Title:

Feature-specific prediction errors for visual mismatch

Abbreviated title:

Feature-specific prediction errors for visual mismatch

Gabor Stefanics [a,b], Klaas Enno Stephan [a,b,c], Jakob Heinzle [a]

[a] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Wilfriedstrasse 6, 8032 Zurich, Switzerland

[b] Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Blümlisalpstrasse 10, 8006 Zurich, Switzerland

[c] Max Planck Institute for Metabolism Research, Cologne, Germany

Corresponding author:

Gabor Stefanics. Address: Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Wilfriedstrasse 6, 8032 Zurich, Switzerland. Tel.: +41 (0) 44 634 9121, email: stefanics@biomed.ee.ethz.ch

Number of pages: 18

Number of figures: 5, tables: 1. Color should NOT be used for any figures in print

Number of words for Abstract: 229

Declaration of interest: none.

Acknowledgements:

We acknowledge support by the University of Zurich (KES), the René and Susanne Braginsky Foundation (KES), and the Clinical Research Priority Program “Multiple Sclerosis” (GS, KES).

Abstract

Predictive coding (PC) theory posits that our brain employs a predictive model of the environment to infer the causes of its sensory inputs. A fundamental but untested prediction of this theory is that the same stimulus should elicit distinct precision weighted prediction errors (pwPEs) when different (feature-specific) predictions are violated, even in the absence of attention. Here, we tested this hypothesis using functional magnetic resonance imaging (fMRI) and a multi-feature roving visual mismatch paradigm where rare changes in either color (red, green), or emotional expression (happy, fearful) of faces elicited pwPE responses in human participants. Using a computational model of learning and inference, we simulated pwPE and prediction trajectories of a Bayes-optimal observer and used these to analyze changes in blood oxygen level dependent (BOLD) responses to changes in color and emotional expression of faces while participants engaged in a distractor task. Controlling for visual attention by eye-tracking, we found pwPE responses to unexpected color changes in the fusiform gyrus. Conversely, unexpected changes of facial emotions elicited pwPE responses in cortico-thalamo-cerebellar structures associated with emotion and theory of mind processing. Predictions pertaining to emotions activated fusiform, occipital and temporal areas. Our results are consistent with a general role of PC across perception, from low-level to complex and socially relevant object features, and suggest that monitoring of the social environment occurs continuously and automatically, even in the absence of attention.

Keywords: predictive coding; precision weighted prediction error; color perception; emotion recognition; perception; perceptual inference

Highlights

Changes in color or emotion of physically identical faces elicit prediction errors

Prediction errors to such different features arise in distinct neuronal circuits

Predictions pertaining to emotions are represented in multiple cortical areas

Feature-specific prediction errors support predictive coding theories of perception

Introduction

Predictive coding (PC) postulates that perceptual inference rests on probabilistic (generative) models of the causes of the sensory input (Rao and Ballard, 1999; Friston, 2005; Clark, 2015). The theory emphasizes the active nature of perceptual inference: in contrast to theories that view perception as a reactive, feed-forward analysis of bottom-up sensory information (Hubel and Wiesel, 1965; Riesenhuber and Poggio, 2000), PC regards the brain as actively predicting the sensory signal, based on a hierarchical probabilistic model of the causes of its sensory signals (Egner et al., 2010; Friston, 2010; Lochmann et al., 2012; Bogacz, 2017). According to this theory, perception involves inferring the most likely cause of the sensory signals by integrating incoming sensory information at a given level in the hierarchy with predictions generated at the level above (Rao and Ballard, 1999; Lee and Mumford, 2003; Friston, 2005), where the latter derive from prior information. In this framework a unified perceptual representation of an object involves a set of hierarchical predictions that relate to the object's different attributes, such as spatiotemporal coordinates but also intrinsic structure. At each hierarchical level, incoming signals from the level below are compared to predictions from the level above, and the ensuing prediction errors (PEs) are passed to the higher level in order to update predictions.

PC thus offers a framework to describe how object representations emerge during hierarchical perceptual inference: segregation and integration of predicted lower-level and more abstract attributes take place in a probabilistic network bound together by passing messages between hierarchical levels that most effectively minimize perceptual PEs (Friston, 2005; Bogacz, 2017). In this framework, unexpected stimuli trigger PE responses which subside as stimuli become predictable, for example through repeated presentation.

PC has become one of the most influential theories of perception, and many of its implications have been confirmed experimentally (e.g., Smith and Muckli, 2010; Wacongne et al., 2011; Kok et al., 2012a,b; Durschmid et al., 2016; Sedley et al., 2016; Ehinger et al., 2017; Gordon et al., 2017; Schwiedrzik and Freiwald, 2017). One central question about the implementation of PC is whether the same physical stimulus elicits separable feature-specific PE responses when distinct predictions about its various attributes exist, regardless whether such attributes are behaviorally relevant. To our knowledge, this has only been studied under attention (Jiang et al., 2016), but not for automatic processing, in the absence of attention and task-relevance. To answer this question, we used a roving standard paradigm (Fig.1A) to systematically manipulate predictions of two attributes of complex stimuli, the color and emotional expression of faces. Based on prior event-related brain potential (ERP) studies, we used a visual mismatch paradigm (for reviews, see Stefanics et al., 2014; Kremlacek et al., 2016) to study brain responses reflecting PEs and model updating processes elicited by unexpected changes in color and facial emotion while participants engaged in a distractor task.

We used the Hierarchical Gaussian Filter (HGF, Mathys et al., 2011; Mathys et al., 2014) to simulate belief trajectories of an ideal Bayesian observer. The HGF is a computational model that allows for inferring an agent's beliefs and uncertainty about hidden states of the world that generate sensory information. The model tracks the beliefs of the agent about the probability of each stimulus feature and updates its inference as new information is presented trial-by-trial. The HGF implements a form of PC in the temporal

domain and has been used in multiple studies to investigate predictive processes in the brain (e.g., Iglesias et al., 2013; Schwartenbeck et al., 2015; Vossel et al., 2015; Auksztulewicz et al., 2017; Diaconescu et al., 2017; Lawson et al., 2017; Powers et al., 2017; Adams et al., 2018; Katthagen et al., 2018; Stefanics et al., 2018a).

In this paper we used a similar experimental paradigm, computational modeling and analysis approach as in a previous single-trial EEG study that allowed us to study the time course of event-related brain potentials (ERP) to unexpected color and emotion changes associated with pwPEs (Stefanics et al., 2018a). In this previous study, we found that both kind of changes elicited brain responses that were better explained with pwPEs as parametric regressors than regressors encoding categorical stimulus changes in a general linear modeling (GLM) analysis. Here, we used fMRI to identify the brain regions associated with feature-specific predictions and pwPEs to human faces. Critically, our paradigm independently manipulated the color and emotional expression of face stimuli (Fig. 1B, C), allowing us to model predictions and pwPEs to violations of emotion expectations separately from predictions and pwPEs elicited by changes in color. This enabled us to study predictive processes pertaining to low versus high level object features for physically identical stimuli.

Methods

Ethics Statement

The experimental protocol was approved by Cantonal Ethics Commission of Zurich (KEK 2010-0327). Written informed consent was obtained from all participants after the procedures and risks were explained. The experiments were conducted in compliance with the Declaration of Helsinki.

Subjects

Thirty-nine healthy, right-handed subjects participated in this experiment. One subject was excluded due to incomplete data, and three subjects' data of one scanning day were lost during transfer due to a technical failure. The final sample comprised 35 subjects (mean age=23.06ys, sd=3.02ys, 15 females). All subjects had normal or corrected-to-normal vision.

Paradigm

Faces were presented in four peripheral quadrants of the screen (Fig. 1A) on a grey background with a fixation cross in the center. Each stimulus panel contained four faces of different identity expressing the same emotion. Stimulus duration was 200ms. The stimuli were presented after an inter-stimulus interval of 550ms during which only the fixation cross was present. A change detection task was presented at the central fixation cross. Roving paradigms have frequently been used to study automatic sensory expectation effects (Haenschel et al., 2005; Garrido et al., 2008; Costa-Faidella et al., 2011; Moran et al., 2013; Auksztulewicz and Friston, 2015; Stefanics et al., 2018a,b). Here, we used a factorially structured multi-feature visual 'roving standard' paradigm to elicit PE responses by unexpected changes either in color (red, green), or emotional expression (happy, fearful) of human faces, or both. Importantly, this allowed us to study how brain responses to physically identical stimuli differed, depending on the degree

of expectations about color and emotion, respectively. A diagram of the transitions between stimulus types is shown in Fig. 1B.

