

# Context-specific behavioral surprise is differentially correlated with activity in anterior and posterior brain systems

Michael J. Tobia<sup>a,b</sup>, Jan Gläscher<sup>a</sup> and Tobias Sommer<sup>a</sup>

This experiment investigated whether behavioral surprise, an information-theoretic measure of the amount of memory and information integration associated with a response, is correlated with neural activity during decision making. A total of 30 participants (age 18–30) were scanned with functional MRI while completing 240 trials of a sequential decision-making task in which they selected an amount to wager from four possible values on each trial. Behavioral surprise was computed trial by trial using both context-free and context-specific formulations, and was used as a parametric modulator in functional MRI analyses. Whereas context-free surprise was not significantly correlated, two sets of clusters ( $P < 0.005$ ; cluster size  $> 156$  voxels) were differentially modulated by context-specific behavioral surprise. An anterior system comprised of the inferior frontal gyrus and anterior cingulate (each bilaterally), and left caudate, was positively modulated. A posterior system comprised of the posterior cingulate, parahippocampal gyrus and posterior hippocampus (each bilaterally), and left angular gyrus, was negatively modulated. These anticorrelated systems indicate that more surprising (resource demanding) actions recruit greater activity from

the anterior system and less activity from the posterior system and less surprising actions (memory-guided) recruit greater activity from the posterior system and less activity from the anterior system. These results show that context-specific behavioral surprise is a unique neural signal and may be related to mechanisms for both cognitive control and memory-guided behavior, and support contemporary theories that the brain is a statistical observer of external and internal events. *NeuroReport* 27:677–682 Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved.

*NeuroReport* 2016, 27:677–682

**Keywords:** behavioral surprise, context, information theory, memory-guided behavior, response selection

<sup>a</sup>Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany and <sup>b</sup>Department of Radiology, Center for NMR Research, Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

Correspondence to Michael J. Tobia, PhD, Center for NMR Research, Pennsylvania State Hershey College of Medicine, 500 University Drive, Hershey, PA 17033, USA  
Tel: +1 717 531 5858; fax: +1 717 531 8486; e-mail: mtobia@hmc.psu.edu

Received 16 March 2016 accepted 6 April 2016

## Introduction

Information theoretic surprise is a measure of the amount of information conveyed by a single event embedded within a stream of events (i.e. a signal), and novel or infrequent events are surprising [1]. Various formulations of surprise derived from information theory have been applied successfully in cognitive and neuroscience research to quantify the information in sensory events, including events that capture attention [2], cue behavioral responses [3–5], as well as the consequences of behavioral responses [6–8]. These formulations are typically interpreted as prediction errors with respect to sensory events that violate an expectation and are indicative that the brain functions as a computational learning machine that observes the statistics of exogenous sensory events [9,10].

In contrast to these formulations of surprise on the basis of sensory events, behavioral surprise can also be derived from information theory. It measures the amount of information conveyed by a particular response and implies different amounts of memory and processing demands associated with response selection processes. Responses that occur more frequently are less surprising, more strongly associated with memory, and less demanding of cognitive resources (i.e. responses that exploit knowledge),

whereas infrequent responses that are less associated with memory are more resource demanding. It was described previously (i.e. mathematically formulated) in relation to a theory of executive control [11], where it was designated as the amount of information integration required to emit a contextually appropriate response.

Context is an important variable in theories of memory [12] and cognitive control [11,13]. It refers to the conjunction of stimuli or stimulus conditions that are encoded into memory, and allows for a more precise description of the higher-order contingencies that bias behavior than the description provided by a single stimulus condition. Behavioral surprise is greater when response selection deviates from an established pattern of responding in a given context. For example, in the context of a sudden change in weather conditions, it may be expected to alter one's typical driving route to avoid potential hazards (i.e. slippery or flooded roads). The decision to maintain the typical driving route despite this context would be surprising because it deviates from the expected course of behavior. As such, context-specific behavioral surprise could be likened to a prediction error for response selection that is independent of prediction errors for expected consequences.

Despite its central role in a prominent theory of response selection [11] and its potential relation to other contemporary theories of cognitive brain functioning [9], behavioral surprise has not been formally computed and applied to study the neural correlates of response selection. In this study, we computed behavioral surprise trial by trial for a series of responses in a sequential choice task with both context-free and context-specific formulations, and used these as parametric modulators to identify brain systems involved in response selection.

