Feature-specific prediction errors for visual mismatch support predictive coding accounts of perception

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] Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom

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: 19

Number of figures: 4, tables: 1

Number of words for Abstract: 211

Declaration of interest: none.

1

1 **Abstract**

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

17

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

20

21 **Highlights**

22 Predictive coding (PC) implies that brain responses should reflect transient precision weighted
23 prediction errors (pwPEs). By using fMRI we show that rare changes in color or emotional expression of
24 human faces elicit pwPE responses in dedicated neuronal circuits with known specialization in
25 processing color and emotion-related information, respectively. Our results demonstrate that physically
26 identical stimuli elicit pwPEs in distinct neuronal circuits when different (feature-specific) predictions are
27 violated, depending on current sensory expectations based on prior stimulus context. The
28 demonstration of pwPEs in visual areas as well as in emotion processing structures lends experimental
29 support to theoretical accounts of PC in color and social perception, respectively.

30

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: Instead of being a purely reactive, feed-forward analyzer of bottom-up sensory information (Hubel and Wiesel, 1965; Riesenhuber and Poggio, 2000), the brain is thought to actively predict 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; Spratling, 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 an elegant 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 on the implementation of PC is whether the same physical stimulus elicits separable feature-specific PE responses when distinct predictions about its various attributes exist and regardless whether such attributes are behaviorally relevant. To our knowledge, this was only 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 inferring an observer's belief and uncertainty about the hidden state of the world that generates the sensory information reaching the senses of the observer. The model tracks the beliefs of the observer about the probability of each stimulus feature and updates its inference as new information is presented trial-by-

70   trial. The HGF implements PC in the temporal domain and has been used in multiple studies to
71   investigate PE responses in the brain (e.g., Iglesias et al., 2013; Hauser et al., 2014; Schwartenbeck et al.,
72   2015; Vossel et al., 2015; Diaconescu et al., 2017; Lawson et al., 2017; Powers et al., 2017; Adams et al.,
73   2018; Katthagen et al., 2018; Stefanics et al., 2018a).

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

84   **Methods**

85   Ethics Statement

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

89   Subjects

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

94   Paradigm

95   Faces were presented in four peripheral quadrants of the screen (Fig. 1A) on a grey background with a
96   fixation cross in the center. Each stimulus panel contained four faces of different identity expressing the
97   same emotion. Stimulus duration was 200ms. The stimuli were presented after an inter-stimulus interval
98   of 550ms during which only the fixation cross was present. A change detection task was presented at
99   the central fixation cross. Roving paradigms have frequently been used to study automatic sensory
100   expectation effects (Haenschel et al., 2005; Garrido et al., 2008; Costa-Faidella et al., 2011; Moran et al.,
101   2013; Auksztulewicz and Friston, 2015; Stefanics et al., 2018a,b). Here, we used a multi-feature visual
102   'roving standard' paradigm to elicit PE responses by unexpected changes either in color (red, green), or
103   emotional expression (happy, fearful) of human faces, or both. Importantly, this allowed us to study
104   how brain responses to physically identical stimuli differed, depending on the degree of expectations

4

105   about color and emotion, respectively. A diagram of the transitions between stimulus types is shown in
106   Fig. 1B.
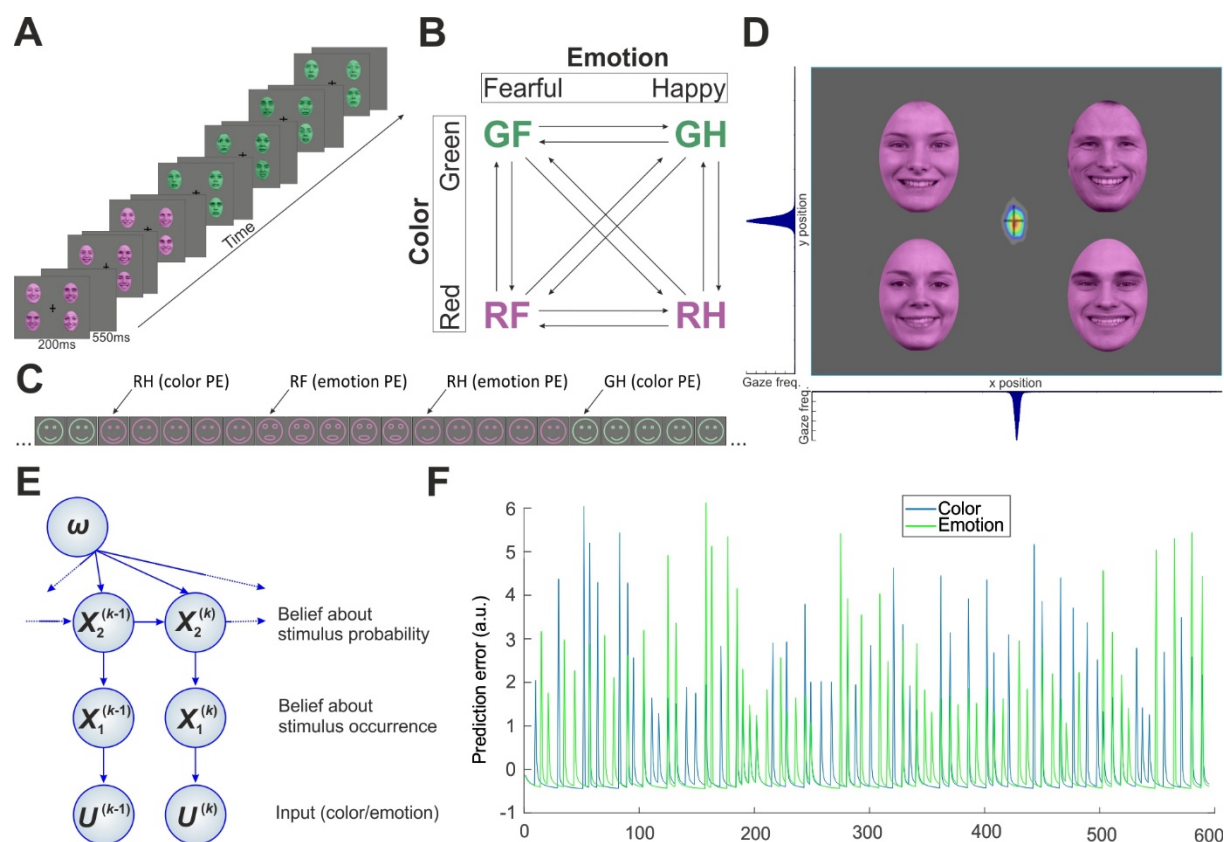
Fig. 1. Experimental design and eye-tracking results. A) Four individual photographs of the same color displaying the same facial affect were presented in each stimulus panel for 200ms in a roving standard paradigm. Each panel was followed by an empty grey screen presented for 550ms. The vertical and 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.