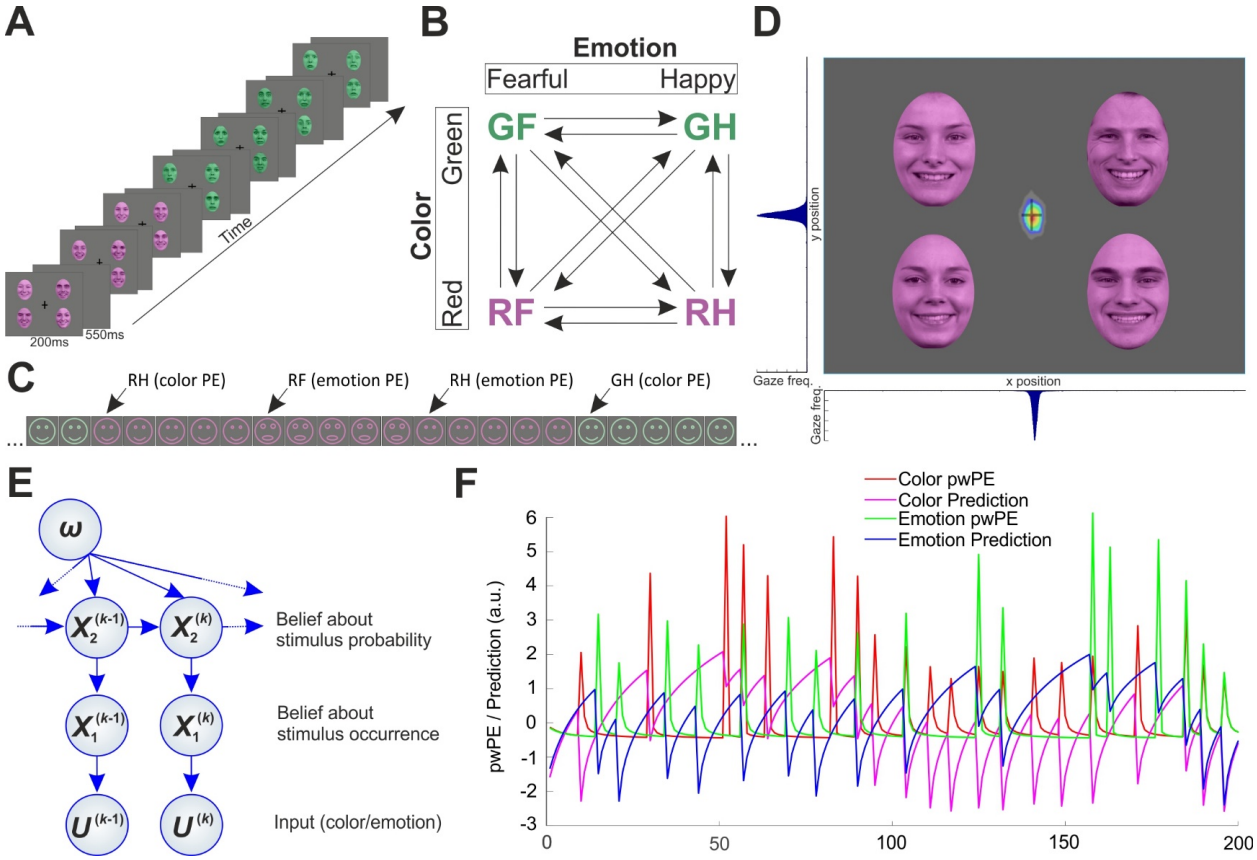

Fig. 1. Experimental design and eye-tracking results. A) Four individual photographs of the same color displaying the same facial affect were presented in each stimulus panel for 200ms in a roving standard paradigm. Each panel was followed by an empty grey screen presented for 550ms. The vertical and horizontal lines of the fixation cross occasionally flipped during this interstimulus interval. The subjects' task was to press a button when the cross flipped. B) Schematic contingency table showing the four equally probable stimulus types (GF: green fearful, GH: green happy, RF: red fearful, RH: red happy faces). After 5-9 presentations each stimulus type was followed by any of the other three types. Arrows indicate transitions with equal overall probability between stimulus types during the experiment. C) Schematic illustration of a stimulus sequence showing transitions between stimulus types. Note physically identical stimuli evoking different PEs depending on expectations established by prior stimulus context. D) Heatmap of normalized gaze position frequency overlaid on a stimulus panel. Warmer colors represent more frequent gaze position. Normalized histograms below and left to the heatmap show the same data projected onto the x and y axis, respectively. Faces were reproduced with permission of the Radboud Faces Database (www.rafd.nl). E) A graphical model of the Hierarchical Gaussian Filter with two levels. F) Model-based pwPE trajectories from one experimental block used as regressors in the GLM.

Images were taken from the Radboud Faces Database (Langner et al., 2010). Ten female and ten male Caucasian models were selected based on their high percentage of agreement on emotion categorization (98% for happy, 92% for fearful faces). A Wilcoxon rank sum test indicated that categorization agreement on the emotional expressions did not differ between happy and fearful faces (Z=-0.63, p=0.53). To control low-level image properties, we equated the luminance and the spatial frequency content of grayscale

images of the selected happy and fearful faces using the SHINE toolbox (Willenbockel et al., 2010). The resulting images were used to create the colored stimuli.

Behavioral task

Similar to previous studies (e.g., Astikainen et al., 2009; Kimura et al., 2012; Müller et al., 2010; Stefanics et al., 2011, 2012, 2018a,b; Kreegipuu et al., 2013; Kuldkepp et al., 2013; Kovács-Bálint et al., 2014; Farkas et al., 2015) we used a behavioral task to engage participants' attention and thus reduce attentional effects on the processing of face stimuli across participants. The task involved detecting changes in the length of the horizontal and vertical lines of a fixation cross presented in the center of the visual field. At random times, the cross became wider or longer (Fig. 1A), at a rate of 8 flips per minute on average. The cross-flips were unrelated to the changes of the unattended faces. The task was to quickly respond to the cross-flips with a right hand button-press. Reaction times were recorded.

Eye-tracking

Participants were explicitly asked to fixate at the cross in the center of the screen. To make sure that participants did not direct their overt attention to the face stimuli, we used an Eyelink 1000 eye-tracking system to record gaze position at 250 Hz during the experiment. After removal of intervals immediately before and after, as well as during blinks, heatmap of x-y data points for all subjects were plotted using the EyeMMV toolbox (Krassanakis et al., 2014). A Gaussian filter (SD=3 pixels) was applied to smooth the final image. The heatmap was normalized to have maximum value of 1, and gaze position histograms for x and y coordinates were plotted (Fig.1D).

Data acquisition and preprocessing

FMRI data was acquired on a Philips Achieva 3 Tesla scanner using an eight channel head-coil (Philips, Best, The Netherlands) at the Laboratory for Social and Neural Systems Research at the University of Zurich. A structural image was acquired for each participant with a T1-weighted MPRAGE sequence: 181 sagittal slices, field of view (FOV): 256 × 256 mm2, Matrix: 256 x 256, resulting in 1 mm^3 resolution. Functional imaging data was acquired in six experimental blocks. In each block 200 whole-brain images were acquired using a T2*-weighted echo-planar imaging sequence with the following parameters. 42 ascending transverse plane slices with continuous in-plane acquisition (slice thickness: 2.5 mm; in-plane resolution: 3.125 x 3.125 mm; inter-slice gap: 0.6 mm; TR = 2.451 ms; TE = 30 ms; flip angle = 77; field of view = 220 x 220 x 130 mm; SENSE factor = 1.5; EPI factor = 51). We used a 2nd order pencil-beam shimming procedure provided by Philips to reduce field inhomogeneities during the functional scans. All functional images were reconstructed with 3 mm isotropic resolution. Functional data acquisition lasted approximately 1 hour. During fMRI data acquisition, respiratory and cardiac activity was recorded using a breathing belt and an electrocardiogram, respectively.

We used statistical parametric mapping (SPM12, v6470; RRID: SCR_007037; Friston et al., 2007) for fMRI data analysis. First, functional images were slice time corrected, realigned to correct for motion and co-registered with the subject's own anatomical image. Next, we normalized structural images to MNI space using the unified segmentation approach and applied the same warping to normalize functional images.

The functional images were smoothed with a 6 mm full-width at half maximum Gaussian kernel and resampled to 2 mm isotropic resolution. We used RETROICOR (Glover et al., 2000) as implemented in the PhysIO-Toolbox (Kasper et al., 2017) from the open source software TAPAS (http://www.translationalneuromodeling.org/tapas) to create confound regressors for cardiac pulsations, respiration, and cardio-respiratory interactions. These confound regressors were entered into the general linear model (GLM; see below). The data and code used in this study are available from the corresponding author, upon reasonable request.

