Detecting (un)seen change: The neural underpinnings of (un)conscious prediction errors

Authors: Elise G. Rowe[1,2,4,5], Naotsugu Tsuchiya[1,2,6,7,8 *], Marta I. Garrido[3,4,5,8 *],

[1]School of Psychological Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, 3168, Victoria, Australia

[2]Turner Institute for Brain and Mental Health, Monash University, Clayton, 3168, Victoria, Australia

[3]Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, 3010, Victoria, Australia

[4]Queensland Brain Institute, University of Queensland, St Lucia, 4072, Queensland, Australia

[5]Centre for Advanced Imaging, University of Queensland, St Lucia, 4072, Queensland, Australia

[6]Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Suita, Osaka 565-0871, Japan

[7]Advanced Telecommunications Research Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan.

[8]ARC Centre of Excellence for Integrative Brain Function, 770 Blackburn Rd, Clayton, 3800, Victoria, Australia

*: equal contribution

Corresponding author: Elise G. Rowe, Turner Institute for Brain and Mental Health, 770 Blackburn Road, Monash University, Melbourne, Victoria, Australia, elise.rowe@monash.edu

ABSTRACT

Detecting changes in the environment is fundamental for our survival. According to predictive coding theory, detecting these irregularities relies both on incoming sensory information and our top-down prior expectations (or internal generative models) about the world. Prediction errors (PEs), detectable in event-related potentials (ERPs), occur when there is a mismatch between the sensory input and our internal model (i.e., a surprise event). Many changes occurring in our environment are irrelevant for survival and may remain unseen. Such changes, even if subtle, can nevertheless be detected by the brain without emerging into consciousness. What remains unclear is how these changes are processed in the brain at the network level. Here, we used a visual oddball paradigm, in which participants engaged in a central letter task during electroencephalographic (EEG) recordings while presented with task-irrelevant high- or low-coherence background, random-dot motion. Critically, once in a while, the direction of the dots changed. After the EEG session, we confirmed that changes in motion direction at high- and low-coherence were visible and invisible, respectively, using psychophysical measurements. ERP analyses revealed that changes in motion direction elicited PE regardless of the visibility, but with distinct spatiotemporal patterns. To understand these responses, we applied Dynamic Causal Modelling (DCM) to the EEG data. Bayesian Model Averaging showed visible PE relied on a release from adaptation (repetition suppression) within bilateral MT+, whereas invisible PE relied on adaptation at bilateral V1 (and left MT+). Furthermore, while feedforward upregulation was present for invisible PE, the visible change PE also included downregulation of feedback between right MT+ to V1. Our findings reveal a complex interplay of modulation in the generative network models underlying visible and invisible motion changes.

Keywords: consciousness, prediction errors, DCM, vMMN, EEG

This is a provisional file, not the final typeset article

INTRODUCTION

Detecting changes in the environment is fundamental for our survival because these may indicate potential rewards or threats (LeDoux, 1996; Panksepp, 1998; Rensink, 2002). According to predictive coding theory, we detect these irregularities by comparing incoming (bottom-up) sensory information with an internal model based either on prior sensory information or on (top-down) beliefs or expectations (Friston, 2005; Friston & Stephan, 2007; Hohwy, 2013; Clark, 2013). Mismatch between the sensory input and our internal model results in a surprise event, signalled by a prediction error (PE) response in event-related potentials (ERPs). These PE responses represent the difference in the neural activation between expected (frequent or 'standard') and unexpected (rare/surprising or 'deviant') events and have been used extensively in the auditory modality to yield the mismatch negativity (MMN; Näätänen et al., 1978; Näätänen 1992; Garrido et al., 2007). Over the last decade, protocols for inducing visual PE (or visual MMN, 'vMMN', more specifically) have been developed, with a focus on situations where individuals are aware of a change (Stefanics et al., 2014; Pazo-Alvzrez et al., 2003; Kremláček et al., 2016). In our everyday lives, however, many changes in the sensory environment that evoke PEs may go unnoticed (Czigler, 2014; Stefanics et al., 2014). What remains unclear is how such sensory changes are processed by the brain at the network level.

One of the advantages in investigating non-conscious PE using visual, compared to auditory, stimuli is a richer set of psychophysical methods available to render otherwise salient visual stimuli invisible to participants (Faivre et al., 2017). Employing such tools, researchers have demonstrated PE responses to changes without conscious awareness of these changes. These PE responses in EEG include those using backward masking (Czigler et al., 2007; Kogai et al., 2011), binocular rivalry (Jack et al., 2015; 2017), rapid unmasked presentation (1 ms, Bernat et al., 2001, and 10 ms; Brazdil et al., 2001), and attentional blink (Berti, 2011). The researchers isolated PE ERPs by comparisons between standard and deviant trials, finding stronger and more widespread activation associated with conscious than non-

conscious visual PE. ERP analyses alone, however, cannot inform us of the underlying neurocircuitry underpinning the PE to changes that do, and do not, emerge into consciousness.

Dynamic Causal Modeling (DCM) is a technique useful for inferring the underlying causal network of dynamical systems (Friston, 2003). Boly et al., (2011), applied DCM to auditory PEs in EEG recorded from healthy participants and two populations of brain damaged patients: minimally conscious and vegetative state patients to show that feedback connectivity (or top-down modulation) was reduced in unconscious vegetative patients compared to minimally conscious patients (and healthy controls). This study suggests a potential involvement of feedback modulations in regulating the level of conscious awareness of PE generating stimuli. However, no studies have performed network level examination of non-conscious processing in fully awake healthy participants and whether this also abolishes top-down feedback connections.

To understand the differences in the neural circuitry between sensory changes that can and cannot be consciously perceived, we aimed to elicit visual PE using visible and invisible changes in motion direction. To achieve the desired level of visibility of motion stimuli, we manipulated the coherence of random-dot motion. We used DCM to estimate effective connectivity and examined the involvement of feedback connections for visible and invisible changes.

This is a provisional file, not the final typeset article

MATERIALS AND METHODS

Participants

Twenty-eight healthy university students participated in this study (10 females, aged 18-40, *M* = 22.44, *SD* = 4.30). All participants reported no history of neurological or psychiatric disorder or previous head trauma resulting in unconscious comatose states. All participants gave written informed consent in accordance with the guidelines of the University of Queensland's ethics committee.

Experimental design

Continuous EEG data were recorded during the main task using a Biosemi Active Two system with 64 Ag/AgCl scalp electrodes arranged according to the international 10-10 system. Data were recorded at a sampling rate of 1024 Hz. After the EEG was setup, participants practiced the 1-back letter task for 1 min (see below) before being tested in the main task. The main task was followed by psychophysical tests to assess the discriminability of motion stimuli (see below). We did not record EEG during these follow-up psychophysical tests. The entire experiment took under two hours, including the setup of the EEG.

Display

The experiment was performed using a Macbook Pro and external screen with a resolution of 1920 x 1080 pixels with a 60 Hz refresh rate. All participants were seated at a viewing distance of 50 cm (without a chinrest). The experiments were programmed using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) for MATLAB (version 2014b).