121   Images were taken from the Radboud Faces Database (Langner et al., 2010). Ten female and ten male
122   Caucasian models were selected based on their high percentage of agreement on emotion
123   categorization (98% for happy, 92% for fearful faces). A Wilcoxon rank sum test indicated that
124   categorization agreement on the emotional expressions did not differ between happy and fearful faces
125   (Z=-0.63, p=0.53). To control low-level image properties, we equated the luminance and the spatial
126   frequency content of grayscale images of the selected happy and fearful faces using the SHINE toolbox
127   (Willenbockel et al., 2010). The resulting images were used to create the colored stimuli.

128 Behavioral task

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

137 Eye-tracking

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

145 Data acquisition and preprocessing

146 FMRI data was acquired on a Philips Achieva 3 Tesla scanner using an eight channel head-coil (Philips,
147 Best, The Netherlands) at the Laboratory for Social and Neural Systems Research at the University of
148 Zurich. A structural image was acquired for each participant with a T1-weighted MPRAGE sequence: 181
149 sagittal slices, field of view (FOV): 256 × 256 mm2, Matrix: 256 x 256, resulting in 1 mm^3 resolution.
150 Functional imaging data was acquired in six experimental blocks. In each block 200 whole-brain images
151 were acquired using a T2*-weighted echo-planar imaging sequence with the following parameters. 42
152 ascending transverse plane slices with continuous in-plane acquisition (slice thickness: 2.5 mm; in-plane
153 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
154 view = 220 x 220 x 130 mm; SENSE factor = 1.5; EPI factor = 51). We used a 2nd order pencil-beam
155 shimming procedure provided by Philips to reduce field inhomogeneities during the functional scans. All
156 functional images were reconstructed with 3 mm isotropic resolution. Functional data acquisition lasted
157 approximately 1 hour. During fMRI data acquisition, respiratory and cardiac activity was recorded using
158 a breathing belt and an electrocardiogram, respectively.

159 We used statistical parametric mapping (SPM12, v6470; RRID: SCR_007037; Friston et al., 2007) for fMRI
160 data analysis. First, functional images were slice time corrected, realigned to correct for motion and co-
161 registered with the subject's own anatomical image. Next, we normalized structural images to MNI
162 space using the unified segmentation approach and applied the same warping to normalize functional
163 images. The functional images were smoothed with a 6 mm full-width at half maximum Gaussian kernel
164 and resampled to 2 mm isotropic resolution. We used RETROICOR (Glover et al., 2000) as implemented
165 in   the   PhysIO-Toolbox   (Kasper   et   al.,   2017)   from   the   open   source   software   TAPAS

6

166 (http://www.translationalneuromodeling.org/tapas) to create confound regressors for cardiac

167 pulsations, respiration, and cardio-respiratory interactions. These confound regressors were entered in

168 the general linear model (GLM; see below).

169 <u>Modeling belief trajectories</u>

170 In order to include parametric regressors of precision weighted prediction errors (pwPE) in the GLM, we

171 simulated trajectories of belief update in a Bayesian generative model of perceptual inference, the

172 Hierarchical Gaussian Filter (HGF; Mathys et al., 2011; 2014). We followed the approach described in

173 details in Stefanics et al. (2018a) using the HGF toolbox version v2.2 contained in TAPAS

174 (http://www.translationalneuromodeling.org/tapas), a collection of algorithms and software tools to

175 support computational modeling. Briefly, we simulated the perceptual model of a two-level HGF for the

176 input traces given by the two features of the face stimuli: color (red vs. green) and emotion (fearful vs.

177 happy). Inversion of the HGF (Fig. 1E) infers the hidden states (x) of the world that generate the sensory

178 input (u). The belief states are updated after each trial following a generic update rule: The posterior

179 mean $\mu_2^{(k)}$ of state x_2 at trial k changes its value according to a precision-weighted PE $\varepsilon_2^{(k)}$:

180
$$\mu_2^{(k)} - \mu_2^{(k-1)} \propto \varepsilon_2^{(k)} \qquad\qquad (1)$$

181 Classical reinforcement learning models (e.g. Rescorla and Wagner, 1972) follow the same form:

182
$$\text{prediction}^{(k)} = \text{prediction}^{(k-1)} + \text{learning rate} \times \text{PE} \qquad\qquad (2)$$

183 In the HGF the learning rate is determined by the precision of the state and the precision at the level

184 below. For the simulations we assumed that color and emotion were processed by two separate,

185 independent HGFs. We estimated the parameters of the model assuming an ideal Bayes-optimal

186 observer (Mathys et al., 2011) that minimizes surprise of the incoming input stream. Figure 1F displays

187 example traces of the absolute value of ε_2 which entered the GLM as described below.

188 <u>General linear model analysis</u>

189 The fMRI data was analyzed with two separate GLMs. One included the gradually changing (absolute)

190 pwPE derived from the HGF as modulatory regressors while the other incorporated an all or none

191 categorical change detection (CD; see Lieder et al. 2013). For the GLM based on the CD model, we

192 included stick functions as parametric modulators for each stimulus on those trials when a change

193 occurred in the stimulus sequence. The GLMs were estimated for each participant individually. Both the

194 pwPE as well as the CD modulatory regressors were computed separately for color and emotion. In

195 addition the GLM included modulatory regressors for red vs. green and happy vs. fearful, respectively.