Modeling belief trajectories

In order to include parametric regressors of precision weighted prediction errors (pwPE) in the GLM, we simulated trajectories of belief update in a generative model of perceptual inference, the Hierarchical Gaussian Filter (HGF; Mathys et al., 2011; 2014). We followed the approach described in details in Stefanics et al. (2018a) using the HGF toolbox version v2.2 contained in TAPAS (http://www.translationalneuromodeling.org/tapas). Briefly, we simulated the perceptual model of a two-level HGF for the input traces given by the two features of the face stimuli: color (red vs. green) and emotion (fearful vs. happy). Inversion of the HGF (Fig. 1E) infers the hidden states (x) of the world that generate the sensory input (u). The belief states are updated after each trial following a generic update rule: The posterior mean $\mu_2^{(k)}$ of state x_2 at trial k changes its value according to a precision-weighted PE $\varepsilon_2^{(k)}$, where the precision-weighting changes trial by trial and can be regarded as dynamic learning rate:

$$\mu_2^{(k)} - \mu_2^{(k-1)} \propto \varepsilon_2^{(k)} \tag{1}$$

Note that the sigmoid transform of the tendency $\mu_2^{(k-1)}$ constitutes the prediction (probability of observing an input 1 on trial k), while $\mu_2^{(k)}$ is the tendency after it was updated according to the input on trial k. Here, we refer to $\mu_2^{(k)}$ as prediction. For comparison, classical associative and reinforcement learning models (e.g., Rescorla and Wagner, 1972) follow a similar form but use a fixed learning rate:

$$\text{prediction}^{(k)} = \text{prediction}^{(k-1)} + \text{learning rate} \times \text{PE} \tag{2}$$

For the simulations we assumed that color and emotion were processed by two separate, independent HGFs. However, we considered an interaction between color and emotion PEs within a GLM. Investigating possible interactions at the level of the perceptual model of the HGF would require establishing a novel version of the HGF that incorporates interactions between hidden beliefs, which was beyond the scope of our current study. We estimated the parameters of the model assuming an ideal Bayes-optimal observer (Mathys et al., 2011) that minimizes surprise of the incoming input stream. Figure 1F displays example traces of the absolute value of μ_2 and ε_2 which entered the GLM as described below.

General linear model analysis

The fMRI data was analyzed with two separate GLMs. One GLM included the gradually changing (absolute) pwPEs and "prediction strength" given by the absolute value of μ_2 derived from the HGF as modulatory

221 regressors while the other GLM incorporated a regressor representing categorical stimulus change. The
222 latter served for comparison, implementing a simpler alternative than PC, i.e., change detection (CD; see
223 Lieder et al., 2013). For the GLM based on the CD model, we included stick functions as parametric
224 modulators for each stimulus on those trials when a change occurred in the stimulus sequence. The GLMs
225 were estimated for each participant individually. The pwPE and prediction strength as well as the CD
226 modulatory regressors were computed separately for color and emotion. In addition the GLM included
227 modulatory regressors for red vs. green and happy vs. fearful, respectively. Hence, for each run of the
228 experiment the design matrix included the following experimental regressors: i) a main regressor for the
229 onset of each stimulus display, ii) two modulatory regressors encoding color (red = -1, green = 1) and
230 emotion (happy = -1, fearful = 1), respectively, and iii) two modulatory regressors with the absolute pwPE
231 (or CD) for color and emotion, and iv) two modulatory regressors with the absolute value of the tendency
232 ($|\mu_2|$) for color and emotion (only, in the case of the HGF based model). The modulatory regressors were
233 mean centered and normalized to unit variance. In addition to these regressors of interest, button presses
234 to cross-flips of the visual attention task were also included in the model. All regressors were convolved
235 with a canonical hemodynamic response function (HRF). Movement regressors and physiological
236 confounds were included in the first level GLM (Kasper et al., 2017) which was estimated for each
237 participant individually. Please note that the sign of colors and emotions in ii) was arbitrarily chosen.
238 Finally, in order to assess whether there was any interaction between color and emotion PEs, we fitted an
239 additional GLM, where we included the Hadamard (element-wise) product of the color and emotion pwPE
240 as an additional regressor.

241 On the group level, we used F-tests to find regions whose response showed significant correlation with
242 pwPE or stick regressors. The resulting statistical parametric maps (SPM) were family-wise error (FWE)
243 corrected at the cluster level ($p<0.05$) with a cluster defining threshold (CDT) of $p<0.001$ (Woo et al., 2014;
244 Flandin and Friston, 2017). We used probabilistic anatomical labels and cytoarchitectonic maps in the SPM
245 Anatomy toolbox (v2.2c; RRID: SCR_013273, Eickhoff et al., 2005) to identify the anatomical
246 areas/structures where we observed significant effects. We summarize activations in terms of anatomical
247 labeling by reporting all local maxima within each cluster in Table 1. This provides an overview over the
248 activations in terms of commonly used anatomical labels.

249 **Results**

250 Fixation and behavioral responses

251 Gaze position data (Fig. 1D) confirmed that participants complied with task instructions and fixated the
252 central fixation cross throughout the task. Thus, participants engaged in the detection task and were not
253 overtly attending the faces. Mean reaction time to cross-flips was 484ms (standard deviation:
254 SD=106.9ms), and mean hit rate was 78% (SD=7.34%).

255 First-level GLMs

256 We fitted two GLMs on the single-subject level, incorporating parametric regressors that represented two
257 hypotheses about the decay of pwPE/prediction responses following a change in color of emotional
258 expression of the faces. Similar to the model comparison procedure described in our previous study, our

original aim was to create a functionally defined mask of significant voxels showing PE responses under both models at the group level (Stefanics et al., 2018a). However, while similar activation clusters were obtained using the pwPE/prediction and CD regressors to color changes, significant clusters to changes in emotion were only found using the pwPE/prediction regressors. In other words, the beta estimates obtained using CD were not consistent enough across subjects to yield significant activation clusters at the group level. The lack of significant group-level results for the CD regressors prevented us from creating an unbiased mask comprising significant voxels for color and emotion ("logical AND" conjunction). Furthermore, the additional analysis which included the interaction (product) of color and emotion pwPE did not reveal any evidence for an interaction between the two. We thus restrict ourselves to reporting the results obtained at the group-level analysis using the HGF-based pwPE/prediction model.

Effect of color pwPE

A whole-brain analysis of color changes showed significant activation for color pwPE in fusiform areas (Fig. 2A). Post hoc inspection of the contrast estimates (Fig. 2B) revealed an increased response to pwPE. Predictions pertaining to color did not yield significant activations. Detailed information about anatomical labels, cluster size, and MNI coordinates for the maxima of significant voxel clusters are listed in Table 1.

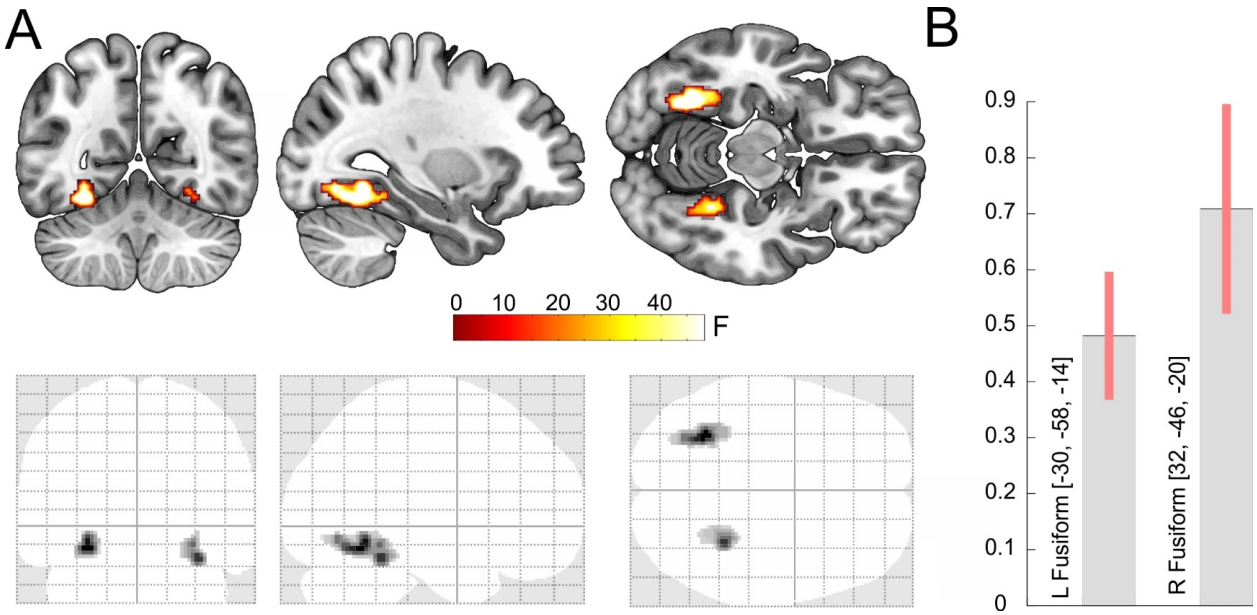

Fig. 2. Effect of color pwPE. A) Top: Activation map (p<0.05 cluster-level whole-brain FWE corrected, with a CDT of p<0.001) overlaid onto the MNI152 standard-space T1-weighted average structural template. Slices show activation in the left fusiform gyrus (MNI-coordinates: [-30 -58 -14]. Bottom: Glass brain showing the results of the F-test. B) Contrast estimates (arbitrary units) for color pwPEs in the left and right fusiform gyrus. Bars indicate 90% C.I.