## Methods

### Participants

Data from 30 healthy young adult men (age 18–30 years) from a previous experiment were used in this study and detailed analyses of task performance are reported elsewhere specifically in the context of a reinforcement learning computational model [14]. This work is a novel, alternative, and independent analysis of those data. All participants provided informed consent and protocols were carried out in accordance with the declaration of Helsinki.

### Strategic sequential investment task

Participants completed 240 trials of the strategic sequential investment task. Figure 1 (top) shows screenshots of an example strategic sequential investment task trial. Briefly, participants select an amount to wager (0, 1, 2 or 3) in a fictitious stock market on each trial. The fictitious stock market contains seven fictitious stocks, each indicated by a unique background stimulus. Only one fictitious stock is presented per trial, but each is associated with a different probability of winning/losing, and different payout, which together defines its expected value. Participants need to learn which stocks have higher expected values to maximize their winnings. Thus, on each trial, a participant may select one of four responses, which corresponds to the amount wagered that the fictitious stock will pay out rather than lose value.

### Computing context-free and context-specific behavioral surprise

Behavioral surprise was computed over the series of wagers in both context-free and context-specific formulations. Context-free surprise was calculated for the ordinal sequence of choices as the  $-\log_2 p(x)$ , where  $p(x)$  is the probability of the current response (amount wagered). Context-specific surprise was computed as the  $-\log_2 p(\text{choice}|\text{context})$ , where  $p(\text{choice}|\text{context})$  is the probability of the current wager conditional on the current stock. The procedure for computing behavioral surprise is shown in Fig. 1 (middle). Values for surprise were initialized to zero and accrued a frequency distribution of events throughout the task. The first 30 trials were not included to allow the stability of the statistical information to develop.

## MRI

MR images were acquired using a Siemens 3T Trio (Siemens, Munich, Germany) using a 32-channel coil. The scanning protocol included a T1-weighted MPRAGE from each participant (voxel resolution of  $1 \times 1 \times 1 \text{ mm}^3$ , coronal orientation, left–right phase encode,  $192 \times 256 \text{ mm}$  FoV, 240 slices, 1100 ms inversion time, TE = 2.98 ms, TR = 2300 ms, and  $9^\circ$  flip angle), and eight runs ( $\sim 10 \text{ min/run}$ ) of functional MRI using a T2\*-weighted GRAPPA EPI (TR = 2360 ms, TE = 25 ms, anterior–posterior phase encode, 40 slices acquired in the descending noninterleaved axial plane with  $2 \times 2 \times 2 \text{ mm}^3$  voxels;  $204 \times 204 \text{ mm}$  FoV; skip factor = 0.5).