196 Hence, for each run of the experiment the design matrix included the following experimental regressors:

197 i) a main regressor for the onset of each stimulus display, ii) two modulatory regressors encoding color

198 (red = -1, green = 1) and emotion (happy = -1, fearful = 1), respectively, and iii) two modulatory

199 regressors with the absolute pwPE (or CD) for color and emotion, respectively. The modulatory

200 regressors were mean centered. In addition to these regressors of interest, button presses to cross-flips

201 of the visual attention task were also included in the model. All regressors were convolved with a

7

202 canonical hemodynamic response function (HRF). Movement regressors and physiological confounds
203 were included in the first level GLM (Kasper et al., 2017) which was estimated for each participant
204 individually. Please note that the sign of colors and emotions in ii) was arbitrarily chosen.

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

213 **Results**

214 <u>Fixation and behavioral responses</u>

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

219 First-level GLMs

220 We fitted two GLMs on the single-subject level, incorporating parametric regressors that represented
221 two hypotheses about the decay of PE responses following a change in color of emotional expression of
222 the faces. Similar to the model comparison procedure described in our previous study, our aim was to
223 create a functionally defined mask of significant voxels showing PE responses under both models at the
224 group level (Stefanics et al., 2018a). However, while similar activation clusters were obtained using the
225 pwPE and CD regressors to color changes, significant clusters to changes in emotion were only found
226 using the pwPE regressors. In other words, the beta estimates obtained using CD were not consistent
227 enough to yield significant activation clusters at the group level. The lack of significant group-level
228 results for the stick regressors prevented us from creating an unbiased mask comprising significant
229 voxels for color and emotion ("logical AND" conjunction). We thus restrict ourselves to report the results
230 obtained at the group-level analysis using the HGF-based pwPE model.

231 <u>Effect of color pwPEs</u>

232 A whole-brain analysis of color changes showed significant activation for color pwPE in fusiform areas
233 (Fig. 2A). Post inspection of the activation (Fig. 2B) revealed an increased response to pwPE. Detailed
234 information about anatomical labels, cluster size, and MNI coordinates for the maxima of significant
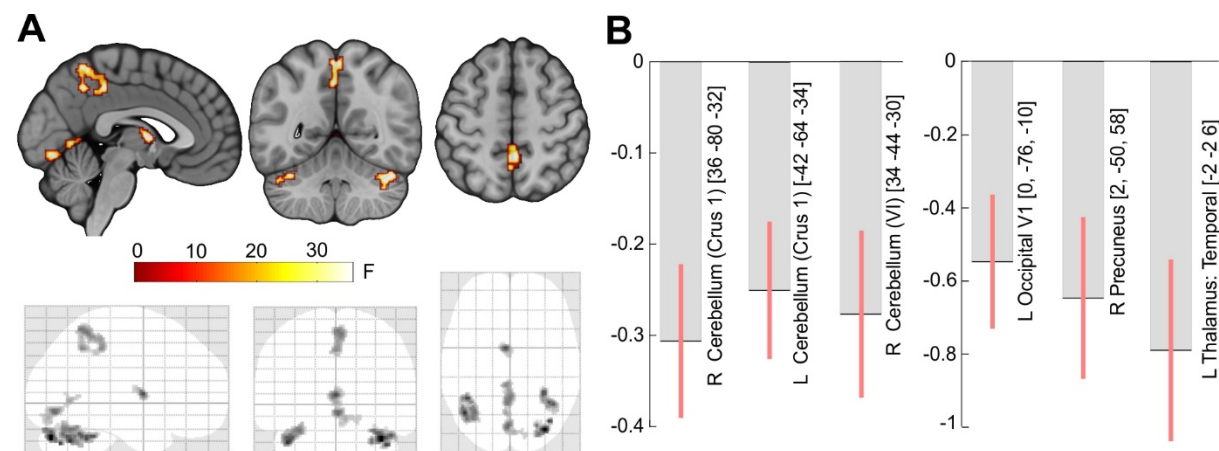235 voxel clusters are listed in Table 1.
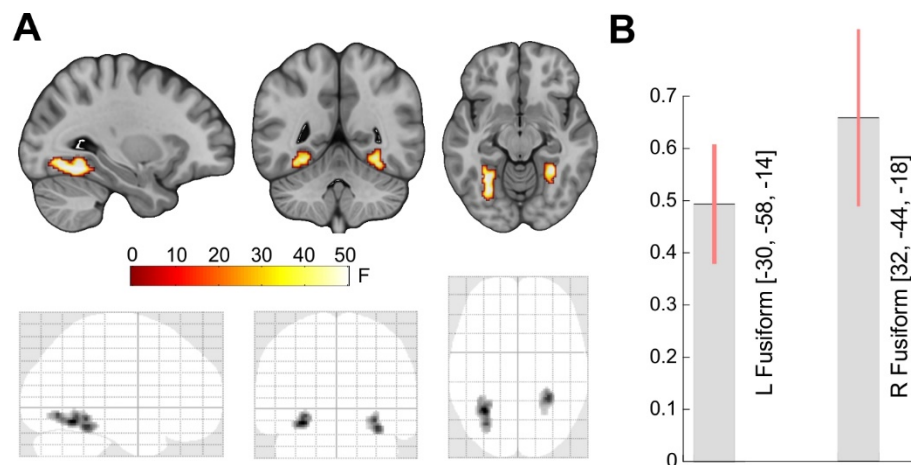
Fig. 2. Effect of color pwPEs. A) Top: Activation map overlaid onto the MNI152 standard-space T1-weighted average structural template. Slices show activation in the left fusiform gyrus (MNI-coordinates: [-30 -50 -12]. **Bottom:** Glass brain showing the results of the F-test. **B)** Contrast estimates (arbitrary units) for color pwPEs in the left and right fusiform gyrus. Bars indicate 90% C.I.