Effect of emotion prediction and pwPE

A whole-brain analysis of emotion PEs showed significant effects in bilateral cerebellum, cuneus, lingual gyrus, precuneus, thalamus, and right supramarginal gyrus (extending into superior and middle temporal gyrus) as well as right posterior medial frontal cortex (Fig. 3A).

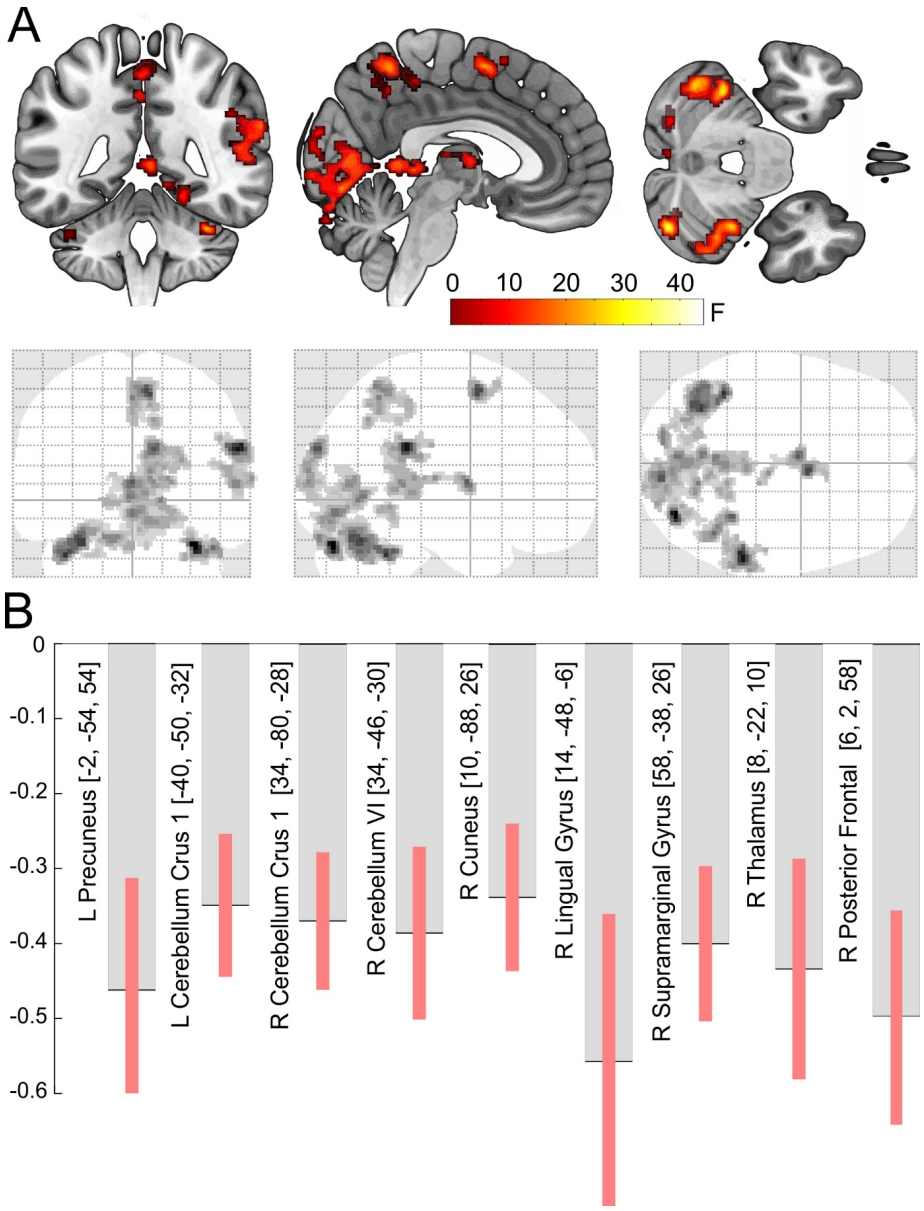

Fig. 3. Main effect of emotion pwPE. A) Top: Activation map (p<0.05 cluster-level whole-brain FWE corrected, with a CDT of p<0.001) overlaid onto the MNI152 standard-space T1-weighted average structural template. Slices show activations at coordinates [4, -45, -31] cutting through the right anterior precuneus. Bottom: Glass brain showing the results of the F-test (whole-brain FWE cluster-level corrected at p<0.05, with a cluster-defining threshold of p<0.001). B) Contrast estimates (arbitrary units) for the emotion pwPEs in the left and right cerebellum, left precuneus, right cuneus, lingual and supramarginal gyrus, thalamus, and posterior frontal cortex. Bars indicate 90% C.I. Note that bar plots are shown for illustration only. Statistical significance was assessed at the whole-brain level described above.

We found significant activations pertaining to emotion predictions in three cortical clusters: two in bilateral fusiform gyri (extending into lingual gyri) and one in right the superior and middle temporal cortex (Fig. 4A). A post hoc analysis of the contrast estimates in these regions revealed that all areas showed a negative effect of emotion pwPEs (Fig. 3B) and predictions (Fig. 4B).

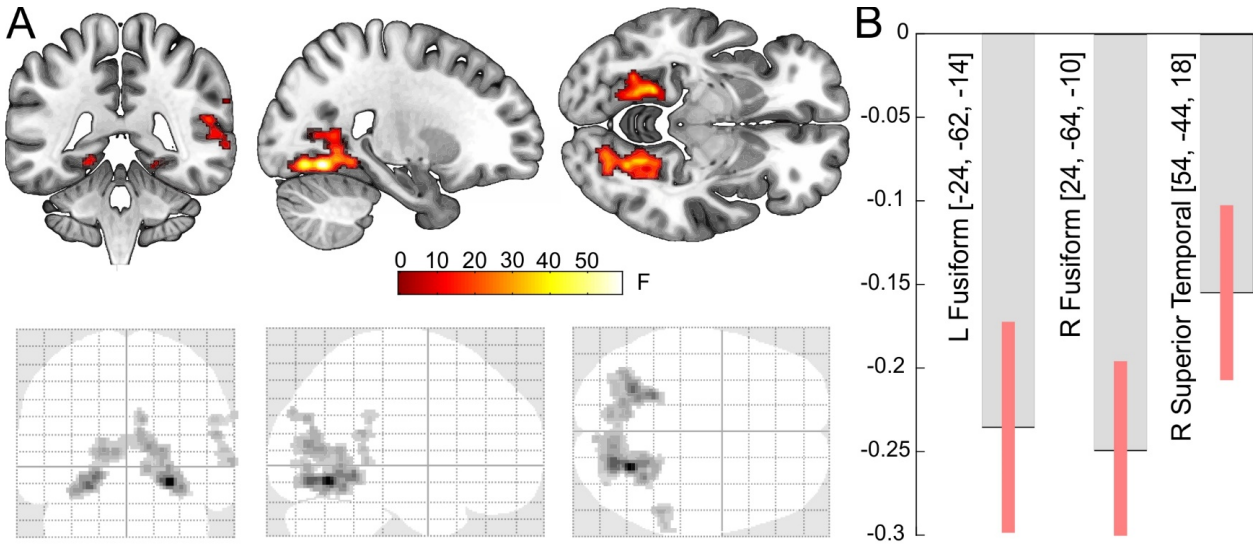

Fig. 4. Main effect of emotion prediction. A) Top: Activation map (p<0.05 cluster-level whole-brain FWE corrected, with a CDT of p<0.001) overlaid onto the MNI152 standard-space T1-weighted average structural template. Slices show activations at coordinates [24, -42, -8] cutting through the right fusiform gyrus. Bottom: Glass brain showing the results of the F-test (whole-brain FWE cluster-level corrected at p<0.05, with a cluster-defining threshold of p<0.001). B) Contrast estimates (arbitrary units) for the emotion prediction in the left and right fusiform gyrus, and right superior temporal gyrus. Bars indicate 90% C.I. Note that bar plots are shown for illustration only. Statistical significance was assessed at the whole-brain level described above.

Discussion

We used the Hierarchical Gaussian Filter, a computational model for learning and inference, to simulate belief trajectories of an ideal Bayesian observer presented with a sequence of face stimuli. The trial by trial update of internal hidden belief states in the HGF relies on precision weighted prediction errors. Traces of predictions and pwPEs pertaining to color and emotional expression of faces served as regressors in a GLM which yielded brain structures where activation showed a significant relationship to those computational quantities. We manipulated sensory expectations towards color and emotional expression of faces independently. Crucially, emotion and color pwPEs/predictions were evoked by physically identical stimuli; only the specific expectation (statistical regularity) that was violated on any given trial, differed between the two conditions. While our previous EEG study reported the scalp distribution and time-course of pwPE responses (Stefanics et al., 2018a), here we used fMRI to find BOLD correlates of pwPEs and predictions in generator structures. We found BOLD correlates of pwPEs to color changes in bilateral fusiform gyrus, whereas pwPEs to changes of emotional expressions activated a different set of areas including the bilateral cerebellum, lingual gyrus, precuneus, thalamus, and right supramarginal gyrus as well as right posterior medial frontal cortex. We observed activations pertaining to emotion predictions in bilateral fusiform and the right supramarginal gyrus (Fig. 5A).