### MRI data processing

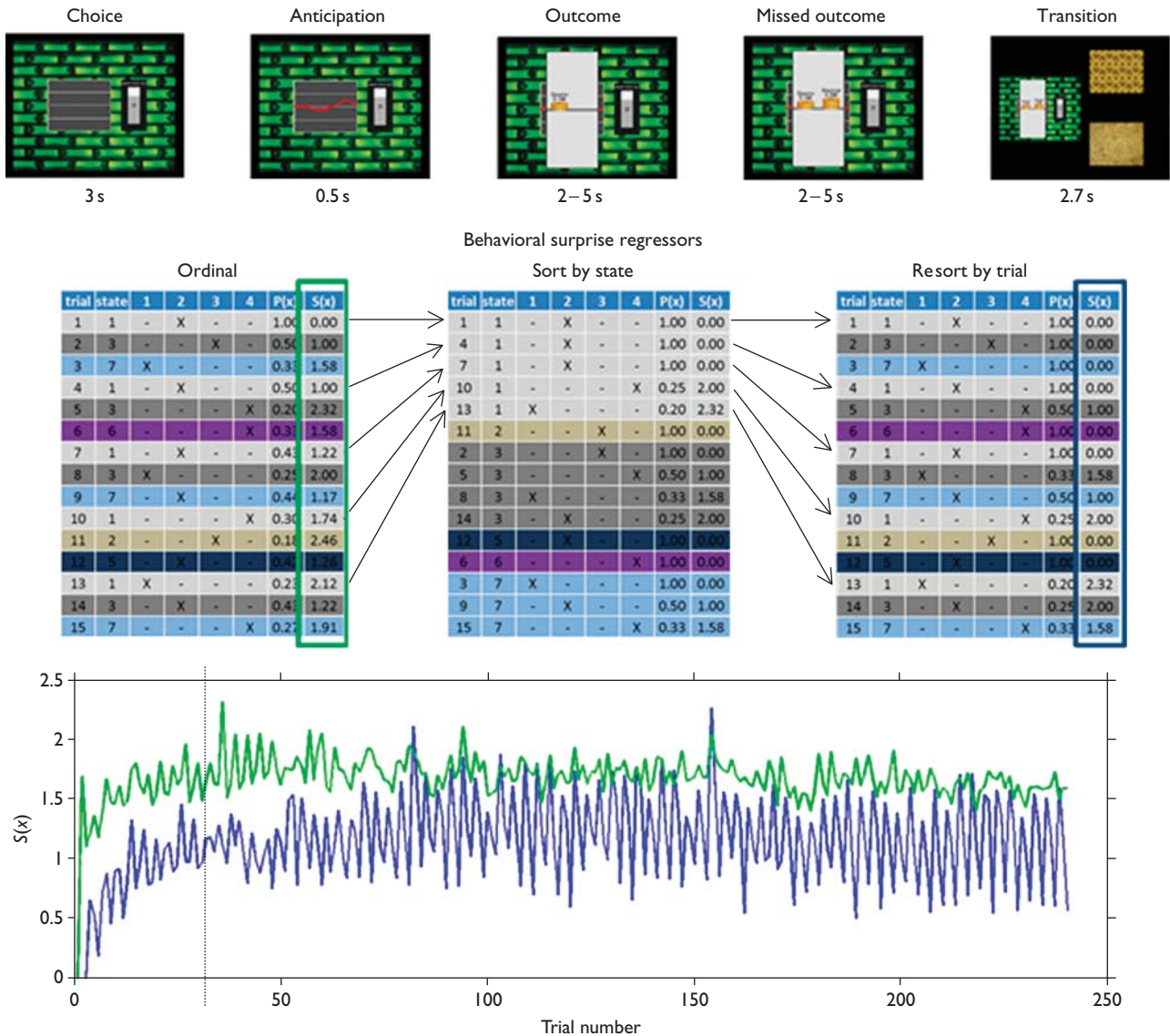
MR image analyses were carried out in SPM8 (Wellcome Trust, London, UK). Anatomical images were segmented and transformed into Montreal Neurological Institute (MNI) standard space. Functional images were corrected for slice-timing acquisition offsets, realigned, and corrected for the interaction of motion and distortion using unwarped toolbox, coregistered to anatomical images and transformed into MNI space using DARTEL, and smoothed ( $8 \text{ mm}^3$  full width at half maximum).

First-level analyses included regressors at the onset of each event (excluding the anticipation event) as well as parametric modulators for nominal value of the wager (0–3) and behavioral surprise at the choice event, and reward value at the outcome event. Coincident parametric modulators were orthogonalized as by default in SPM8 (i.e. behavioral surprise was orthogonalized with respect to choice value). Context-free and context-specific behavioral surprise regressors were analyzed in separate first-level models while all other regressors remained the same. Second-level analyses were carried out using a series of one-sample *t*-tests. Results are reported at a whole-brain cluster threshold correction using AlphaSim [15], which indicated that a single-voxel threshold of *P* less than 0.005 and a cluster of 156 voxels achieved a whole-brain statistical correction to *P* less than 0.05. Spatial coordinates (*x*, *y*, *z*) reported in the results are in MNI space (neurological convention) and correspond to the peak voxel from significant activity for each cluster from behavioral surprise and the whole-brain peak voxel for choice (wager) and reward values.

## Results

Parametric modulators for choice value, reward value, and behavioral surprise each correlated with neural activity. Context-free behavioral surprise was negatively correlated with a small cluster of activity (20 voxels; not shown) in the right posterior parietal cortex (peak: 54, –66, 34), which did not survive the whole-brain statistical correction. Figure 2 shows results for context-specific behavioral surprise, which correlated positively with an anterior network and negatively with a posterior network. The positively modulated anterior network consisted of