116

117 **Task and visual stimuli during EEG measurement**

118 To achieve the desired level of visibility and to induce visual PE to visible and invisible
119 changes, we manipulated the coherence of random-dot motion (Britten et al., 1992). In short, it is well-
120 known that the direction of motion can be consciously discriminated when the level of motion
121 coherence of random dots is sufficiently high. As the level of coherence approaches zero, participants
122 can no longer consciously see the direction of motion (Watanabe et al., 2011). We exploited this useful
123 feature of motion perception and designed a paradigm to elicit prediction errors from visible and
124 invisible changes in the motion direction of a cloud of random dots (Figure 1, demo:
125 https://figshare.com/s/76484519f510ba74891b). These motion stimuli were never task-relevant during
126 the main EEG experiment. Instead, participants were instructed to focus on a central 1-back task and
127 be as quick and accurate as possible.

128 Our motion stimuli comprised 750 white dots enclosed within a 30° diameter circular area on a
129 black background. During each 500 ms trial, each dot moved in a straight trajectory at 14°/sec. When
130 any dot went outside of the stimulus window, the dot was redrawn in the opposite location to keep the
131 density of the dots constant. At the first frame of each 500 ms trial, we removed all dots and redrew
132 them at new random locations at the same time (i.e. no blank interval between trials). Please refer to
133 the exemplar video provided for further clarity of the stimuli.

134 To elicit PE responses, we utilised a roving visual oddball paradigm (i.e. 5-8 repetitions before
135 a change) with dot motion at 5 possible non-cardinal directions (70°, 140°, 210°, 280° and 350°; with
136 0° representing motion to the right) and 2 possible coherence levels (high: 50% and low 5%). Motion
137 direction was manipulated in a 2 x 2 design comparing the *coherence* levels: high and low, and *motion*
138 *direction*: standard (frequent: no change) and deviant (rare/surprising: change of direction). Mean

number of low-coherence standard and deviant trials was 829 (SD = 65) and 136 (SD = 12), respectively. Mean number of high-coherence standard and deviant trials was 832 (SD = 64) and 136 (SD = 13), respectively.

One experimental block comprised 180 trials. At the beginning of the block, both the direction and the coherence of the motion was chosen randomly from the five possible directions and two coherence levels. The direction and coherence level remained unchanged for 5 to 8 trials before we changed the global motion direction randomly into one of the other 4 directions. Throughout the paper, we refer to the trials with unchanged motion direction as the standard trials and to the trials in which the direction changed as the deviant trials. Each deviant trial was then followed by another 5 to 8 trials without a change in motion direction, which we considered as the new standard trials. In some deviant trials, (once in every 26 to 30 trials), we also changed the coherence level. We did not include any of these 'double deviant' trials in our EEG or behavioural analysis.

At the centre of the visual display, we presented a 1-back task within a 5° diameter black circle. Here, a sequence of randomly selected white letters (~1.8^{o} visual angle in size) was presented every 450 ms (i.e. 150 ms on and 300 ms off). The first of these letter stimuli was presented at the onset of the first motion trial but became (mostly) desynchronized in time after this (except for every 9 motion trials where the onset times realigned, see Figure 1b). Participants were required to attend to this letter stream and press the spacebar as quickly as possible whenever they detected the repetition of any letter (which could occur every 10 to 15 letters; Figure 1b). On average, the number of letter repeat targets per participant across the experiment was 191 (range from 177 to 196, SD = 3.69).

Behavioral analysis on the 1-back task during the EEG measurement

We examined participants' accuracy during high- and low-coherence motion, defined as the hit and false alarm rates (and the sensitivity measure of d'; Green & Swets, 1966). For this, we defined a 'hit'

162 as a response made between 200 to 1,000 ms after the onset of a letter repeat. Responses outside of this
163 were considered a false alarm. Due to the fast presentation rate of our target events, we chose to use
164 the method described by Bendixen & Andersen (2012; 'Method E', pp. 930) to calculate our false
165 alarm rate. Here, the number of non-target events (the denominator for calculating the false alarm rate)
166 were defined as the duration of the experiment divided by the average time it would take to execute a
167 response (i.e. 300 ms) minus the number of target events. In using this method, we aimed to overcome
168 the bias in using traditional signal detection theory methods in paradigms with high event rates. Two
169 participants failed to correctly detect above 50% of the targets across the entire experiment. We
170 removed these participants from further analyses.

171 ********** **FIGURE 1 ABOUT HERE** **********

172 ***Follow-up psychophysics tasks***

173 To ensure the visibility and invisibility of high- and low-coherence motion direction changes,
174 respectively, we tested each participant's performance with two follow-up psychophysics tasks. The
175 first task required participants to make a direction discrimination at high- or low-coherence levels. The
176 second psychophysics task, performed in a subset of participants (N=8 out of 28 participants), required
177 them to report when they 'felt' or 'sensed' a change in direction occurred at both high- or low-
178 coherence levels.

179 ***Follow-up psychophysics task 1: direction discrimination task***

180 To estimate the discriminability of the direction of motion, we ran a four alternative-forced choice
181 (4AFC) direction discrimination task based on a study design with similar stimuli to ours (Tsushima
182 et al., 2006). To conservatively estimate the discriminability, we reduced the direction alternatives to
183 4 possibilities (80°, 160°, 240° and 320°), which were different from the motion directions used in the

184 main EEG experiment in order to reduce the chance of perceptual learning which can occur even to
185 sub-threshold dot motion (Watanabe et al., 2011; Tsushima et al., 2006). Furthermore, we aimed to
186 avoid any habituation effects that may have arisen if we had used the same motion directions in both
187 experiments. These effects, in turn, could have improved or reduced task performance. Rather than
188 inducing these uncertain results, we opted for slightly different stimuli as an alternative solution. In
189 each trial, we presented the motion for 915 ms without the central letter task (i.e. the small, black
190 central circle remained but the letters were absent). We selected this extended presentation time based
191 on the study by Tsushima et al., (2006) and in order to conservatively estimate the motion direction
192 discriminability. We did not redraw all the dots after 500 ms as in the EEG experiment but, instead,
193 the dots remained moving for 915 ms (except when it came to the boundary, see above). We randomly
194 selected a motion direction in each trial and pseudo-randomly intermixed 4 coherence levels (2.5%,
195 5%, 25% and 50%) in equal proportions across 120 trials. We incorporated two additional motion
196 coherence levels to reduce the chance of implicit learning based on the motion coherence level. At the
197 end of every trial, participants reported the perceived direction of motion from the 4 possible
198 alternatives.

199 **Behavioral results psychophysics task 1: *Confirming visibility and invisibility of high- and low-***
200 ***coherence motion direction changes in each individual***

201 We used performance in our follow-up psychophysics direction discrimination task to confirm that
202 high- and low-coherence motion direction changes were visible and invisible, respectively. Based on
203 the performance, we excluded participants from the following EEG analysis based on two criteria: 1)
204 performing above chance at low (5%) coherence motion condition or 2) performing below chance at
205 high (50%) coherence motion condition (Figure 2a and 2b). To estimate if the observed discrimination
206 accuracy was above chance, we performed a bootstrap analysis (with replacement) over 1,000
207 repetitions per participant. For each participant, at a given coherence level, we generated a surrogate

sequence of correct vs incorrect discriminations across 30 trials (e.g. correct, incorrect, correct, correct, correct....) which reflected the proportion of correct responses at the number of trials we tested. We then used 2.5%ile and 97.5%ile of the bootstrapped distribution as the confidence interval. That is, for the low (5%) coherence condition, if the bottom end of the confidence interval was above chance (25% correct), we assumed that the participant may have been aware of the direction of motion during the EEG experiment (Figure 2b).