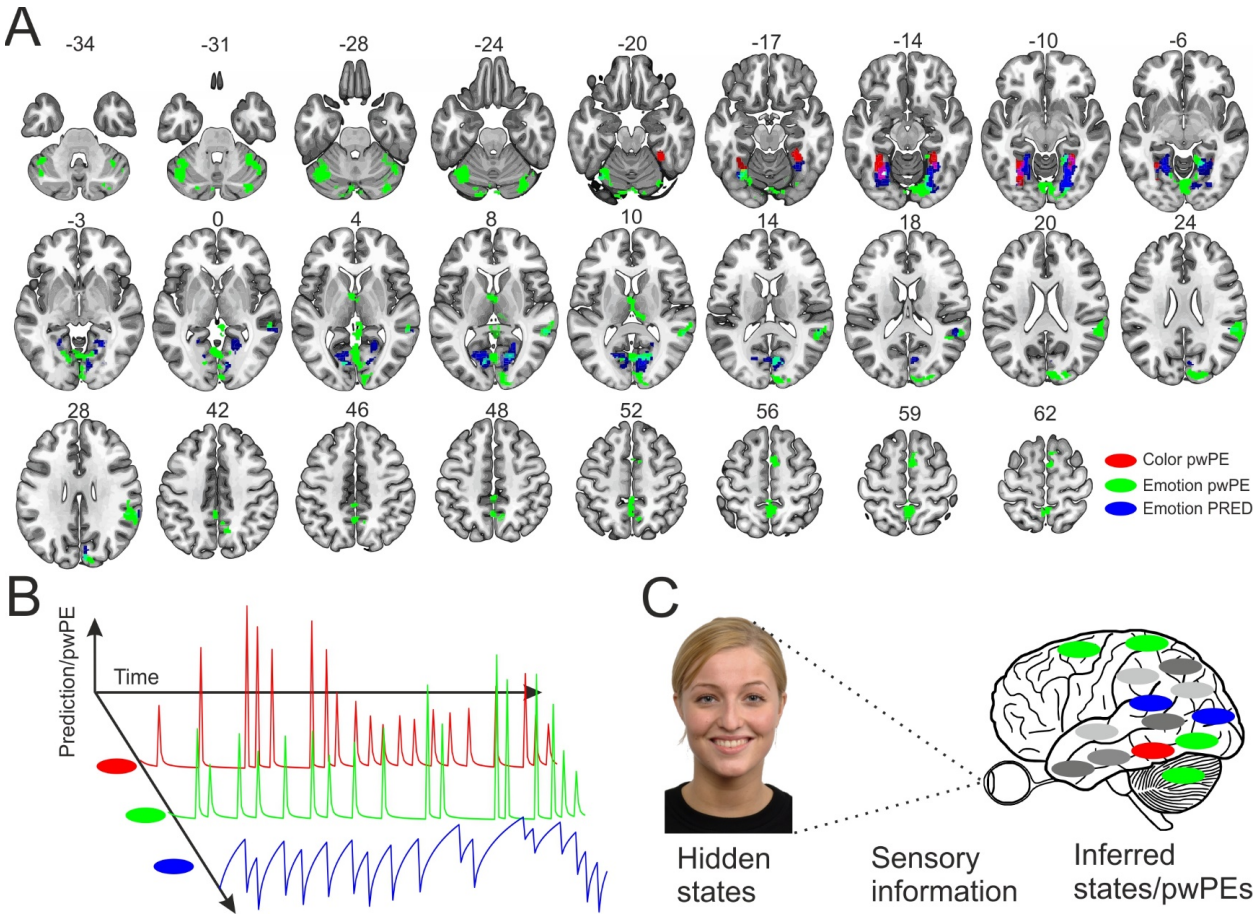

Fig. 5. Overview of the results and PC framework for perceptual prediction errors. A) Colored areas mark main clusters related to color pwPEs (red), and emotion pwPEs (green). Note the dissociation of PEs for color and emotion changes. B) pwPE- and prediction-related activations for different sensory features arise and are updated, respectively, during Bayesian inference as properties of the hidden states that cause the sensory information dynamically change over time. Prediction and pwPEs to color and emotion are marked as in A), additional features are marked with grey. C) Schematic depicting functional segregation in the nervous system, as distinct features of the world are inferred and predicted by distinct neural structures specializing in the given features. Image of a model used in our study reproduced with permission of the Radboud Faces Database (Langner et al., 2010).

According to recent hierarchical formulations of PC (Friston, 2005), creating and maintaining our internal model of the world is a process during which predictive object representations about the likely properties of the hidden objects are updated using precision-weighted PEs (e.g., Moran et al., 2013; Stefanics et al., 2018a) that signals mismatch between the expectations based on prior information and the current sensory data (Fig. 5B). In the present study, the demonstration of activations correlated to pwPE in ventral visual areas as well as in emotion processing structures suggests a role for PC in color and emotion perception. Importantly, we manipulated stimulus sequences to induce automatic expectations about the occurrence of different stimulus features, using the same faces to elicit distinct emotion and color pwPEs. In line with our hypothesis, color and emotion pwPEs were reflected by activity in brain structures known to be dedicated to color and emotion processing. A hypothetical generalization of our results is shown in

342 Fig. 5C, which illustrates functional segregation of inferring hidden causes of sensory information for
343 different features, including color and emotional expression of faces.

344 Here, we studied predictions and pwPEs to unattended and task-irrelevant stimuli. We used a primary
345 task independent of the facial stimuli to ensure that participants did not attend to the faces and verified
346 their attentional focus by eye-tracking. Thus, predictions and pwPEs were elicited under an automatic
347 recognition processes and minimized confounding variations in attentional contributions.

348 It is important to note that due to the lack of significant group level results for the emotion stick regressor,
349 we were not able to directly compare models using the approach presented in Stefanics et al. (2018a).
350 Hence, we could not use model comparison to assess whether the pwPE or the stick regressor traces
351 provided the formally better model. Notably, the latter are equivalent to a very quick adaptation without
352 precision weighting. However, the second level results suggest that the representation of pwPEs is more
353 consistent across subjects, leading to a significant group effect. In addition, while we use computational
354 quantities to model neural activity in the GLM, our method (fMRI) does not allow us to make a direct
355 statement about the neuronal implementation, e.g., neuronal fatigue, suppressive effects in single
356 neurons, or network effects (e.g., Solomon and Kohn; 2014, Stefanics et al., 2016). Based on the current
357 analysis it is not possible to reject some form of adaptation (e.g., fatigue) as a potential mechanism as
358 opposed to a more general model based on hierarchical Bayesian inference. Thus, adaptation could be an
359 alternative explanation of our findings.

360 To our knowledge this is the first fMRI study using a Bayesian observer model to describe automatic
361 predictions and pwPEs to violations of expectations to different features of the same objects, in the
362 absence of focal attention and task-relevance. Both expectation based on stimulus probability and
363 attention based on task-relevance have been suggested to modulate sensory PEs (e.g., Summerfield and
364 de Lange, 2014; Auksztulewicz and Friston, 2015; Auksztulewicz et al., 2017). Attentional effects have
365 been suggested to increase synaptic gain of PE coding neurons (Kok et al., 2012; Wyart et al., 2012; Jiang
366 et al., 2013; Vossel et al., 2014; Auksztulewicz and Friston, 2015), whereas expectation effects manifest in
367 reduced neuronal responses (Grotheer and Kovács, 2015; Auksztulewicz and Friston, 2016; Stefanics et
368 al., 2018a,b). Recent formulations of PC suggest that attention serves to optimize precision estimates of
369 specific PEs. By increasing the weight that is put on PEs, the role of attention is to influence subsequent
370 inference and learning (Friston, 2009; Feldman and Friston, 2010; den Ouden et al., 2012; Parr and Friston,
371 2018). Furthermore, a previous study also found that PEs spread across object features in the visual cortex
372 (Jiang et al., 2016). Here, we extend these previous findings by showing that (i) pwPEs can also be elicited
373 in spatially remote neural structures that specialize in the processing of distinct stimulus attributes and
374 (ii) in the absence of attention. Notably, Jiang et al. (2016) studied PEs to attended and task-relevant
375 random dot stimuli, while in our study face stimuli were task-irrelevant and not attended, as verified by
376 eye tracking. The differences between our current and their results suggest that the role of focal attention
377 in perception might not only be to enhance but also spread PEs across features at the object level (Jiang
378 et al., 2016) which is in line with the feature-integration theory of attention (Treisman and Gelade, 1980).
379 Thus, while the visual system likely represents statistical relationships across features and automatically
380 structures them into objects (Müller et al., 2009, 2011), our results suggest that PEs to violations of specific
381 features are processed mostly in different regions. Clusters in the cerebellum, thalamus, precuneus,

posterior medial frontal cortex, and right temporal areas were activated exclusively for predictions and/or pwPEs pertaining to emotions. However, activations in the fusiform gyrus for color and emotion showed some overlap (Fig. 5A). In addition, we could not find any evidence for an interaction between PEs for different features when they are task-irrelevant and unattended. However, we only considered an interaction between color and emotion PEs at the level of the GLM and did not investigate possible interactions at the level of the perceptual model of the HGF. This would require establishing a novel HGF that incorporates interactions between hidden beliefs.