Effect of emotion PEs

A whole-brain analysis of emotion PEs showed significant effects in bilateral cerebellar areas, bilateral precuneus, bilateral lingual gyrus (LG), and left thalamus (Fig. 3A). A post hoc analysis of the contrast estimates in these regions (Fig. 3B) revealed that all areas showed a negative effect of emotion pwPEs.


Fig. 3. Main effect of emotion pwPEs. A) Top: Activation map overlaid onto the MNI152 standard-space T1-weighted average structural template. Slices show activations overlaid onto the MNI152 standard-space T1-weighted average structural template at coordinates [-3, -55, 55] cutting through the left 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 lingual gyrus, right precuneus, and left thalamus. 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

9

256 trial update of internal hidden belief states in the HGF relies on precision weighted prediction errors.
257 Hence, traces of the latter served as regressors in a GLM which yielded brain structures where activation
258 showed a significant relationship to precision-weighted PEs of color and emotional expression of faces,
259 respectively. We manipulated sensory expectations towards color and emotional expression of faces
260 independently. Crucially, emotion and color pwPEs were evoked by physically identical stimuli; only the
261 expectation, i.e. violation of regularity, differed between the two conditions. While our previous EEG
262 study reported the scalp distribution and time-course of pwPE responses (Stefanics et al., 2018a), here
263 we used fMRI to find BOLD correlates of pwPEs in generator structures. We found BOLD correlates of
264 pwPEs to color changes in bilateral fusiform gyrus. pwPEs to changes of emotional expressions activated
265 a different set of areas including the bilateral cerebellum, lingual gyrus, precuneus, and the left
266 thalamus (Fig. 4A).

267
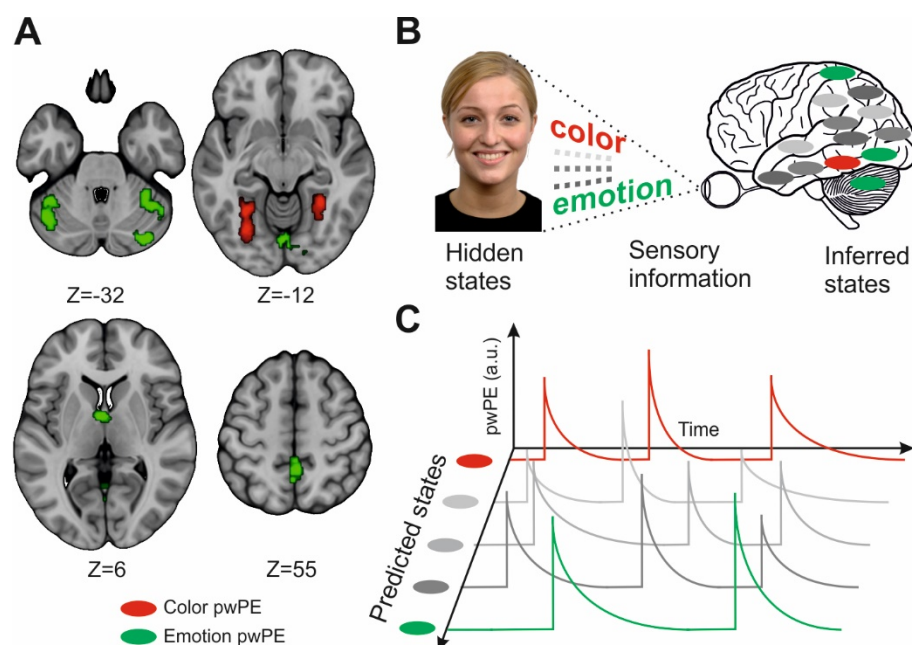
Fig. 4. Overview of the results and PC framework for perceptual prediction errors. A) Colored areas mark main clusters related to color pwPEs (red), and emotion pwPEs (green). Note the dissociation of PEs for color and emotion changes. B) Schematic depicting functional segregation in the nervous system, as distinct features of the world are inferred and predicted by distinct neural structures specializing in the given features. To-be-inferred hidden states (causes) for color and emotion are marked in red and green as in A), additional features (to-be-inferred hidden states) are marked with grey. Image of a model used in our study reproduced with permission of the Radboud Faces Database (Langner et al., 2010). C) pwPEs for different features arise and are explained away during Bayesian inference as properties of the hidden cause that generate the sensory information dynamically change over time.

276 The demonstration of activations correlated to pwPE in ventral visual areas as well as in emotion
277 processing structures lends experimental support to theoretical accounts of PC in color and emotion
278 perception. From the perspective of PC, functional architectures exist to infer hidden causes of specific
279 sensory information and compare predictions based on inference to observed sensory input (Friston,
280 2002). Importantly, we manipulated stimulus sequences to induce automatic expectations about the
281 occurrence of different stimulus features, using the same faces to elicit distinct emotion and color
282 pwPEs. In line with our hypothesis, color and emotion pwPEs were reflected by activity in brain

283 structures known to be dedicated to color and emotion processing. A hypothesized generalization of our
284 results is shown in Fig. 4B, which illustrates functional segregation of inferring and predicting hidden
285 causes of sensory information for different features, including color and emotional expression of faces.
286 According to PC, creating and maintaining our internal model of the world is a process during which
287 predictive object representations about the likely properties of the hidden objects are updated using
288 precision-weighted PEs (e.g., Moran et al., 2013; Stefanics et al., 2018a) that signals mismatch between
289 the expectations based on prior information and the current sensory data (Fig. 4C).