Based on our bootstrapping analyses, we confirmed the visibility and invisibility of the high- and low-coherence motion direction changes, and excluded two participants due to criterion 1 and one due to criterion 2. Due to technical error (corrupted data files), we could not perform bootstrapping analysis for four more participants and we conservatively chose to remove the EEG data of these four participants from further analysis as well.

In summary, we used the data from N=19 participants for ERP, source, and DCM analysis.

********** **FIGURE 2 ABOUT HERE** **********

Follow-up psychophysics task 2: Detectability of motion direction changes

Finally, as another follow-up experiment, we ran a detection task to exclude the possibility that participants were able to 'sense' the change of motion direction even in the absence of the perception of motion direction per se, but based on the metacognitive 'feeling of change' (Rensink, 2004). To test this, we repeated one block from the main experiment; after removing the letters for the 1-back task but leaving the small, black central circle (Figure 1). The direction of motion and coherence levels were the same as the main experiment. We instructed participants to press the spacebar as quickly as possible whenever they detected a change in the global motion direction of the cloud of dots. Participants were instructed that even if they were unsure, they should report a feeling of change. We

This is a provisional file, not the final typeset article

regarded a change of global motion direction within a coherence level (which was either 5% or 50%) as a target event. There were 20 events per coherence level. To analyze this behavioral data, we computed the sensitivity measure of d' using the same signal detection theory method as our main experiment.

Behavioral results psychophysics task 2: *Confirming lack of 'feeling' of change*

We found that d' was significantly above zero for high-coherence motion direction changes (Figure 2c; one-sided one sample t-test, $M = 2.77$, $SD = 1.68$, $p > 0.001$, $df = 7$) but not for low-coherence motion direction changes (one-sided one sample t-test, $M = -0.47$, $SD = 1.60$, $p = 0.781$, $df = 7$). We took this as evidence that this subset of participants could not even sense direction changes at low-coherence, while they were able to do so at high-coherence.

EEG preprocessing

We used SPM12 (http://www.fil.ion.ucl.ac.uk/spm/) to pre-process the data. We first re-referenced the raw EEG recordings to the average of all electrodes, down-sampled to 200 Hz and high-pass filtered above 0.5 Hz. We epoched within a trial time window of 100 ms before to 400 ms after the onset of each motion trial. We then removed eyeblink artefacts using the VEOG channels with a bad channel maximum rejection threshold of 20% and by thresholding all channels at 100uV. In the following EEG analysis, we present data without the removal of trials in which a target for the 1-back task occurred or a button response was made. Given that the task and button presses were uncorrelated with the conditions of interest (i.e. motion trials), these are unlikely to have an effect on our results (which we confirmed; data not shown). To obtain the mean ERP per participant per condition, we averaged the epoched data using robust averaging (Wager et al., 2005) with a built-in function in SPM12. This robust averaging process down-weighted each time-point within a trial according to how different it was from the median across trials. Next, because high frequency noise can be introduced during the robust

253 averaging process, we further low-pass filtered the processed data at 40 Hz. Finally, we baseline
254 corrected the data for each participant and condition by subtracting the (robustly averaged) ERP
255 between -100 to 0 ms.

256 **ERP analysis**

257 We analysed ERPs corresponding to each of the experimental conditions across participants (as well
258 as the visual PE, in the form of the 'vMMN': derived from subtracting the standard waveform from
259 the deviant waveform within a condition). We combined channel clusters over left anterior (FC3, F1,
260 F3, F5, AF3), right anterior (FC4, F2, F4, F6, AF4), left central (C3, C5, CP1, CP3, CP5), right central
261 (C4, C6, CP2, CP4, CP6), left posterior (P7, PO3, PO7, PO9, O1) and right posterior (P8, PO4, PO8,
262 PO10, O2) electrodes (see Jack et al., 2015). The significant vMMN time periods were those where
263 the difference waves mean amplitude was below zero across participants (after FDR correction for
264 multiple comparisons at each time point).

265 **Spatio-temporal statistical maps**

266 We obtained spatiotemporal images of the ERP for each participant and condition across the scalp
267 within the window of -100 to 400 ms. To obtain one 2D image, for each of the 101 time bins, we
268 projected the EEG electrode locations onto a plane and interpolated the electrode locations linearly
269 onto the 32 x 32 pixel grid. We then stacked each 2D image over time to obtain a 3D volume (32 x 32
270 x 101) and smoothed with a 3D Gaussian kernel of full-width at half maximum: 12 mm x12 mm x 20
271 ms.

272 The 3D spatio-temporal image volumes were analysed, on a participant-by-participant basis, with a
273 mass univariate general linear model (GLM) as implemented in SPM12. We estimated the main effects
274 of surprise (i.e., standards vs deviants) and coherence (i.e., high- vs low-coherence), their interaction,

This is a provisional file, not the final typeset article

and contrasts between standards vs deviants separately within the high- or low-coherence conditions using between-subject F-contrasts. All spatio-temporal effects are reported at a threshold of $P<0.05$ at the cluster-level with family-wise error (FWE) correction for multiple comparisons over the whole spatio-temporal volume. F-statistics are reported as the maximum at the peak-level.

Source reconstruction

We obtained source estimates on the cortical mesh by reconstructing scalp activity with a single-sphere boundary element method head model, and inverting a forward model with multiple sparse priors (MSP) assumptions for the variance components under group constraints (Friston et al., 2008). One benefit of using MSP for the source reconstruction process is that through the use of empirical Bayes one can select multiple sources to build covariance components for both sparse or distributed source configurations (Friston et al., 2008). Because our main effects of coherence and surprise (and their interaction) spanned the entire epoch we decided to consider the whole peristimulus time window (0 to 400 ms) in the MSP procedure. This allowed for inferences on the most likely cortical regions that generated the sensor-level data across the entire trial time window. We obtained volumes from these reconstructions for each of the four conditions for every participant. These images were smoothed at full-width at half maximum: 12x12x12 mm^3. We then computed the main effects of (1) coherence, (2) surprise, (3) the interaction (coherence x surprise), as well as the (4) high- and (5) low-coherence PE (standards vs. deviants) using conventional SPM analysis between-subject F-contrasts.

Dynamic Causal Modelling

We used DCM to estimate a generative causal model that most parsimoniously explained the observed ERPs at the selected source locations with minimal model complexity (Friston, 2003). It is important to note, here, that DCM is a statistical inference of causal models, and 'causality' is not established through perturbation or other intervention (Pearl, 2000). We used a data-driven approach combined

298 with a priori locations drawn from the visual motion processing literature, to explain best the observed
299 PE related signals for visible and invisible motion changes. The data-driven spatial location of left
300 inferior temporal gyrus (ITG) was selected based on our source-level analysis. We also included the
301 sensory input nodes of bilateral primary visual cortices (V1) and middle temporal cortex (MT+), which
302 are likely to be the initial cortical stages for visual motion processing (Born & Bradley, 2005), and the
303 bilateral posterior parietal cortices (PPC), which are known for higher-level visual motion processing
304 (Anderson, 1989; Ilg et al., 2004). In total, we assumed seven sources: bilateral V1, MT+, PPC and left
305 ITG (see Results for Montreal Neurological Institute (MNI) coordinates).