Color PEs

Color processing involves the ventral visual pathway (Mesulam, 1998; Bartels and Zeki, 2000), where fMRI studies have shown strong color-related activations (Brewer et al., 2005; Solomon and Lennie, 2007; Barbur and Spang, 2008; Brouwer and Heeger, 2009). The location of the fusiform activation in our experiment is in agreement with "color-biased" regions in the ventral occipito-temporal cortex (Lafer-Sousa et al., 2016). To our knowledge, there have been no previous investigations of color processing from a PC-related perspective. Our results suggest the importance of pwPEs, as a putative signature of PC, for color perception.

Emotion PEs and predictions

Facial emotions are non-verbal acts of communication that express emotional states and intentions, and are fundamental in social interactions (Fridlund, 1994; Frith, 2009). The social environment is not constant, and detecting changes in the emotional valence of facial expressions in our social space is important for socially successful behavior. Prior ERP studies (Susac et al., 2004; Kimura et al., 2012; Li et al., 2012; Csukly et al., 2013; Stefanics et al., 2012, 2018a; Astikainen et al., 2013; Fujimura and Okanoya, 2013; Xu et al., 2018) suggest that emotional expressions are processed in a few hundred milliseconds and stored in predictive memory representations. We found emotion pwPEs in a set of areas including the bilateral cerebellum, precuneus, thalamus, right lingual and supramarginal gyrus, as well as right posterior medial frontal cortex. We observed activations pertaining to emotion predictions in bilateral fusiform and the right superior temporal gyrus. Details of significant clusters are provided in Table 1. This pattern of results (Fig. 5A) is in line with the notion that emotion processing involves a mosaic-like set of affective, motor-related and sensory components (Bastiaansen et al., 2009). More specifically, it demonstrates pwPE/prediction activations in areas that previous work identified as activated by the processing of emotional faces (Fusar-Poli et al., 2009; E et al., 2014; Adamaszek et al., 2017) and theory of mind tasks, in particular the Mind in the Eyes task (Schurz et al., 2014).

In our current study we observed positive and negative betas for color and emotion pwPEs, respectively, which might reflect complementary neural mechanisms for predictive processing across distinct features. The notion that predictive coding across features can be mediated by qualitatively different mechanisms (Auksztulewicz et al., 2018) suggests domain-specific predictive signaling. As fMRI does not allow to measure detailed neural firing but rather represents the bulk signal of excitation and inhibition within a region (Logothetis, 2008), we cannot draw conclusions about specific mechanisms that could lead to this difference in the PE signal.

In summary, our findings demonstrate that the same physical stimulus can elicits separate feature-specific pwPE/prediction responses, depending on distinct predictions about its various attributes. This is in agreement with PC theories of perception. In future extensions of this work, models of effective connectivity could examine the signaling of pwPEs/predictions in cortical networks as postulated by PC.

References

Adamaszek M, D'Agata F, Ferrucci R, Habas C, Keulen S, Kirkby KC, Leggio M, Mariën P, Molinari M, Moulton E, Orsi L, Van Overwalle F, Papadelis C, Priori A, Sacchetti B, Schutter DJ, Styliadis C, Verhoeven J (2017) Consensus Paper: Cerebellum and Emotion. Cerebellum 16:552-576. https://doi.org/10.1007/s12311-016-0815-8

Adams RA, Napier G, Roiser JP, Mathys C, Gilleen J (2018) Attractor-like dynamics in belief updating in schizophrenia. J Neurosci doi: 10.1523/JNEUROSCI.3163-17.2018.

Astikainen P, Hietanen JK (2009) Event-related potentials to task-irrelevant changes in facial expressions. Behav Brain Funct 5:30.

Astikainen P, Cong F, Ristaniemi T, Hietanen JK (2013) Event-related potentials to unattended changes in facial expressions: detection of regularity violations or encoding of emotions? Front Hum Neurosci 7:557.

Auksztulewicz R, Friston K (2015) Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. Cereb Cortex 25:4273-4283.

Auksztulewicz R, Friston K (2016) Repetition suppression and its contextual determinants in predictive coding. Cortex 80:125-40. doi: 10.1016/j.cortex.2015.11.024.

Auksztulewicz R, Friston KJ, Nobre AC (2017) Task relevance modulates the behavioural and neural effects of sensory predictions. PLoS Biol 15(12): e2003143. doi.org/10.1371/journal.pbio.2003143

Auksztulewicz R, Schwiedrzik CM, Thesen T, Doyle W, Devinsky O, Nobre AC, Schroeder CE, Friston KJ, Melloni L (2018) Not All Predictions Are Equal: "What" and "When" Predictions Modulate Activity in Auditory Cortex through Different Mechanisms. J Neurosci. 38(40):8680-8693. doi: 10.1523/JNEUROSCI.0369-18.2018.

Barbur JL, Spang K (2008) Colour constancy and conscious perception of changes of illuminant. Neuropsychologia 46:853-863.

Bartels A, Zeki S (2000) The architecture of the colour centre in the human visual brain: new results and a review. Eur J Neurosci 12:172-193.

Bastiaansen JA, Thioux M, Keysers C (2009) Evidence for mirror systems in emotions. Philos Trans R Soc Lond B Biol Sci 364:2391-2404.

Bogacz R (2017) A tutorial on the free-energy framework for modelling perception and learning. J Math Psychol. 76(Pt B):198-211.

Brewer AA, Liu J, Wade AR, Wandell BA (2005) Visual field maps and stimulus selectivity in human ventral occipital cortex. Nat Neurosci 8:1102-1109.

Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. J Neurosci 29:13992-14003.

Clark A (2015) Surfing Uncertainty: Prediction, Action, and the Embodied Mind. In. Oxford: Oxford University Press.

Costa-Faidella J, Baldeweg T, Grimm S, Escera C (2011) Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. J Neurosci 31:18590-18597.

Csukly G, Stefanics G, Komlósi S, Czigler I, Bitter I, Czobor P (2013) Emotion-related visual mismatch responses in schizophrenia: Impairments and correlations with emotion recognition. PLoS ONE, 8(10): e75444.

den Ouden HE, Kok P, de Lange FP (2012) How prediction errors shape perception, attention, and motivation. Front Psychol. 3:548. doi: 10.3389/fpsyg.2012.00548.

Diaconescu AO, Mathys C, Weber LAE, Kasper L, Mauer J, Stephan KE (2017) Hierarchical prediction errors in midbrain and septum during social learning. Soc Cogn Affect Neurosci 12(4):618-634. doi: 10.1093/scan/nsw171.

Durschmid S, Edwards E, Reichert C, Dewar C, Hinrichs H, Heinze HJ, Kirsch HE, Dalal SS, Deouell LY, Knight RT (2016) Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc Natl Acad Sci U S A 113:6755-6760.

E KH, Chen SH, Ho MH, Desmond JE (2014) A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. Hum Brain Mapp 35:593-615.

Egner T, Monti JM, Summerfield C (2010) Expectation and surprise determine neural population responses in the ventral visual stream. J Neurosci 30:16601-16608.

Ehinger BV, Hausser K, Ossandon JP, Konig P (2017) Humans treat unreliable filled-in percepts as more real than veridical ones. Elife 6:ARTN e21761.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25:1325-1335.

Farkas K, Stefanics G, Marosi C, Csukly G (2015) Elementary sensory deficits in schizophrenia indexed by impaired visual mismatch negativity. Schizophrenia Research 166: 164-170.

Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci. 4:215. doi: 10.3389/fnhum.2010.00215.

Flandin G, Friston KJ (2017) Analysis of family-wise error rates in statistical parametric mapping using random field theory. Hum Brain Mapp doi: 10.1002/hbm.23839

Fridlund AJ (1994) Human facial expression: An evolutionary view. San Diego, CA: Academic Press.