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

294 To our knowledge this is the first fMRI study using a computational model that simulates the belief
295 trajectories of an optimal Bayesian observer to describe automatic pwPEs to violations of expectations
296 to different features of the same objects. Our paradigm allowed us to study pwPEs in the absence of
297 focal attention and task-relevance. A previous study found that PEs spread across object features in the
298 visual cortex (Jiang et al., 2016). Here, we show that (i) pwPEs can also be elicited in spatially remote
299 neural structures that specialize in the processing of distinct stimulus attributes and (ii) in the absence of
300 attention. Notably, Jiang et al. (2016) studied PEs to attended and task-relevant random dot stimuli,
301 while in our study face stimuli were task-irrelevant and not attended, as verified by eye tracking. The
302 differences between our current and their results suggest that the role of focal attention in perception
303 might not only be to enhance (e.g., Auksztulewicz and Friston, 2015) but also spread PEs across features
304 at the object level (Jiang et al., 2016) which is in line with the feature-integration theory of attention
305 (Treisman and Gelade, 1980).

306 Color PEs

307 Color processing involves the ventral visual pathway (Mesulam, 1998; Bartels and Zeki, 2000), where
308 fMRI studies have shown strong color-related activations (Brewer et al., 2005; Solomon and Lennie,
309 2007; Barbur and Spang, 2008; Brouwer and Heeger, 2009). The location of the fusiform activation in
310 our experiment is in agreement with "color-biased" regions in the ventral occipito-temporal cortex
311 (Lafer-Sousa et al., 2016). The abundance of reciprocal connections in cortex (Felleman and Van Essen,
312 1991; Markov et al., 2013) indicates that information flow is likely to be bi-directional in the cortical
313 hierarchy. This is in line with PC, where backward connections convey prior knowledge about the
314 hierarchical causal structure of the world (Friston, 2005) whereas the role of forward connections is to
315 convey PE to higher levels and update the internal predictive model of the world. One limitation of our
316 current study is that we cannot separate the putative contributions of bottom-up and top-down
317 mechanisms to the fMRI activations. In sum, our results emphasize the importance of pwPEs and
318 represent an important advance by providing previously unavailable support for a PC account of color
319 perception, given that our parameter estimates likely captured BOLD correlates of pwPEs and model
320 updating processes within color-biased regions of the fusiform cortex.

11

321   Emotion PEs

322   Facial emotions are non-verbal acts of communication that express emotional states and intentions, and
323   are fundamental in social interactions (Fridlund, 1994; Frith, 2009). The social environment is not
324   constant, and detecting changes in the emotional valence of facial expressions in our social space is
325   important for socially successful behavior. Prior ERP studies (Susac et al., 2004; Kimura et al., 2012; Li et
326   al., 2012; Csukly et al., 2013; Stefanics et al., 2012, 2018a; Astikainen et al., 2013; Fujimura and
327   Okanoya, 2013; Xu et al., 2018) suggest that emotional expressions are processed in a few hundred
328   milliseconds and stored in predictive memory representations. We found emotion pwPEs in a set of
329   areas including the bilateral cerebellum, lingual gyrus, precuneus, and left thalamus. This pattern of
330   results (Fig. 4A) is in line with the notion that emotion processing involves a mosaic-like set of affective,
331   motor-related and sensory components (Bastiaansen et al., 2009).

332   Processing of emotional faces involves the cerebellum, lingual gyrus and precuneus (Fusar-Poli et al.,
333   2009; E et al., 2014; Adamaszek et al., 2017), among other structures. The cerebellar areas showing
334   emotion pwPE activity included bilateral lobules VI and VIIa (Crus 1), supramodal zones which are
335   functionally connected with prefrontal and parietal cortex including the precuneus (Allen et al., 2005;
336   Habas et al., 2009; Buckner et al., 2011). Furthermore, we observed emotion pwPEs in early visual areas
337   known to be activated by emotional faces (Fusar-Poli et al., 2009), as well as the left thalamus. Emotion
338   pwPEs in the lingual gyrus are consistent with our hypothesis that PEs should emerge in structures that
339   likely encode low-level structural attributes in response to changes in facial emotions. The precuneus is
340   a major association area active for a wide range of cognitive processes, in particular somatosensory and
341   visuomotor integration (Bruner et al., 2017). It is functionally connected to sensorimotor areas (Cavanna
342   and Trimble, 2006) as well as early visual areas (Zhang and Li, 2012), thus it is ideally suited to transform
343   sensory information from the lingual gyrus about observed changes in the superficial geometry of faces
344   into a somatosensory representation. The cerebellum does not contain a representation of early visual
345   areas (Buckner et al., 2011) therefore we suggest that cerebellar pwPEs observed in our study were
346   computed based on input from the precuneus. Subcortical connections of the precuneus target the basis
347   pontis allowing the precuneus to access multiple cerebellar circuits (Cavanna and Trimble, 2006).
348   Cerebellar output controls various parts of the thalamo-cortical network (Gomati et al., 2018) which
349   possibly allows the cerebellum to modulate precuneus activity. The cerebellum is thought to support
350   computations of forward and inverse internal models and error signals (Ito, 2008; Roth et al., 2013; Popa
351   et al., 2014; Van Overwalle and Mariën, 2016; Sokolov et al., 2017). While a forward model relies on
352   motor commands and reproduces the dynamics of the controlled object, inverse models in the
353   cerebellum infer motor commands. Thus, the cerebellum's role in emotion recognition might involve
354   inferring the motor commands that resulted in the complex, coordinated position of a number of facial
355   muscles which, in turn, resulted in the observed superficial geometry of the face (emotional expression).
356   Accordingly, the precuneus might infer the internal states generating the observed facial expressions,
357   i.e., emotions and intentions, relying on information inferred by the cerebellum about motor
358   commands. We speculate that pwPEs generated in the cerebellum might signal a mismatch between the
359   expected somatosensory information provided by the precuneus, based on previously observed
360   emotional expressions, and the information for an unexpected facial emotion on a current trial. Based

12

361   on the observed mosaic-like set of regions showing pwPEs we hypothesize that automatic emotion
362   recognition is supported by a thalamo-cortico-cerebellar network of regions that interact to infer the
363   hidden internal states underlying the observed facial expression. The logic here is similar to that of PC
364   theories of social perception (Kilner et al., 2007; Friston et al., 2011; Caligiore et al., 2013; Ishida et al.,
365   2015), suggesting that inverse models can be used as recognition models and therefore can infer the
366   cause of an observed action. Given that static facial emotional expressions are the outcome of specific
367   communicative actions by facial muscles to convey a message about the internal state of the agent
368   (Ekman and Friesen, 1978; Redcay, 2008; Furl et al., 2010; Cattaneo and Pavesi, 2014), we suggest that
369   emotion recognition might involve inference about the hierarchical structure of hidden causes (change
370   in light reaching the retina ← change in superficial face geometry ← change in underlying musculature
371   ← motor commands ← internal emotional state) generating observable facial emotions.

