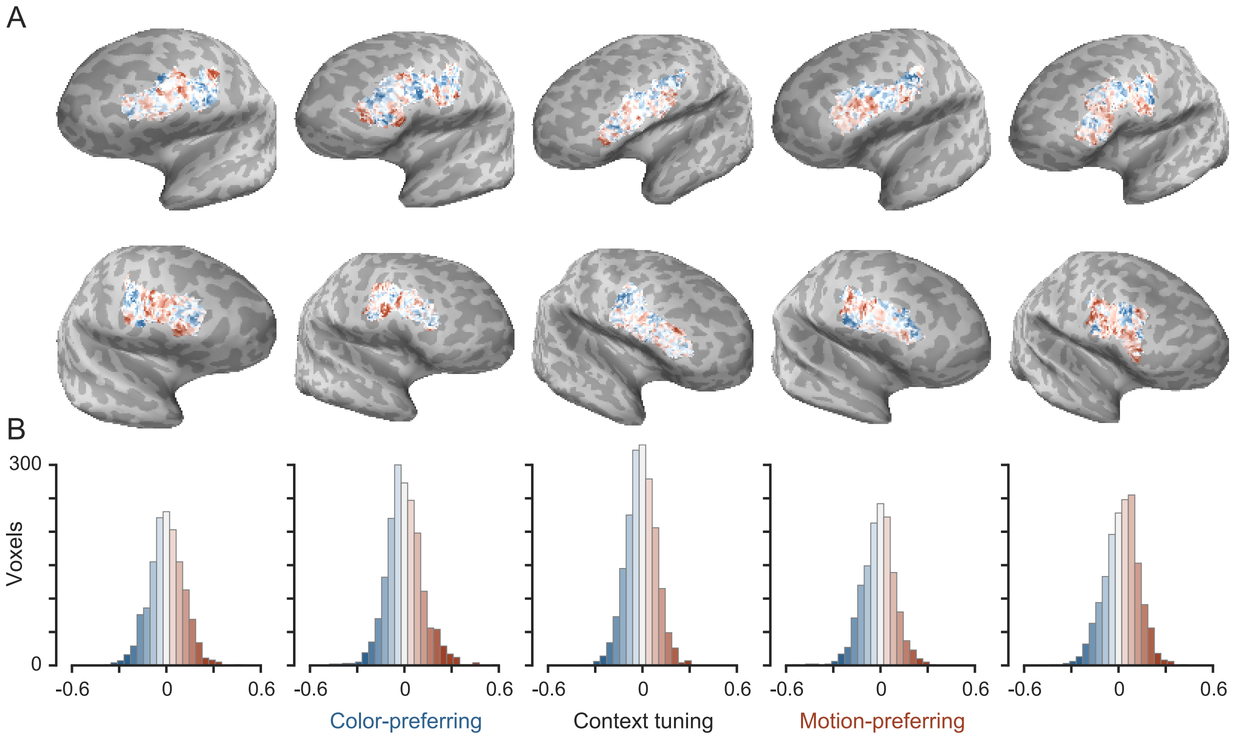

Figure 7: Context tuning has a fine-scale organization in prefrontal cortex. (A) Context tuning coefficients in the IFS are plotted on the inflated cortical mesh for five representative subjects. Voxels with similar context preferences tended to cluster together, although each subject had multiple clusters of voxels with strong context tunings, and no consistent between-subjects organization was apparent. (B) Distribution of the coefficients plotted in (A). The color of each bar corresponds to the colormap used in (A). These distributions were unimodal, centered on 0, and generally symmetric.