Fig. 1

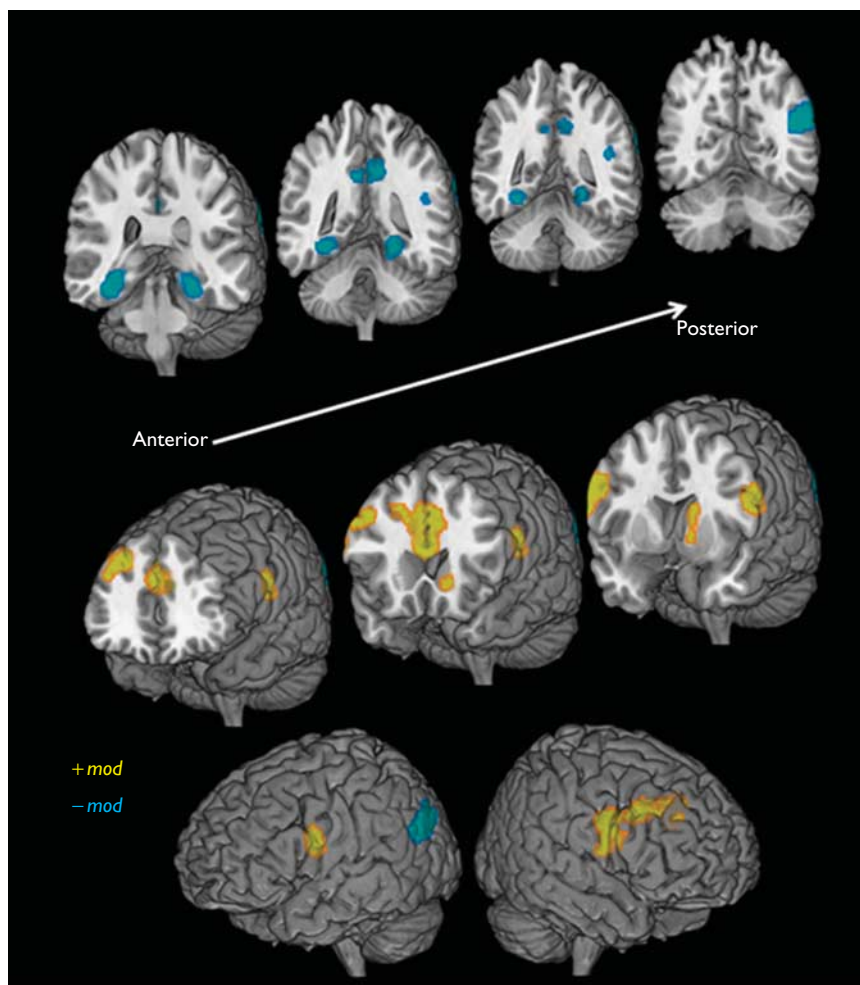


Top: screenshots of an example trial from the SSIT show five stimulus events and their durations per trial. Participants make a choice by selecting an amount to wager during the choice event, and then anticipate and subsequently observe the outcome. Counterfactual outcomes and state transitions are then presented. Middle: computation of the context-free and context-specific behavioral surprise regressors follows two procedures. The procedure for computing the context-free behavioral surprise is shown on the left, for which the actual sequence of trial-by-trial choices is used. The sorting procedure for generating context-specific behavioral surprise is shown in the center. These values are then resorted (right) to produce a trial-by-trial time series of context-specific behavioral surprise values. Bottom: the group average context-free (green) and context-specific (blue) behavioral surprise time series are shown. The dashed vertical line shows that the first 30 trials were excluded from fMRI analyses to avoid the effects of an initial transient. fMRI, functional MRI; SSIT, strategic sequential investment task.

the left caudate (peak:  $-18, 10, 14, t=3.54$ ), anterior cingulate cortex (ACC) (peak:  $-2, 24, 18; t=5.25$ ) and posterior portions of the inferior frontal gyrus (left peak:  $-60, 6, 24; t=3.72$ ; right peak:  $58, 6, 26; t=4.25$ ). The negatively modulated posterior network consisted of the hippocampus and the parahippocampal gyrus (left peak:  $-22, -38, -8; t=-4.45$ ; right peak:  $26, -36, -8; t=-4.01$ ), as well as the posterior cingulate cortex (PCC) (peak:  $-12, 40, 42; t=-3.69$ ) and left angular gyrus (peak:  $-54, -70, 30; t=-4.07$ ). Figure 3 shows the results for the

choice and reward value parametric modulators, as well as the context-specific behavioral surprise regressor to show the spatial independence and overlap of each. Choice value (green clusters) correlated with activity in the ventral striatum (peak:  $10, 8, -6; t=3.62$ ) and the ACC. Reward value (red clusters) correlated with activity in a widely distributed network including the ventral visual cortex (occipital and temporal), the posterior parietal cortex, medial temporal lobes, the striatum (peak:  $12, 14, -10; t=6.63$ ), ventral PFC, ACC, PCC, as well as anterior

Fig. 2



Context-specific behavioral surprise positively correlated with an anterior network (yellow clusters) and negatively correlated with a posterior network (blue clusters). Activity is shown at a single-voxel threshold of  $P < 0.005$  thresholded by 156 contiguous voxels and superimposed on a rendering of the group average T1 image in MNI space (rendered with MRICroGL, v.12 <http://www.mccauslandcenter.sc.edu/mricrogl/>). Images are shown in neurological convention (left = left). MNI, Montreal Neurological Institute.

and posterior lateral PFC. The figure shows that each of the three regressors are primarily spatially independent; however, all three overlap in the ACC.