372   In summary, our findings demonstrate that the same physical stimulus can elicits separate feature-
373   specific pwPE responses, depending on distinct predictions about its various attributes. This is in
374   agreement with PC theories of perception. Future extensions of our work could involve using
375   computational models of effective connectivity in order to examine the network dynamics generating
376   pwPEs as postulated by PC.

13

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.

Allen G, McColl R, Barnard H, Ringe WK, Fleckenstein J, Cullum CM (2005) Magnetic resonance imaging of cerebellar-prefrontal and cerebellar-parietal functional connectivity. Neuroimage 28:39-48.

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.

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.

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.

Bruner E, Preuss TM, Chen X, Rilling JK (2017) Evidence for expansion of the precuneus in human evolution. Brain Struct Funct. 222:1053-1060. doi: 10.1007/s00429-015-1172-y.

Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BT (2011) The organization of the human cerebellum estimated by intrinsic functional connectivity. J Neurophysiol 106:2322-2345.

Caligiore D, Pezzulo G, Miall RC, Baldassarre G (2013) The contribution of brain sub-cortical loops in the expression and acquisition of action understanding abilities. Neurosci Biobehav Rev 37:2504-2515.

Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional anatomy and behavioural correlates. Brain 129:564-583.

Cattaneo L, Pavesi G (2014) The facial motor system. Neurosci Biobehav Rev 38:135-159.

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.

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.

Ekman P, Friesen W (1978) Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press.

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.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

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

434 Fridlund AJ (1994) Human facial expression: An evolutionary view. San Diego, CA: Academic Press.
435 Friston K (2002) Functional integration and inference in the brain. Progress in Neurobiology 68:113–143.
436 Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815-836.
437 Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127-138.
438 Friston KJ, Ashburner J, Kiebel SJ, Nichols TE, Penny WD, eds. (2007) Statistical Parametric Mapping: The Analysis of Functional
439   Brain Images. Academic Press.
440 Friston K, Mattout J, Kilner J (2011) Action understanding and active inference. Biol Cybern 104:137-160.
441 Friston K, Adams R, Montague R (2012a) What is value-accumulated reward or evidence? Front Neurorobot 6:11.
442 Friston K, Adams RA, Perrinet L, Breakspear M (2012b) Perceptions as hypotheses: saccades as experiments. Front Psychol
443   3:151.
444 Frith C (2009) Role of facial expressions in social interactions. Philos Trans R Soc Lond B Biol Sci 364:3453-3458.
445 Fujimura T, Okanoya K (2013) Event-related potentials elicited by pre-attentive emotional changes in temporal context. PLoS
446   One 8:e63703.
447 Furl N, van Rijsbergen NJ, Kiebel SJ, Friston KJ, Treves A, Dolan RJ (2010) Modulation of perception and brain activity by
448   predictable trajectories of facial expressions. Cereb Cortex 20:694-703.
449 Fusar-Poli P, Placentino A, Carletti F, Landi P, Allen P, Surguladze S, Benedetti F, Abbamonte M, Gasparotti R, Barale F, Perez J,
450   McGuire P, Politi P (2009) Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105
451   functional magnetic resonance imaging studies. J Psychiatry Neurosci 34:418-432.
452 Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM (2008) The functional anatomy of the MMN: a DCM study of
453   the roving paradigm. Neuroimage 42:936-944.
454 Glover GH, Li TQ, Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI:
455   RETROICOR. Magn Reson Med 44:162-167.
456 Gordon N, Koenig-Robert R, Tsuchiya N, van Boxtel JJ, Hohwy J (2017) Neural markers of predictive coding under perceptual
457   uncertainty revealed with Hierarchical Frequency Tagging. Elife 6.
458 Gornati SV, Schäfer CB, Eelkman Rooda OHJ, Nigg AL, De Zeeuw CI, Hoebeek FE (2018) Differentiating Cerebellar Impact on
459   Thalamic Nuclei. Cell Rep 23:2690-2704. doi: 10.1016/j.celrep.2018.04.098.
460 Habas C, Kamdar N, Nguyen D, Prater K, Beckmann CF, Menon V, Greicius MD (2009) Distinct cerebellar contributions to
461   intrinsic connectivity networks. J Neurosci 29:8586-8594.
462 Haenschel C, Vernon DJ, Dwivedi P, Gruzelier JH, Baldeweg T (2005) Event-related brain potential correlates of human auditory
463   sensory memory-trace formation. J Neurosci 25:10494-10501.
464 Hubel DH, Wiesel TN (1965) Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat.
465   J Neurophysiol 28:229-289.
466 Ishida H, Suzuki K, Grandi LC (2015) Predictive coding accounts of shared representations in parieto-insular networks.
467   Neuropsychologia 70:442-454.
468 Ito M (2008) Control of mental activities by internal models in the cerebellum. Nat Rev Neurosci. 9:304-313.
469 Jiang J, Summerfield C, Egner T (2016) Visual Prediction Error Spreads Across Object Features in Human Visual Cortex. J
470   Neurosci 36:12746-12763.
471 Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinzle J, Iglesias S, Hauser TU, Sebold M, Manjaly ZM, Pruessmann KP,
472   Stephan KE (2017) The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. J Neurosci Methods 276:56-72.
473 Katthagen T, Mathys C, Deserno L, Walter H, Kathmann N, Heinz A, Schlagenhauf F (2018) Modeling subjective relevance in
474   schizophrenia and its relation to aberrant salience. PLoS Comput Biol 14(8): e1006319.
475   https://doi.org/10.1371/journal.pcbi.1006319
476 Kilner JM, Friston KJ, Frith CD (2007) Predictive coding: an account of the mirror neuron system. Cogn Process 8:159-166.
477 Kimura M, Kondo H, Ohira H, Schroger E (2012) Unintentional temporal context-based prediction of emotional faces: an
478   electrophysiological study. Cereb Cortex 22:1774-1785.
479 Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP (2012) Attention reverses the effect of prediction in silencing sensory signals.
480   Cereb Cortex 22:2197-2206.
481 Krassanakis V, Filippakopoulou V, Nakos B (2014) EyeMMV toolbox: An eye movement post-analysis tool based on a two-step
482   spatial dispersion threshold for fixation identification. J Eye Movement Res 7.
483 Kreegipuu K, Kuldkepp N, Sibolt O, Toom M, Allik J, Näätänen R (2013) vMMN for schematic faces: automatic detection of
484   change in emotional expression. Front Hum Neurosci 7:714.
485 Kremlacek J, Kreegipuu K, Tales A, Astikainen P, Poldver N, Naatanen R, Stefanics G (2016) Visual mismatch negativity (vMMN):
486   A review and meta-analysis of studies in psychiatric and neurological disorders. Cortex 80:76-112.
487 Kuldkepp N, Kreegipuu K, Raidvee A, Näätänen R, Allik J (2013) Unattended and attended visual change detection of motion as
488   indexed by event-related potentials and its behavioral correlates. Front Hum Neurosci 7:476. doi:
489   10.3389/fnhum.2013.00476.
490 Lafer-Sousa R, Conway BR, Kanwisher NG (2016) Color-Biased Regions of the Ventral Visual Pathway Lie between Face- and
491   Place-Selective Regions in Humans, as in Macaques. J Neurosci 36:1682-1697.