306 We connected our candidate nodes using the same architecture and then exhaustively tested all
307 possible combinations for the direction of modulation(s) amongst these nodes (with one exception, see
308 below). Our model architecture comprised: (1) recurrent (forwards and backwards) connections from
309 V1 to MT+ and MT+ to PPC within the left and right hemispheres, and left MT+ to left ITG, (2) each
310 node was laterally connected with the corresponding node in the other hemisphere (e.g. left V1 laterally
311 connected to right V1), (3) intrinsic (or within-region) modulation only at V1 and MT+, which are
312 known to strongly adapt to repeated visual motion in the same direction (Kohn & Movshon, 2004), and
313 (4) left ITG and left PPC laterally connected. We decided to assign the equal hierarchical relationship
314 between ITG and PPC due to the inconsistent relationship between them reported in the literature (e.g.,
315 DeYoe & Van Essen, 1988; Felleman & Van Essen, 1991).

316 Next, although there could be a huge number of possible DCM modulations based on our 7
317 identified nodes, we decided to reduce the possible space for modulations based on anatomical
318 information as much as possible. By anatomical criteria, we decided to examine models that always
319 contained recurrent (forwards and backwards) modulation between the input sources of bilateral V1
320 and MT+ and intrinsic modulations at these input nodes; we refer to this as the 'minimal' model (i.e. a
321 base model for all other models). Using this reduced number of potential modulation directions, we

This is a provisional file, not the final typeset article

322 were left with 7 connections that could be modulated beyond the minimal model: 4 connections
323 between MT+ and PPC (forwards and backwards in each hemisphere), 2 connections between MT+
324 and IT (forwards and backwards in left hemisphere) and 1 connection between PPC and IT (one lateral
325 connection in left hemisphere). Note, that we never modulated the lateral connections between the
326 hemispheres. Thus, in total, we tested 129 models (i.e. $2^7 = 128$ + a null model with no modulation at
327 or between any nodes) comprising all combinations of modulation directions (see Figure 7a).

328 We performed DCM analyses to estimate the parameters of each model, separately for each
329 effect of interest: (1) the visible PE and (2) the invisible PE. For the visible PE, we used the between-
330 trial effect (condition weights) of [0, 1] for the high-coherence standard and deviant trials, respectively.
331 For the invisible PE, we used the weights of [0, 1] for the low-coherence standard and deviant trials,
332 respectively. DCM analyses started from estimation of the log model evidence for each of our 129
333 generative models fitted for every participant. The log model evidence quantified how well each model
334 could generate ERPs similar to the observed (pre-processed) ERPs for that participant. Importantly,
335 the iterative free-energy minimization process used to calculate the log model evidence penalised
336 models with greater complexity to avoid overfitting (Friston et al., 2008). Once estimated, we used a
337 random-effects (RFX) approach to determine the winning model across participants via Bayesian
338 Model Selection. RFX assumes that a (potentially) different model underpins each participant's
339 responses; making it robust to outliers and best suited for studying perceptual processes whose
340 underlying network structure is likely to be varied across participants (Stephan et al., 2009). To
341 compute a weighted average of the parameter estimates across all models, we employed Bayesian
342 Model Averaging (BMA) (Penny et al., 2010). BMA weights the estimated parameter with the
343 probability of each model associated with that parameter for all participants. In this way, all models
344 contributed to the final connectivity estimate, with the most probable model having the greatest weight
345 and the least probable model contributing the least to the final estimates (Stephan et al., 2010). In a

346 follow-up DCM analyses, we compared the visible and invisible PE by modelling their interaction
347 using the between-trial effect (condition weights) of [0,1] to contrast the mean ERPs for the visible PE
348 and invisible PE per participant.

349 ********** **FIGURE 3 ABOUT HERE** **********

350 **RESULTS**

351 ***No difference in distraction by visible or invisible motion directions***

352 After exclusion based on the follow-up psychophysics task, we analysed the performance of the
353 remaining N = 19 participants in the 1-back task to check if they appropriately focused on the central
354 letter task and ignored the background motion during the EEG session of the main task. Based on prior
355 studies (e.g., Tsushima et al., 2006), we expected that our low-coherence background motion would
356 distract participants more than high-coherence motion to degrade the performance of the 1-back task.

357 Between the high- and low-coherence conditions, we observed a difference in the percentage
358 of hits and false alarms but no difference in d' or reaction times. The proportion of hits during the high-
359 coherence condition (M = 69.54%, SD = 11.68%) was significantly lower than in the low-coherence
360 condition (M = 73.87%, SD = 12.34%; two-tailed paired t-test, $t(18) = -3.27$, $p = 0.004$; Figure 3a).
361 Mean number of false alarms across participants was also significantly lower in the high-coherence
362 condition (M = 4.95, SD = 4.71) than the low-coherence condition (M = 6.95, SD = 5.34, two-tailed
363 paired t-test, $t(18) = -2.39$, $p = 0.028$; Figure 3b). Combining the proportion of hits and false alarms,
364 we computed the d' measure. Using this measure, we found no significant differences between d' for
365 performance on the 1-back task during the high- (M = 3.32, SD = 0.49) and low-coherence conditions
366 (M = 3.31, SD = 0.49, two-tailed paired t-test, $t(18) = 0.09$, p = 0.931) (Figure 3c). Furthermore, we
367 found no differences in the reaction times between the high (M = 557 ms, SD = 59 ms) and low (M =

This is a provisional file, not the final typeset article

555 ms, SD = 64 ms, two-tailed paired t-test, $t(18) = 0.722$, $p = 0.48$) coherence conditions (Figure 3d).

Overall, we did not find the results to be consistent with our expectation of greater distraction effects during low-coherence motion trials. We suggest this may be due to the differences between our study and that by Tsushima et al. (2006; such as the different behavioural task and the inclusion of the PE generating stimulus sequence), leading to lower task difficulty and comparable distractibility effects between the two coherence levels. We interpret these results as: (1) participants were able to maintain their focal attention to the 1-back task, (2) participants did not trade off speed for accuracy differently between the two coherence conditions and (3) that the behavioral effects of high- or low-coherence motion were comparable.

********** **FIGURE 4 ABOUT HERE** **********

Scalp-level ERP analysis

Non-conscious prediction errors occur earlier than conscious prediction errors

We extracted the ERPs from scalp-level electrode clusters to examine the vMMN for visible and invisible PE responses (defined as sustained negativity of vMMN difference waves after correcting for multiple comparisons at each time point, $q < 0.05$). We found that vMMN was evoked from both visible and invisible motion direction deviants. For visible PE responses (Figure 4a and 4b), we observed vMMN ($p_{FDR} < 0.05$) at left central and anterior channels clusters between 285-295 and 275-400 ms, respectively. For the invisible PE responses (Figure 4c), we observed vMMN at left posterior channels at 205 ms.