Friston K (2002) Functional integration and inference in the brain. Progress in Neurobiology 68:113–143.
Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815-836.
Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn. Sci. 13(7):293-301. doi: 10.1016/j.tics.2009.04.005.
Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127-138.
Friston KJ, Ashburner J, Kiebel SJ, Nichols TE, Penny WD, eds. (2007) Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press.
Friston K, Adams R, Montague R (2012a) What is value-accumulated reward or evidence? Front Neurorobot 6:11.
Friston K, Adams RA, Perrinet L, Breakspear M (2012b) Perceptions as hypotheses: saccades as experiments. Front Psychol 3:151.
Frith C (2009) Role of facial expressions in social interactions. Philos Trans R Soc Lond B Biol Sci 364:3453-3458.
Fujimura T, Okanoya K (2013) Event-related potentials elicited by pre-attentive emotional changes in temporal context. PLoS One 8:e63703.
Fusar-Poli P, Placentino A, Carletti F, Landi P, Allen P, Surguladze S, Benedetti F, Abbamonte M, Gasparotti R, Barale F, Perez J, McGuire P, Politi P (2009) Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. J Psychiatry Neurosci 34:418-432.
Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM (2008) The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42:936-944.
Glover GH, Li TQ, Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med 44:162-167.
Gordon N, Koenig-Robert R, Tsuchiya N, van Boxtel JJ, Hohwy J (2017) Neural markers of predictive coding under perceptual uncertainty revealed with Hierarchical Frequency Tagging. Elife 6.
Grotheer M, Kovács G (2015) The relationship between stimulus repetitions and fulfilled expectations. Neuropsychologia. 67:175-82. doi: 10.1016/j.neuropsychologia.2014.12.017.
Haenschel C, Vernon DJ, Dwivedi P, Gruzelier JH, Baldeweg T (2005) Event-related brain potential correlates of human auditory sensory memory-trace formation. J Neurosci 25:10494-10501.
Hubel DH, Wiesel TN (1965) Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. J Neurophysiol 28:229-289.
Jiang J, Summerfield C, Egner T (2013) Attention sharpens the distinction between expected and unexpected percepts in the visual brain. J Neurosci. 33(47):18438-18447. doi: 10.1523/JNEUROSCI.3308-13.2013.
Jiang J, Summerfield C, Egner T (2016) Visual Prediction Error Spreads Across Object Features in Human Visual Cortex. J Neurosci 36:12746-12763.
Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinzle J, Iglesias S, Hauser TU, Sebold M, Manjaly ZM, Pruessmann KP, Stephan KE (2017) The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. J Neurosci Methods 276:56-72.
Katthagen T, Mathys C, Deserno L, Walter H, Kathmann N, Heinz A, Schlagenhauf F (2018) Modeling subjective relevance in schizophrenia and its relation to aberrant salience. PLoS Comput Biol 14(8): e1006319. https://doi.org/10.1371/journal.pcbi.1006319
Kimura M, Kondo H, Ohira H, Schroger E (2012) Unintentional temporal context-based prediction of emotional faces: an electrophysiological study. Cereb Cortex 22:1774-1785.
Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP (2012) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197-2206.
Krassanakis V, Filippakopoulou V, Nakos B (2014) EyeMMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. J Eye Movement Res 7.
Kreegipuu K, Kuldkepp N, Sibolt O, Toom M, Allik J, Näätänen R (2013) vMMN for schematic faces: automatic detection of change in emotional expression. Front Hum Neurosci 7:714.
Kremlacek J, Kreegipuu K, Tales A, Astikainen P, Poldver N, Naatanen R, Stefanics G (2016) Visual mismatch negativity (vMMN): A review and meta-analysis of studies in psychiatric and neurological disorders. Cortex 80:76-112.
Kuldkepp N, Kreegipuu K, Raidvee A, Näätänen R, Allik J (2013) Unattended and attended visual change detection of motion as indexed by event-related potentials and its behavioral correlates. Front Hum Neurosci 7:476. doi: 10.3389/fnhum.2013.00476.
Lafer-Sousa R, Conway BR, Kanwisher NG (2016) Color-Biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans, as in Macaques. J Neurosci 36:1682-1697.
Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A (2010) Presentation and validation of the Radboud Faces Database. Cognition & Emotion 24:1377-1388.
Lawson RP, Mathys C, Rees G (2017) Adults with autism overestimate the volatility of the sensory environment. Nat Neurosci 20:1293-1299.
Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am 20.

Li X, Lu Y, Sun G, Gao L, Zhao L (2012) Visual mismatch negativity elicited by facial expressions: new evidence from the equiprobable paradigm. Behav Brain Funct 8:7.
Lochmann T, Ernst UA, Deneve S (2012) Perceptual inference predicts contextual modulations of sensory responses. J Neurosci 32:4179-4195.
Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453(7197):869-878. doi: 10.1038/nature06976.
Mesulam MM (1998) From sensation to cognition. Brain 121 (Pt 6):1013-1052.
Moran RJ, Campo P, Symmonds M, Stephan KE, Dolan RJ, Friston KJ (2013) Free energy, precision and learning: the role of cholinergic neuromodulation. J Neurosci 33:8227-8236.
Müller D, Widmann A, Schröger E (2013) Object-related regularities are processed automatically: evidence from the visual mismatch negativity. Front Hum Neurosci. 7:259. doi:10.3389/fnhum.2013.00259
Müller D, Winkler I, Roeber U, Schaffer S, Czigler I, Schröger E (2010) Visual object representations can be formed outside the focus of voluntary attention: evidence from event-related brain potentials. J Cogn Neurosci 22:1179-1188.
Parr T, Friston KJ (2018) The Anatomy of Inference: Generative Models and Brain Structure. Front Comput Neurosci. 12:90. doi: 10.3389/fncom.2018.00090.
Powers AR, Mathys C, Corlett PR (2017) Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. Science 357:596-600.
Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79-87.
Riesenhuber M, Poggio T (2000) Models of object recognition. Nat Neurosci 3 Suppl:1199-1204.
Schurz M, Radua J, Aichhorn M, Richlan F, Perner J (2014) Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci Biobehav Rev 42:9-34.
Schwartenbeck P, FitzGerald TH, Mathys C, Dolan R, Friston K (2015) The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. Cereb Cortex 25:3434-3445.
Schwiedrzik CM, Freiwald WA (2017) High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. Neuron 96:89-97 e84.
Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD (2016) Neural signatures of perceptual inference. Elife 5:e11476.
Smith FW, Muckli L (2010) Nonstimulated early visual areas carry information about surrounding context. P Natl Acad Sci USA 107:20099-20103.
Solomon SG, Kohn A (2014) Moving sensory adaptation beyond suppressive effects in single neurons. Curr Biol. 24(20):R1012-22. doi: 10.1016/j.cub.2014.09.001.
Solomon SG, Lennie P (2007) The machinery of colour vision. Nat Rev Neurosci 8:276-286.
Stefanics G, Kimura M, Czigler I (2011) Visual mismatch negativity reveals automatic detection of sequential regularity violation. Front Hum Neurosci 5:46.
Stefanics G, Kremlacek J, Czigler I (2014) Visual mismatch negativity: a predictive coding view. Front Hum Neurosci 8:666.
Stefanics G, Kremlacek J, Czigler I (2016) Mismatch negativity and neural adaptation: Two sides of the same coin. Response: Commentary: Visual mismatch negativity: a predictive coding view. Front Hum Neurosci 10:13. doi: 10.3389/fnhum.2016.00013
Stefanics G, Heinzle J, Horvath AA, Stephan KE (2018a) Visual Mismatch and Predictive Coding: A Computational Single-Trial ERP Study. J Neurosci 38:4020-4030.
Stefanics G, Heinzle J, Czigler I, Valentini E, Stephan KE (2018b) Timing of repetition suppression of event-related potentials to unattended objects. Eur J Neurosci doi:10.1111/ejn.13972
Stefanics G, Csukly G, Komlosi S, Czobor P, Czigler I (2012) Processing of unattended facial emotions: a visual mismatch negativity study. Neuroimage 59:3042-3049.
Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci. 15(11):745-56. doi: 10.1038/nrn3838.
Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97-136.
Susac A, Ilmoniemi RJ, Pihko E, Supek S (2004) Neurodynamic studies on emotional and inverted faces in an oddball paradigm. Brain Topogr 16:265-268.
Vossel S, Mathys C, Daunizeau J, Bauer M, Driver J, Friston KJ, Stephan KE (2014) Spatial Attention, Precision, and Bayesian Inference: A Study of Saccadic Response Speed. Cereb Cortex. 24(6):1436-50. doi: 10.1093/cercor/bhs418.
Vossel S, Mathys C, Stephan KE, Friston KJ (2015) Cortical Coupling Reflects Bayesian Belief Updating in the Deployment of Spatial Attention. J Neurosci 35:11532-11542.
Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proceedings of the National Academy of Sciences of the United States of America 108:20754-20759.