## Discussion

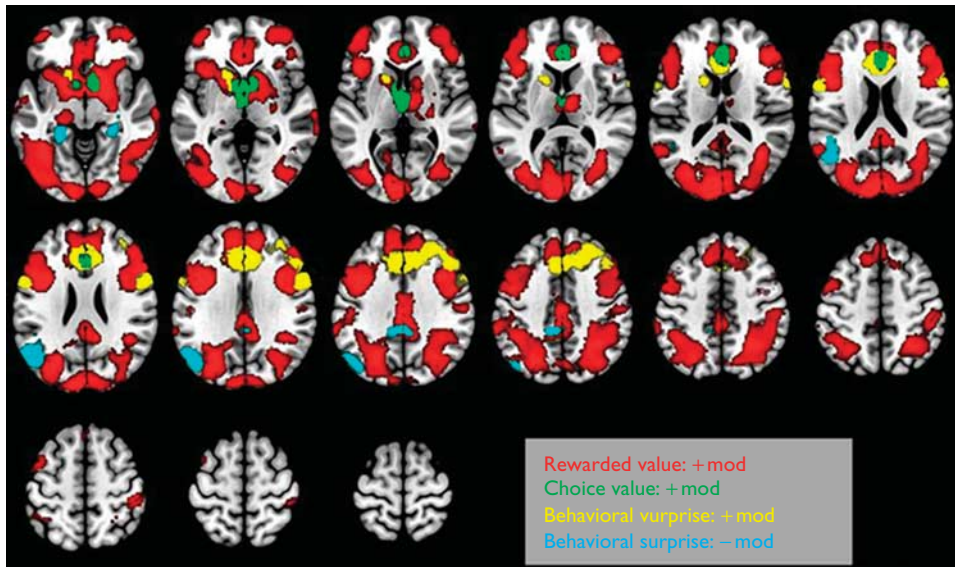
The most noteworthy finding from this study is that context-specific behavioral surprise is not only correlated with neural activity, but that it is differentially correlated with anterior and posterior brain systems associated previously with executive control and memory-guided behavior, respectively. The positively correlated anterior system indicates that more surprising response selection recruits greater activity (i.e. more resource demanding) and less surprising actions recruit less activity. The negatively correlated posterior system indicates that more surprising responses recruit less activation and less surprising responses recruit more activation. Thus, response

selection may involve a balance of activation between anterior and posterior systems.

Koechlin and Summerfield [11] proposed that context-specific behavioral surprise corresponds to the magnitude of information processing (or integration) necessary for response selection on a given trial. In their information theoretical approach, context-free behavioral surprise corresponds to sensory control [i.e.  $-\log_2 p(\text{choice})$ ], for which we did not find a strong neural representation, although a few voxels in the posterior parietal cortex were negatively correlated (did not survive the cluster threshold). Context-specific behavioral surprise corresponds to contextual cognitive control [i.e.  $-\log_2 p(\text{choice}|\text{context})$ ], for which Koechlin and Summerfield predicted activity in the posterior inferior frontal gyrus, precisely as we found. Our findings therefore support their information theoretical approach to prefrontal executive functions.



Fig. 3



Context-specific behavioral surprise (yellow and blue clusters) show spatial independence with respect to the nominal value of the wager (green clusters), and reward value (red clusters), suggesting that it is a unique neural signal. The only region where all three signals overlap is in the anterior cingulate cortex. Activity is shown at a single-voxel threshold of  $P < 0.005$  thresholded by 156 contiguous voxels and superimposed on a rendering of the group average T1 image in MNI space (rendered with MRICroGL, v.12). Images are shown in neurological convention (left = left). MNI, Montreal Neurological Institute.

The hippocampus and parahippocampal gyrus, along with the PCC and angular gyrus, comprise a network for memory-guided behavior [16]. This network, which corresponds to the posterior regions that were negatively correlated with context-specific behavioral surprise in this study, is postulated to bind relationships among entities, actions, and consequences into a schema-like model that guides behavior. The negative correlation with behavioral surprise supports this interpretation because less surprising actions, such as those having greater previous experience and memory associations, recruit greater activation from this system than the anterior system involved in executive control (and more surprising actions recruit less activation from this system).