In addition to examining ERPs at the scalp, we used spatio-temporal statistical map analysis in SPM (see Methods). Here, we obtained the 3D images interpolated from ERP data recorded at the scalp and applied between-subject F-tests to quantify the effect of surprise (i.e., prediction error (PE)) within the high- or low-coherence conditions (i.e. visible and invisible PE). Importantly, we found that PEs were evoked from both visible and invisible motion direction deviants. PE to visible motion direction changes (Figure 5a) disclosed a number of significant clusters, ranging from 290 – 395 ms observed across widespread channels. The earliest cluster at 290 ms was found at the central channels (peak-level F_{max} = 25.44, cluster-level p_{FWE} = 0.024) followed by a cluster at left front-temporal and central channels at 380 ms (peak-level F_{max} = 34.12, cluster-level p_{FWE} < 0.001) and at 395 ms (peak-level F_{max} = 67.29, cluster-level p_{FWE} < 0.001). Compared to visible PE, invisible PE occurred earlier and were less spatially spread (Figure 4b); only at 160 ms in left parietal channels (peak-level F_{max} = 25.65, cluster-level p_{FWE} = 0.024).

Next, we applied between-subject F-tests to quantify the main effects of coherence (Figure 5c), surprise (Figure 5d) and their interaction (Figure 5e). The main effect of coherence (Figure 5c) disclosed three significant clusters. The first cluster at 160 ms was located occipitally (peak-level F_{max} = 34.37, cluster-level p_{FWE} = 0.002), the second at 185 ms occurred in the same location (peak-level F_{max} =35.72, cluster-level p_{FWE} = 0.004) and the third at 295 ms was found at right occipito-parietal and frontal channels (peak- level F_{max} = 32.31, cluster-level p_{FWE} = 0.001). The main effect of surprise (Figure 5d) showed three significant clusters. The first at 80 ms occurred at left parietal channels (peak-level F_{max} = 29.53, cluster- level p_{FWE} = 0.014), the second at 285 ms was located at right central channels (peak-level F_{max} =34.66, cluster-level p_{FWE} < 0.001) and the third at 375 ms was observed in the same location (peak-level F_{max} = 55.29, cluster-level p_{FWE} < 0.001). Finally, we observed an interaction between surprise and coherence (Figure 5e) at central and frontal channels at 395 ms (peak-level F_{max} = 35.93, cluster-level p_{FWE} = 0.002).

This is a provisional file, not the final typeset article

413 ********** **FIGURE 5 ABOUT HERE** **********

414 **Source-level analysis**

415 ***Left ITG as a source for conscious PE***

416 We applied MSP source reconstruction to estimate the cortical regions involved in generating PE to
417 visible and invisible motion direction changes. In Figure 5 (the scalp-level maps), we found that the
418 main effects of surprise, coherence, and the interaction spanned the whole epoch, thus we decided to
419 use the whole epoch data (0-400 ms) rather than to temporally constrain the data for source
420 reconstruction (see Methods for details). Figure 6 shows the significant source-level results for the
421 main effect of surprise and PE to visible changes at an uncorrected threshold of $p<0.001$. We did not
422 find any significant sources for invisible PE, main effects of coherence or interaction when we
423 corrected for multiple comparisons. For the main effect of surprise, we found one significant cluster in
424 the left ITG ([-52 -26 -30, peak-level $F_{max} = 22.19$, cluster-level $p_{FWE} = 0.003$, Figure 6a). For the PE
425 to visible changes, we found a similar cluster in the left ITG ([-48, -12, -32], peak-level $F_{max} = 17.91$,
426 cluster-level $p_{FWE} = 0.039$, Figure 6b). The effect of surprise for the invisible conditions revealed a
427 cluster on the right hemisphere (Figure 6a) which did not survive correction.

428 ********** **FIGURE 6 ABOUT HERE** **********

429 **Dynamic Causal Modelling**

430 We used DCM to examine how the source location identified in the previous step, interacted with
431 cortical locations known to specialize in motion processing, and how the strengths of these interactions
432 are modulated by the visibility of motion changes. Specifically, we identified one region at the FWE
433 corrected ($p<0.05$) threshold (between-subject F-tests) (Anatomy Toolbox; Eickhoff et al., 2005): the
434 left **ITG** (MNI coordinate: [-48, -12, -32], Figure 6). We included the sensory input nodes of bilateral

435 **V1** (MNI coordinates: left [-14, -100, 7] and right [17, -97, 9]) and bilateral **MT+/V5** (MNI
436 coordinates: left [-48, -69, 7] and right [50, -66, 11]) as these regions are essential for visual motion
437 processing (Born & Bradley, 2005; Plomp et al., 2015). We also included the bilateral **PPC** (MNI
438 coordinates: left [-46, -46, 54] and right [52, -42, 50]) because these regions are known for higher-level
439 visual motion processing (Ilg et al., 2004). Based on our 7 identified nodes, we tested 129 models
440 comprising all combinations of modulation directions (building on the fixed 'minimal model'; see
441 Methods for more information). Figure 7a shows the defined model space, including: the fixed model
442 architecture (white arrows) and how each model was allowed to vary in terms of the direction of
443 modulation (black arrows).

444 ********** **FIGURE 7 ABOUT HERE** **********

445 **Optimized DCM modulation parameters explaining differential ERPs for visible and invisible**
446 **PE**

447 Once the log model evidence for every model and participant was estimated, we used RFX Bayesian
448 Model Selection to determine the winning model at the group-level. Figure 7b shows, for both the
449 visible and invisible PE, the exceedance probability of all models at the group-level. The exceedance
450 probability is the probability that a particular model is more likely than any other model given the
451 group data.

452 For the visible PE, there was no single clear winning model amongst our 129 models. This is
453 unsurprising when using RFX analysis as effects can be diluted if a large number of models are
454 compared in a small sample size. In this situation, it is recommended not to single out the best
455 architecture alone, but rather to use BMA to consider all models, taking into account the strength of
456 the evidence for each model. After applying BMA, we found significant positive modulation of the
457 intrinsic self-connections of left MT+ (mean parameter outputs = +0.2004, two-tailed t-test against 0:

This is a provisional file, not the final typeset article

pFDR= 0.0486, *df* = 18), and right MT+ (+0.2644, *pFDR* = 0.0011, *df* = 18). This result confirms what we expected from the nature of our oddball paradigm; the visible change induced PE relies, in part, on intrinsic (self) modulation at MT+ due to a release from adaptation (i.e. repetition suppression) following the onset of deviant motion. In addition to this, we also found two more significantly modulated connections. We found a *backward* connection from right MT+ to right V1 was significantly downregulated (-0.1310, *pFDR* = 0.0453, *df* = 18) and a *forward* connection from left MT+ to left ITG was significantly positively modulated (+0.0875, *pFDR* = 0.0060, *df* = 18).