492 Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A (2010) Presentation and validation of the Radboud
493   Faces Database. Cognition & Emotion 24:1377-1388.
494 Lawson RP, Mathys C, Rees G (2017) Adults with autism overestimate the volatility of the sensory environment. Nat Neurosci
495   20:1293-1299.
496 Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am 20.
497 Li X, Lu Y, Sun G, Gao L, Zhao L (2012) Visual mismatch negativity elicited by facial expressions: new evidence from the
498   equiprobable paradigm. Behav Brain Funct 8:7.
499 Lochmann T, Ernst UA, Deneve S (2012) Perceptual inference predicts contextual modulations of sensory responses. J Neurosci
500   32:4179-4195.
501 Markov NT, Ercsey-Ravasz M, Van Essen DC, Knoblauch K, Toroczkai Z, Kennedy H (2013) Cortical high-density counterstream
502   architectures. Science 342:1238406.
503 Mesulam MM (1998) From sensation to cognition. Brain 121 (Pt 6):1013-1052.
504 Moran RJ, Campo P, Symmonds M, Stephan KE, Dolan RJ, Friston KJ (2013) Free energy, precision and learning: the role of
505   cholinergic neuromodulation. J Neurosci 33:8227-8236.
506 Müller D, Winkler I, Roeber U, Schaffer S, Czigler I, Schröger E (2010) Visual object representations can be formed outside the
507   focus of voluntary attention: evidence from event-related brain potentials. J Cogn Neurosci 22:1179-1188.
508 Popa LS, Hewitt AL, Ebner TJ (2014) The cerebellum for jocks and nerds alike. Front Syst Neurosci. 8:113.
509 Powers AR, Mathys C, Corlett PR (2017) Pavlovian conditioning-induced hallucinations result from overweighting of perceptual
510   priors. Science 357:596-600.
511 Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-
512   field effects. Nat Neurosci 2:79-87.
513 Redcay E (2008) The superior temporal sulcus performs a common function for social and speech perception: implications for
514   the emergence of autism. Neurosci Biobehav Rev 32:123-142.
515 Riesenhuber M, Poggio T (2000) Models of object recognition. Nat Neurosci 3 Suppl:1199-1204.
516 Roth MJ, Synofzik M, Lindner A (2013) The cerebellum optimizes perceptual predictions about external sensory events. Curr Biol
517   23:930-935. doi: 10.1016/j.cub.2013.04.027.
518 Schwartenbeck P, FitzGerald TH, Mathys C, Dolan R, Friston K (2015) The Dopaminergic Midbrain Encodes the Expected
519   Certainty about Desired Outcomes. Cereb Cortex 25:3434-3445.
520 Schwiedrzik CM, Freiwald WA (2017) High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing
521   Hierarchy. Neuron 96:89-97 e84.
522 Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD (2016) Neural signatures of perceptual
523   inference. Elife 5:e11476.
524 Smith FW, Muckli L (2010) Nonstimulated early visual areas carry information about surrounding context. P Natl Acad Sci USA
525   107:20099-20103.
526 Sokolov AA, Miall RC, Ivry RB (2017) The Cerebellum: Adaptive Prediction for Movement and Cognition. Trends Cogn Sci
527   21(5):313-332. doi: 10.1016/j.tics.2017.02.005.
528 Solomon SG, Lennie P (2007) The machinery of colour vision. Nat Rev Neurosci 8:276-286.
529 Spratling MW (2017) A review of predictive coding algorithms. Brain Cogn 112:92-97.
530 Stefanics G, Kimura M, Czigler I (2011) Visual mismatch negativity reveals automatic detection of sequential regularity violation.
531   Front Hum Neurosci 5:46.
532 Stefanics G, Kremlacek J, Czigler I (2014) Visual mismatch negativity: a predictive coding view. Front Hum Neurosci 8:666.
533 Stefanics G, Heinzle J, Horvath AA, Stephan KE (2018a) Visual Mismatch and Predictive Coding: A Computational Single-Trial ERP
534   Study. J Neurosci 38:4020-4030.
535 Stefanics G, Heinzle J, Czigler I, Valentini E, Stephan KE (2018b) Timing of repetition suppression of event-related potentials to
536   unattended objects. Eur J Neurosci doi:10.1111/ejn.13972
537 Stefanics G, Csukly G, Komlosi S, Czobor P, Czigler I (2012) Processing of unattended facial emotions: a visual mismatch
538   negativity study. Neuroimage 59:3042-3049.
539 Susac A, Ilmoniemi RJ, Pihko E, Supek S (2004) Neurodynamic studies on emotional and inverted faces in an oddball paradigm.
540   Brain Topogr 16:265-268.
541 Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97-136.
542 Van Overwalle F, Mariën P (2016) Functional connectivity between the cerebrum and cerebellum in social cognition: A multi-
543   study analysis. Neuroimage 124(Pt A):248-255. doi: 10.1016/j.neuroimage.2015.09.001.
544 Vossel S, Mathys C, Stephan KE, Friston KJ (2015) Cortical Coupling Reflects Bayesian Belief Updating in the Deployment of
545   Spatial Attention. J Neurosci 35:11532-11542.
546 Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions
547   and prediction errors in human cortex. Proceedings of the National Academy of Sciences of the United States of
548   America 108:20754-20759.