The ACC is associated with response selection mechanisms when there are strong demands for cognitive control, such as the need to resolve conflict among competing response alternatives [17]. Behavioral surprise, the nominal value of the current action, and subsequent reward processing overlapped prominently in the ACC, but were otherwise spatially independent. This suggests that neural activity in the ACC during response selection represents or integrates diverse types of information from multiple brain systems, including executive control systems and memory-guided behavior systems, which may allow it to weigh evidence from memory alongside evidence from the current environmental conditions so as to best resolve competition.

Importantly, context-specific behavioral surprise is a unique neural signal that is temporally orthogonal and spatially dissociable from the nominal choice wager, and does not appear to overlap with regions representing expected value during choice behavior [14], nor subsequent reward processing. Moreover, it is neurally distinct from exploratory behavioral responses that activate the intraparietal sulcus [18], as well as the rostrolateral PFC that is involved in behavioral exploration driven by uncertainty [19]. As such, behavioral surprise may play an important role in understanding cognitive brain systems that are distinct from the effects of surprise associated with sensory events and exploratory decision making.

## Conclusion

Context-specific behavioral surprise is a unique neural signal that is encoded in at least two brain systems, each associated with a different mechanism of controlling behavior [10–11,16]. Future research should consider whether and how these two brain systems compete for control over response selection and how such competition is resolved, which may involve information integration in the ACC. Furthermore, these findings support theories that the brain is a Bayesian (or statistical) observer [9] not only of exogenous sensory events but also of its own behavior.

## Acknowledgements

### Conflicts of interest

There are no conflicts of interest.

**References**

- 1 Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948; **27**:379–423.
- 2 Itti L, Baldi P. Bayesian surprise attracts human attention. *Vision Res* 2008; **49**:1295–1306.
- 3 Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ. Information theory, novelty and hippocampal response: unpredicted or unpredictable? *Neural Net* 2005; **18**:225–230.
- 4 D'Acromont M, Schultz W, Bossaerts P. The human brain encodes event frequencies while forming subjective beliefs. *J Neurosci* 2013; **32**:10887–10897.
- 5 Zhang J, Rowe JB. The neural signature of information regularity in temporally extended event sequences. *NeuroImage* 2015; **107**:266–276.
- 6 Bornstein AM, Daw ND. Cortical and hippocampal correlates of deliberation during model-based decision for reward in humans. *PLoS Comput Biol* 2013; **9**:e1003387.
- 7 Chumbley JR, Burke CJ, Stephan KE, Friston KJ, Tobler PN, Fehr E. Surprise beyond prediction error. *Hum Brain Mapp* 2014; **35**:4805–4814.
- 8 Roesch MR, Esber GR, Li J, Daw ND, Schoenbaum G. Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain. *Eur J Neurosci* 2012; **35**:1190–1200.
- 9 Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. Action and behavior: a free-energy formulation. *Biol Cybern* 2010; **102**:227–260.
- 10 Fan J. An information theory account of cognitive control. *Front Hum Neurosci* 2014; **8**:680.
- 11 Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn Sci* 2007; **11**:229–235.
- 12 Rugg MD, Vilberg KL, Mattson JT, Yu SS, Johnson JD, Suzuki M. Item memory, context memory, and the hippocampus: fMRI evidence. *Neuropsychologia* 2012; **50**:3070–3079.
- 13 Koechlin E, Ody C, Kounheir F. The architecture of cognitive control in the human prefrontal cortex. *Science* 2003; **302**:1181–1185.
- 14 Tobia MJ, Guo R, Schwarze U, Boehmer W, Gläscher J, Finckh B, et al. Neural systems for choice and valuation with counterfactual learning signals. *NeuroImage* 2014; **89**:57–69.
- 15 Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance images (fMRI): use of a cluster-size threshold. *Magn Reson Med* 1995; **33**:636–647.
- 16 Ranganath C, Ritchey M. Two cortical systems for memory-guided behavior. *Nat Rev Neurosci* 2012; **13**:713–726.
- 17 Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD. Conflict monitoring versus selection-for-action in anterior cingulate. *Nature* 1999; **402**:179–181.
- 18 Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature* 2006; **44**:876–879.
- 19 Badre D, Doll BB, Long NM, Frank MJ. Rostrolateral prefrontal cortex and individual differences in uncertainty-driven behavior. *Neuron* 2012; **73**:595–607.