For the invisible PE, as with our visible PE DCM analyses, in the absence of one single winning model, we again applied BMA finding that, similarly to the visible PE DCM results we observed significant positive modulations for intrinsic modulations within left V1 (+0.1501, *pFDR* < 0.0001, *df* = 18), right V1 (+0.1400, *pFDR* = 0.0081, *df* = 18) and left MT+ (+0.2821, *p FDR* < 0.0001, *df* = 18), reflecting a release from adaptation upon a change in motion direction, despite no awareness of this change occurring. One difference, compared to the visible PE, was the direction of one other significantly positively modulated connection *forwards* from right MT+ to PPC (+0.2048, *pFDR* = 0.0094, *df* = 18). We found no significantly modulated backward connections within our DCM analyses of the invisible PE.

Finally, we directly compared differences in effective connectivity between the visible and invisible PE by modelling their interaction. We applied RFX BMA to determine the model parameters in the absence of one single winning model and found two significant connections (prior to FDR correction) between left ITG and left PPC and at a self-connection of left MT+. However, these connections did not survive correction for multiple comparisons. We interpret these findings as weak but consistent with our source results showing left-hemisphere localisation of the visible PE.

DISCUSSION

In this study, we aimed to elicit PE related neural activity via changes in stimulus statistics that were consciously visible or not. As planned, we successfully manipulated the awareness of the stimulus changes using supra- and sub-threshold motion coherence of a dynamic random-dot display and confirmed the visibility of the motion direction changes with a follow-up psychophysics task. Our ERP analyses (Figure 4 and 5) confirmed robust visible PE responses, which were widespread in space and time and invisible PE responses, which were more confined in space and time. Our source level analysis located the source of the visible PE response in the left ITG. Using the identified location of the left ITG and the cortical nodes known to be critically involved in motion processing (V1, MT+, and PPC), we performed DCM analyses, hoping to reveal the network underpinning visible and invisible PE responses.

Confirmation of motion direction (in)visibility

It is important that we first acknowledge that our method of rendering our stimuli invisible (or weaker stimulus paradigms in general, e.g. Dehaene, 1999; Dehaene & Naccache, 2001; Gaillard et al., 2009, Del Cul et al., 2007, van Vugt et al., 2018) confounds the issue of conscious perception with the issue of stimulus change. We also point out that alternative approaches which eschew this issue (such as binocular-rivalry like paradigms), do suffer from other issues such as the effects of the report. That is, when the physical stimulus input is equated, the effects of reports can be falsely interpreted as the neural correlates of consciousness (Frässle et al., 2014, Tsuchiya et al., 2015, Wilke et al., 2009). Thus, we foresee the combination of both approaches (Tsuchiya et al., 2016) will be important for future studies.

Secondly, we must highlight the factors leading to the design of our follow-up psychophysics task 1 that was used to confirm (in)visibility of motion direction at high- and low-coherence. The

design of this task was motivated by a number of concerns with presenting motion stimuli identical to that used in the main experiment. We opted to change the motion parameters during this task primarily because of our concern on perceptual learning and habituation during the visibility check. To alleviate this issue, we made four critical changes in the stimulus design: we lengthened the visual motion presentation time, removed the distracting central letter task, intermixed 4 coherence levels, and reduced the direction alternatives. All of these changes, in addition to the fact that during the main experiment participants did not pay attention to the motion, should have made our visibility assessment rather conservative; that is, we should have detected any participants who were aware of the low-coherence visual motion direction changes during the main task using this visibility task. Furthermore, while our choice of 30 trials was on the smaller end in terms of the number trials as a visibility check, we performed non-parametric bootstrapping analysis on an individual basis, which is the most sensitive method to detect aware participants. Overall, we are confident in our method of confirming that the motion direction changes were visible or invisible.

Temporal features of the prediction error (PE) responses with visible and invisible motion changes

We were able to elicit PE responses for both visible and invisible motion direction changes. Scalp ERPs (Figure 4) disclosed a significant early component at 150 ms for invisible PE (i.e. the low-coherence vMMN difference wave) at left posterior channels. In contrast, we observed two significant later components for the visible PE (high-coherence vMMN difference waves) between 285-295 and 275-400 ms at left central and left anterior clusters, respectively. Our spatio-temporal scalp-level statistical mapping analysis (Figure 5) supported these findings. Here, after FWE correction, PE evoked by invisible motion direction changes peaked at parietal channels earlier in time (<160 ms) than PE to visible motion direction changes (>290 ms), which were observed at central and fronto-temporal channels.

530 Our finding is consistent with some studies that elicited non-conscious visual PE. In an MEG
531 study using vertical gratings and backward masking of rapidly presented deviants, Kogai et al. (2011)
532 found similar latencies in the non-conscious PE response from 143-154 ms in striate regions. Further,
533 Czigler et al., (2007), backward masked coloured checkerboards at different stimulus onset
534 asynchronies (SOAs) and found mask SOAs at 40 and 53 ms elicited PE between 124-132 ms at
535 occipital channels (but that participants had some level of awareness of the deviant stimuli). Other
536 studies, however, report later PE to invisible oddballs. For example, Jack and colleagues (2017) found
537 slightly later non-conscious PE (to deviant stimuli presented monocularly to the non-dominant eye)
538 that peaked around 250 ms (as did the conscious PE). This longer latency for the non-conscious PE
539 was also echoed in another study from the same group (Jack et al., 2015). We suggest the discrepancies
540 in the latencies arise from both stimulus differences and in the methods used to render stimuli non-
541 conscious. Future paradigms in which both the standard and deviant stimuli can be rendered invisible
542 may help disentangle conscious and non-conscious PE effects.

543 But why does the invisible PE occur earlier than visible PE? Here we offer two possible
544 explanations. The first is related to the effects of attention. Previous visual PE studies that have used
545 consciously *perceivable motion direction* changes found an earlier PE at 142-198 ms regardless of the
546 attentional load on the irrelevant task (Krëmlacek et al., 2013). A separate study found a later PE
547 component emerged when attention was directed to the PE generating stimuli (Kuldkepp et al., 2013).
548 Indeed, in a binocular rivalry vMMN study (using grating stimuli), van Rhijn et al. (2013) observed
549 two early components from 140 to about 220 ms both when attention was directed to the rivalry stimuli
550 and when attention was diverted. But a second late negativity (270 to 290 ms) was observed only when
551 attention was directed to the rival stimuli. In our paradigm, participants may have attended to the visible
552 motion resulting in a delayed PE. Conversely, invisible motion may not have deployed attention, which
553 resulted in an earlier PE. However, this explanation does not speak to the neural mechanisms. An

This is a provisional file, not the final typeset article

554 alternative explanation lies on a putative difference in adaptation (or repetition suppression) for visible
555 and invisible stimuli. It is plausible that stronger adaptation in MT+ to the visible stimuli may mask
556 the earlier PE component in V1. This adaptation effect may be weaker in MT+ for invisible stimuli,
557 although this interpretation does not fully concur with our DCM results showing adaptation in V1 (but
558 not right MT+) for invisible stimuli (see below for further discussion). Additionally, the later PE
559 component for visible stimuli may be more related to the awareness of prediction violation. Further
560 studies that orthogonally manipulate adaptation and prediction might be able to distinguish between
561 these possibilities (Kok et al., 2012; Summerfield & Egner, 2009; Summerfield et al., 2006) and
562 disentangle the two components underlying the ('classical') vMMN; namely, the 'genuine' vMMN PE
563 response and the effects of neural adaptation (O'Shea, 2015).