595 Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: the SHINE
596 toolbox. Behav Res Methods 42:671-684.
597 Woo CW, Krishnan A, Wager TD (2014) Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations.
598 Neuroimage 91:412-419.
599 Wyart V, Nobre AC, Summerfield C (2012) Dissociable prior influences of signal probability and relevance on visual contrast
600 sensitivity. Proc Natl Acad Sci USA. 109(9):3593-8. doi: 10.1073/pnas.1120118109.
601 Xu QR, Ruohonen EM, Ye CX, Li XQ, Kreegipuu K, Stefanics G, Luo WB, Astikainen P (2018) Automatic Processing of Changes in
602 Facial Emotions in Dysphoria: A Magnetoencephalography Study. Frontiers in Human Neuroscience 12:186.
603

604 **Table 1.**

Contrast and Cluster	Structure	Cytoarchitectonic area	Cluster max. (MNI)		
Main effects of color PEs					
Cluster 1 (326 voxels)	**L Fusiform Gyrus**	**Area FG3**	**-30**	**-58**	**-14**
	L Fusiform Gyrus	Area FG1	-28	-70	-8
	L Fusiform Gyrus	Area FG3	-30	-54	-10
Cluster 2 (212 voxels)	**R Fusiform Gyrus**	**Area FG3**	**32**	**-46**	**-20**
	R Fusiform Gyrus	Area FG3	30	-46	-12
	R Fusiform Gyrus	n.a.	26	-54	-12
Main effects of emotion PEs					
Cluster 1 (1417 voxels)	**R Cuneus**	**Area hOc3d [V3d]**	**10**	**-88**	**26**
	L Lingual Gyrus	Area hOc1 [V1]	-2	-78	-8
	Cerebellar Vermis (4/5)	n.a.	-2	-64	0
	R Lingual Gyrus	Area hOc3v [V3v]	14	-66	-2
	L Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	-12	-88	-22
	R Calcarine Gyrus	Area hOc2 [V2]	10	-94	8
	n.a.	Area hOc1 [V1]	6	-82	-14
	R Cuneus	Area hOc2 [V2]	8	-94	14
	L Lingual Gyrus	Area hOc1 [V1]	2	-72	6
	R Calcarine Gyrus	Area hOc1 [V1]	18	-68	10
	R Lingual Gyrus	Area hOc3v [V3v]	18	-86	-14
Cluster 2 (554 voxels)	**L Cerebellum (Crus 1)**	**Lobule VIIa crusI (Hem)**	**-40**	**-50**	**-32**
	L Cerebellum (VI)	Lobule VI (Hem)	-30	-64	-24
	L Cerebellum (Crus 1)	Lobule VIIa crusI (Hem)	-42	-66	-30
	L Cerebellum (VI)	Lobule VI (Hem)	-36	-54	-28
	L Cerebellum (VI)	Lobule VI (Hem)	-24	-74	-20
	L Cerebellum (Crus 1)	Lobule VIIa crusI (Hem)	-20	-80	-24
	L Cerebellum (Crus 1)	Lobule VIIa crusI (Hem)	-22	-80	-32
	L Cerebellum (Crus 1)	Lobule VIIa crusI (Hem)	-26	-80	-32
Cluster 3 (511 voxels)	**R SupraMarginal Gyrus**	**Area PF (IPL)**	**58**	**-38**	**26**
	R Middle Temporal Gyrus	n.a.	56	-40	6
	R Middle Temporal Gyrus	n.a.	58	-42	8
	R SupraMarginal Gyrus	Area PFm (IPL)	64	-44	26
	R Superior Temporal Gyrus	Area PF (IPL)	66	-34	10
	R Superior Temporal Gyrus	Area PFm (IPL)	60	-42	20
	R Middle Temporal Gyrus	n.a.	52	-48	16
	R Superior Temporal Gyrus	n.a.	54	-44	14
	R SupraMarginal Gyrus	Area PFt (IPL)	54	-24	28
	R SupraMarginal Gyrus	n.a.	48	-42	32
	R Superior Temporal Gyrus	n.a.	62	-38	12
Cluster 4 (380 voxels)	**L Precuneus**	**n.a.**	**-2**	**-54**	**54**
	L Precuneus	n.a.	0	-56	60
	R Precuneus	Area 5M (SPL)	2	-50	58
	L Precuneus	n.a.	-2	-54	48
	R Precuneus	n.a.	10	-60	42
	R Precuneus	n.a.	8	-58	50
	L Midcingulate cortex	Area 5M (SPL)	0	-38	52
	L Midcingulate cortex	n.a.	-2	-38	44
	L Midcingulate cortex	n.a.	-2	-44	42
Cluster 5 (178 voxels)	**R Cerebellum (VI)**	**Lobule VI (Hem)**	**34**	**-46**	**-30**
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	44	-56	-28
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	40	-54	-30
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	38	-52	-32
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	48	-60	-32
	R Cerebellum (Crus 1)*	Area FG2	46	-62	-26
Cluster 6 (162 voxels)	**R Cerebellum (Crus 1)**	**Lobule VIIa crusI (Hem)**	**34**	**-80**	**-28**
	R Cerebellum (Crus 1)	Lobule VIIa crusI (Hem)	28	-76	-34
	R Cerebellum (VI)	Lobule VI (Hem)	32	-72	-24
	R Cerebellum (VI)	Lobule VI (Hem)	36	-64	-26
	R Cerebellum (Crus 1)	n.a.	40	-76	-22
Cluster 7 (130 voxels)	L Thalamus***	Thalamus proper	-2	-4	8

	R Thalamus	**Thal: Temporal**	**8**	**-22**	**10**
	R Thalamus	Thal: Temporal	14	-28	10
Cluster 8 (120 voxels)	**R Posterior-Medial Frontal****	**Supplementary motor cortex**	**6**	**2**	**58**
	R Posterior-Medial Frontal**	Supplementary motor cortex	8	12	58
	R Posterior-Medial Frontal	n.a.	10	14	62
Cluster 9 (107 voxels)	n.a.	n.a.	4	-36	2
	Cerebellar Vermis (4/5)	n.a.	4	-46	4
	n.a.	n.a.	-2	-36	8
Cluster 10 (79 voxels)	**R Lingual Gyrus**	**n.a.**	**14**	**-48**	**-6**
	R Lingual Gyrus	n.a.	18	-52	-6
	R Lingual Gyrus	n.a.	22	-48	-8
	R Fusiform Gyrus	n.a.	22	-46	-12
Main effects of emotion PREDICTIONS					
Cluster 1 (1500 voxels)	**R Fusiform Gyrus**	**Area hOc4v [V4(v)]**	**24**	**-64**	**-10**
	R Lingual Gyrus	Area hOc3v [V3v]	22	-74	-10
	R Lingual Gyrus	n.a.	22	-56	-8
	R Fusiform Gyrus	Area FG3	32	-54	-14
	R Calcarine Gyrus	Area hOc1 [V1]	12	-76	10
	R Lingual Gyrus	Area hOc1 [V1]	24	-56	-2
	R Calcarine Gyrus	Area hOc1 [V1]	20	-72	8
	L Calcarine Gyrus	Area hOc2 [V2]	-16	-68	8
	R Lingual Gyrus	n.a.	24	-46	-10
	R Lingual Gyrus	Area hOc2 [V2]	10	-78	-2
	L Calcarine Gyrus	Area hOc1 [V1]	-10	-66	8
Cluster 2 (515 voxels)	**L Fusiform Gyrus**	**Area FG1**	**-24**	**-62**	**-14**
	L Lingual Gyrus	n.a.	-22	-48	-8
	L Lingual Gyrus	n.a.	-22	-52	-8
	L Lingual Gyrus	n.a.	-20	-58	-8
	L Fusiform Gyrus	Area FG1	-28	-72	-12
	L Fusiform Gyrus	Area hOc4v [V4(v)]	-22	-72	-14
	L Fusiform Gyrus	Area hOc4v [V4(v)]	-30	-76	-12
	L Lingual Gyrus	Area hOc4v [V4(v)]	-18	-70	-10
Cluster 3 (147 voxels)	**R Superior Temporal Gyrus**	**n.a.**	**54**	**-44**	**18**
	R Middle Temporal Gyrus	n.a.	62	-40	2
	R Middle Temporal Gyrus	n.a.	58	-42	8
	R SupraMarginal Gyrus	Area PFcm (IPL)	54	-38	26
	R SupraMarginal Gyrus	Area PF (IPL)	66	-38	30

Table 1. Assignment of activations to anatomical and cytoarchitectonic regions (Anatomy Toolbox, v2.2c). In order to characterize the anatomical locations of the cluster we report maxima within the clusters and their assignment to anatomical regions. If a maximum lies within a particular region, this means that the cluster extends into that anatomical region, but does not imply that the entire region is activated or that the entire cluster lies within that anatomical region. Whole brain analyses on the cluster level $p<0.05$ (FWE-corrected) with a cluster defining threshold of $p<0.001$. Contrast estimates from structures in bold font are plotted in Figures 2-4. n.a.: these maxima were not assigned to any region. *The anatomy toolbox labelled this maximum as Cerebellum but assigned it to the fusiform area FG2. **The anatomy toolbox labelled this maximum as Posterior-Medial Frontal cortex but did not assign it. The anatomical label was corrected to Supplementary motor cortex based on Neuromorphometrics labelling in SPM. ***The anatomy toolbox did not label this maximum. The anatomical label of left Thalamus was added based on Neuromorphometrics labelling in SPM.