549 Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: the SHINE
550   toolbox. Behav Res Methods 42:671-684.
551 Woo CW, Krishnan A, Wager TD (2014) Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations.
552   Neuroimage 91:412-419.
553 Xu QR, Ruohonen EM, Ye CX, Li XQ, Kreegipuu K, Stefanics G, Luo WB, Astikainen P (2018) Automatic Processing of Changes in
554   Facial Emotions in Dysphoria: A Magnetoencephalography Study. Frontiers in Human Neuroscience 12:186.
555 Zhang S, Li CS (2012) Functional connectivity mapping of the human precuneus by resting state fMRI. Neuroimage 59:3548-
556   3562.

557

558

559 **Table 1.**

Contrast and Cluster	Structure	Cytoarchitectonic area	Cluster max. (MNI)		
Main effects of color PEs					
Cluster 1 (366voxels)	L Fusiform Gyrus	Area FG3	-30	-58	-14
	L Fusiform Gyrus	Area FG1	-28	-70	-8
	L Fusiform Gyrus	Area FG1	-28	-66	-12
Cluster 2 (333voxels)	R Fusiform Gyrus	Area FG3	32	-44	-18
	R Fusiform Gyrus	Area FG3	30	-46	-12
	R Fusiform Gyrus	Area FG3	28	-60	-10
Main effects of emotion PEs					
Cluster 1 (315 voxels)	L Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	-42	-64	-34
	L Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	-40	-50	-32
	L Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	-40	-70	-28
	L Cerebellum (VI)	Lobule VI (Hem)	-32	-64	-24
	L Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	-44	-60	-30
	L Cerebellum (VI)	Lobule VIIa crus I (Hem)	-34	-68	-24
	L Cerebellum (VI)	Lobule VI (Hem)	-36	-54	-28
	L Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	-38	-60	-28
Cluster 2 (263 voxels)	R Precuneus	Area 5m (SPL)	2	-50	58
	L Precuneus	n.a.	0	-54	60
	L MMC	n.a.	-2	-38	44
	L Precuneus	n.a.	-2	-54	44
	L MMC	Area 5m (SPL)	0	-38	52
	L Precuneus	n.a.	-2	-54	54
	L MMC	n.a.	0	-42	40
Cluster 3 (150 voxels)	Cerebellar Vermis (6)***	L Area hOc1 [V1]	0	-76	-10
	Cerebellar Vermis (4/5)***	L Area hOc1 [V1]	-2	-64	0
	n.a.	R Area hOc1 [V1]	24	-66	38
	Cerebellar Vermis (7)	R Lobule VI (Verm)	6	-82	-14
	R Lingual Gyrus	R Area hOc2 [V2]	12	-82	-14
	L Lingual Gyrus *	Area hOc2 [V2]	0	-58	4
	R Lingual Gyrus	R Area hOc1 [V1]	16	-86	-14
	L Lingual Gyrus	L Area hOc1 [V1]	0	-68	6
Cluster 4 (150 voxels)	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	36	-80	-32
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	40	-72	-36
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	30	-84	-24
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	28	-76	-34
	R Cerebellum (VI)**	R Cerebellum Exterior	30	-74	-20
Cluster 5 (147 voxels)	R Cerebellum (VI)	Lobule VI (Hem)	34	-44	-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 (VI)	Lobule VI (Hem)	34	-52	-30
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	38	-58	-36
	R Cerebellum (Crus 1)	Lobule VIIa crus I (Hem)	48	-62	-32
Cluster 6 (71 voxels)	L Thalamus: Temporal	n.a.	-2	-2	6

560

561 **Table 1.** Assignment of activations to anatomical and cytoarchitectonic regions (Anatomy Toolbox,
562 v2.2c). In order to characterize the anatomical locations of the cluster we report maxima within the
563 clusters and their assignment to anatomical regions. If a maximum lies within a particular region, this
564 means that the cluster extends into that anatomical region, but does not imply that the entire region is
565 activated or that the entire cluster lies within that anatomical region. Whole brain analyses on the
566 cluster level p<0.05 (FWE-corrected) with a cluster defining threshold of p<0.001. n.a.: these maxima
567 were not assigned to any region. *The anatomy toolbox labelled this maximum as Cerebellum but
568 assigned it to the visual cortex. The anatomical label was corrected to L Lingual Gyrus after visual
569 inspection. ** The anatomy toolbox labelled this maximum as Cerebellum but assigned it to the visual

18

570   cortex. The anatomical label was corrected to R Cerebellum Exterior after visual inspection. *** These

571   cluster maxima were assigned to the cerebellar vermis, the anatomical label was corrected to L V1 after

572   visual inspection.