564 **Spatial features of the prediction error (PE) responses with visible motion changes**

565 In terms of spatial characteristics of visible PE, and based on previous studies by Tsushima et al.,
566 (2006), we expected to detect prefrontal source activity to our supra-threshold visible motion stimuli.
567 Instead, our source-level analyses (which pooled responses over time) provided evidence that visible
568 PE, as well as the main effect of coherence (regardless of the level of surprise), were generated by
569 cortical sources within the left ITG. We suggest these source-level differences might have arisen from
570 differences, most likely, in our measurement modalities (fMRI in Tsushima et al, 2006, and EEG in
571 our study), as well as differences in experimental paradigms, namely, our inclusion of the roving
572 oddball design instead of coherence differing on a trial-by-trial basis and our 1-back task.
573

574 According to our literature search, we note one study of visual consciousness that supports our
575 findings of left hemispheric lateralization (O'Shea et al., 2013). In this study, the authors used binocular
576 rivalry to seek brain activity that could predict visual consciousness and found early activity (around

577 180 ms) confined to left parietal-occipital-temporal regions. Beyond this, literature on the lateralization
578 of conscious awareness of specific stimuli to the left hemisphere is scarce; reported only in a handful
579 of studies of emotional processing (Gazzaniga, 2000; Kimura et al., 2004; Meneguzzo et al., 2014;
580 Prete et al., 2015; Shepman et al., 2016; Williams et al., 2006). In these studies, subliminally presented
581 face stimuli activated right amygdala (via a subcortical route) in response to fear but left amygdala for
582 supraliminal fear (Williams et al., 2006), and using masked or unattended affective-stimuli activated
583 the right hemisphere in both the visual (Kimura et al., 2004) and auditory domains (Shepman et al.,
584 2016). However, unlike these studies, we used visual motion stimuli and examined the difference
585 between PE rather than responses to visual stimuli more generally. Another possible explanation for
586 the observed lateralization of conscious perception of motion changes to left hemisphere comes from
587 work in split-brain patients showing the left hemisphere is more adept at monitoring probabilities to
588 infer causal relationships based on series of events over time (Gazzaniga 2000; Roser et al., 2005) and
589 at creating internal models to predict future events (Wolford et al., 2000). This suggests that when
590 stimuli are consciously perceivable, the left ITG generates and updates predictions and PE. However,
591 due to the lack of literature, specifically on visual PE lateralization, further work is needed to fully
592 understand the role of left ITG source for conscious PE processing.

593 **Network level model of causal connections investigated by DCM**

594 Using DCM, we aimed to extend earlier visual PE studies by investigating the network level properties
595 underlying the generation of PE to visible and invisible changes. The primary question of our DCM
596 analysis was whether there was evidence for top-down modulations for PE to visible or invisible
597 change. According to the predictive coding framework, when an unexpected stimulus occurs, the PE
598 signal is propagated from lower to higher brain areas, resulting in upregulation of forward connectivity.
599 This, in turn, is followed by the revised prediction from high- to low-level brain areas, resulting in
600 increased feedback connectivity. Previous studies are consistent with this theory that consciously

This is a provisional file, not the final typeset article

perceived PE lead to increases in both feedforward and backward connectivity (e.g., Boly et al., 2011). Our finding of increased forward connectivity from left MT+ to ITG for visible PE is consistent with the first part of the theoretical prediction. What is puzzling is that conscious PE was accompanied by significant decreases in top-down feedback connectivity from right MT+ to right V1 (Figure 7b). This means that the neural prediction from MT+ to V1 decreased when the motion direction was unexpected, which appears inconsistent with the general framework of predictive coding. One possible explanation is that when the prediction is violated, the system suspends prediction, corresponding to the down regulated prediction from the high- to low-level area. Invisible PE, on the other hand, only induced enhanced feedforward connectivity from right MT+ to PPC.

Common to both the visible and invisible PE were significant modulations of the self-connections within lower-level visual areas of V1 and MT+. This suggests that both types of PE rely (in part) on a release from adaptation (i.e. repetition suppression) upon deviant motion onset (i.e. change in stimulus statistics). Whether the adaptation effects were observed at V1 or MT+ depended upon whether this change was consciously perceived. That is, visible change PE relied on adaptation at bilateral MT+, whilst invisible change PE relied on adaptation at bilateral V1 (and left MT+). One possible explanation for the stronger effects observed for visible PE at MT+ than V1 could be related to the subcortical visual motion pathway that carries visual information directly from the lateral geniculate nucleus to MT+ (Sincich et al., 2004). It is possible that when the motion signal is more coherent (i.e. stronger), this subcortical pathway is also activated (in addition to that between V1 and MT+), leading to a stronger motion signal in MT+. Subsequently, if MT+ is more highly activated compared to V1, this may lead to greater adaptation effects upon deviant motion presentation. Alternatively, when the motion signal is weak (i.e. low-coherence), this subcortical route is not activated, and thus, V1 and MT+ may have comparable levels of adaptation (as observed in our invisible PE DCMs). Findings that the generators of visual PE to motion changes are located in motion

625 centres or the dorsal pathway itself have been suggested previously by other studies of visible motion
626 direction changes (Kremlacek et al., 2006; Pazo-Lavarez et al., 2004). We add to these findings by
627 showing that this still holds when the changes are invisible.

628

629 **CONCLUSION**

630 We provide new insights into the brain mechanisms underpinning visual change detection, even in the
631 absence of awareness, when task reporting is not required. We lend support for visual PE in response
632 to both consciously and non-consciously perceivable changes; with the former evidenced as stronger
633 and more widespread cortical activity. Our findings suggest hemispheric lateralization within the left
634 hemisphere when motion changes were visible. Using DCM, we found that both types of PE were
635 generated via a release from adaptation in sensory areas responsible for visual motion processing. The
636 overall pattern emerging from our study reveals a complex picture of down- and up- regulation of
637 feedforward and feedback connectivity in relation to conscious awareness of changes. To test the
638 generality of our findings, further investigations are necessary, especially with techniques that
639 explicitly manipulate conscious awareness under comparable task conditions testing for the neuronal
640 effects on prediction and surprise.

641

642 **Author contributions:** ER, NT, & MG designed the study; ER collected and analyzed data; ER wrote
643 the first draft of the manuscript; ER, NT, & MG edited the manuscript.

644

645 **Funding:** This work was funded by a University of Queensland Fellowship (2016000071) and the
646 ARC (Australian Research Council) Centre of Excellence for Integrative Brain Function (ARC Centre

This is a provisional file, not the final typeset article

Grant CE140100007) to MG, an ARC Future Fellowship (FT120100619) to NT, as well as ARC Discovery Projects: DP180104128 to MIG and NT and DP180100396 to NT.

Acknowledgements: We thank the volunteers for participating in this study. We also thank both reviewers for their insightful comments, which helped improved the manuscript.

Conflict of Interest Statement: *The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest*.

REFERENCES

Amendedo, E., Pazo-Alvarez, P. & Cadaveira, F. (2007). Vertical asymmetries in pre-attentive detection of changes in motion direction. *International Journal of Psychophysiology, 64,* 184-189.

Anderson, R.A. (1989). Visual and eye movement functions of the posterior parietal cortex. *Annual Reviews of Neuroscience*, *12*, 377-403.

Astikainen, P., Cong, F., Ristaniemi, T. & Hietanen, J.K. (2013). Event-related potentials to unattended changes in facial expressions: detection of regularity violations or encoding of emotions? *Frontiers in Human Neuroscience, 7*(557), 1-10.

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

Bendixen, A., & Andersen, S. K. (2013). Measuring target detection performance in paradigms with high event rates. *Clinical Neurophysiology,* 124, 928-940. https://doi.org/10.1016/j.clinph.2012.11.012

Bernat, E., Shevrin, H. & Snodgrass, M. (2001). Subliminal visual oddball stimuli evoke a P300 component. *Clinical Neurophysiology, 112*, 159-171.

Berti, S. (2011). The attentional bilink demonstrates automatic deviance processing in vision. *NeuroReport,* 22, 6647-667.

Boly, M., Garrido, M.I., Gosseries, O., Bruno M.-A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S. & Friston, K. (2011). Preserved feedforward but impaired top-down processes in the vegatitive state. *Science, 332,* 858-862.

Boly, M., Seth, A.K., Wilke, M., Ingmundson, P., Baars, B., Laureys, S., Edelman, D.B. & Tsuchiya, N., (2013). Consciousness i humans and non-human animals: recent advances and future directions. *Frontiers in Psychology, 4(*625), 1-20.

Bor, D. & Seth, A. (2012). Consciousness and the prefrontal parietal network: Insights from attention, working memory, and chunking. *Frontiers in Psychology, 3*(63), 1-16.

Born, R.T., & Bradley, D.C. (2005). Structure and function of visual area MT. *Annual Reviews of Neuroscience, 28*, 157-189.

Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision 10*:433-436.

Brazdil, M., Rektor, I., Daniel, P., Dufek, M. & Jurak, P. (2001). Intracerebral event-related potentials to subthreshold target stimuli. *Clinical Neurophysiology, 112,* 650-661.

Britten, K.H., Shadlen, M.N., Newsome, W.T. & Movshon, J.A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745-4765.

Clark, A., (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive

This is a provisional file, not the final typeset article

science. *Behavioural and Brain Sciences, 36(*3), 1-73.

Clery, H., Andersson, F., Bonnet-Brilhault, F., Phillippe, A., Wicker, B. & Gomot, M. (2013). fMRI investigation of visual change detection in adults with autism. *NeuroImage:Clinical,* 303-312.

Czigler, I., Weisz, J. & Winkler, I. (2007). Backward masking and visual mismatch negativity: Electrophysiological evidence for memory-based detection of deviant stimuli. *Psychophysiology, 44,* 610-619.

Czigler, I. (2014). Visual mismatch negativity and categorization. *Brain Topography, 27*(4), 590-598.

David, O. & Friston, K.J. (2003). A neural mass model for MEG/EEG. *NeuroImage*. *20*(3), 1743–1755.

Dehaene, S. & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70,* 200-227.

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-P. & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience,* 4, 752-758. doi: https://doi.org/10.1038/89551

Dehaene, S., Sergent, C. & Changeux, J.-P. (2003). A neuronal network model linking subjective reports an objective physiological data during conscious perception. *Proceedings of National Academy of Science, 100*, 8520-8525.

Del Cul, A., Baillet, S., Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology, 5*(10), e260, doi:10.1371/journal.pbio. 0050260

DeYoe E.A. & Van Essen D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience*, 11(2), 19-226.

Dubner, R., & Zeki, S.M., (1971). Response properties and receptive fields of cells in an anatomically

defined region of the superior temporal sulcus in the monkey. *Brain Research, 35*(2), 528-532.

Dykstra, A.R. & Gutschalk, A. (2015). Does the mismatch negativity operate on a consciously accessible memory trace? *Science Advances,* e1500677, 1-10.

Engel, A. K., Fries, P. & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Neuroscience, 2,* 704-717.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage, 25*(4), 1325-1335.

Faivre, N., Arzi, A., Lunghi, C. & Salomon, R. (2017), Consciousness is more than meets the eye: A call for a multisensory study of subjective experience. *Neuroscience of Consciousness, 1-8*

Felleman, D.J. & Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*. 1, 1-47.

Files, B.T., Auer, E.T. & Bernstein, L.E. (2013). The visual mismatch negativity elicited with visual speech stimuli. *Frontiers in Human Neuroscience, 7*(371), 1-18.

Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: Frontal activity related to introspection and action but not to perception. *Journal of Neuroscience, 34*(5), 1738-1747. doi:10.1523/jneurosci.4403-13.2014

Friston, K.J., Harrison, L., Penny, W. (2003). "Dynamic causal modelling". *NeuroImage*. *19*(4), 1273–1302

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B,*

This is a provisional file, not the final typeset article

*360(*1456), 815-836.

Friston, K.J. & Stephan, K.E. (2007). Free-energy and the brain. *Syntheses, 159(*3), 417-458

Friston, K.J., Trujillo-Barreto, N., Daunizeau, J. (2008). DEM: a variational treatment of dynamic systems. *NeuroImage,* 41, 849–885

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C.M., Ziegler, C. & Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *NeuroImage, 128,* 413-431

Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., Cohen, L. & Naccache, L. (2009). Converging intracranial markers of conscious access. *PLoS Biology, 7*(3): e1000061. doi:10.1371/ journal.pbio.1000061

Gayle, L.C., Gal, D.E. & Kieffaber, P.D. (2012). Measuring affective reactivity in individuals with autism spectrum personality traits using the visual mismatch negativity event-related brain potential. *Frontiers in Human Neuroscience, 6*(334), 1-7.

Garrido, M., Kilner, J.M., Kiebel, S.J. & Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of National Academy of Science, 104*(52), 20961-20966

Garrido, M.I., Rowe, E.G., Halasz, V., & Mattingley, J.B. (2017). Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimuli. Cerebral Cortex. doi: 10.1093/cercor/bhx087

Gazzaniga, M.S. (2000). Cerebral specialization and interhemispheric communication. Does the corpus callosum enable the human condition? *Brain, 123,* 1293-1326.

Gilbert, C. & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron,*

54, 677-695.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (Vol. 1). New York: Wiley.

Haroush, K., Deouelle, L.Y. & Hochstein, S. (2011). Hearing whilst blinking: Multisensory attentional blink revisited. *The Journal of Neuroscience, 31*(3), 922-927.

Hohwy, J. (2013). The predictive mind. New York, NY, US: *Oxford University Press.* http://dx.doi.org/10.1093/acprof:oso/9780199682737.001.0001

Ilg, U.J., ASchumann, S. & Thier, P. (2004). Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron, (43),* 145-151.

Jack, B.N., Roeber, U. & O'Shea, R.P. (2015). We make predictions about eye of origin of visual input: Visual mismatch negativity from binocular rivalry. *Journal of Vision, 15*(13), 1-19.

Jack, B., Widmann, A., O'Shea, R.P., Schroger, E., & Roeber, U. (2017). Brain activity from stimuli that are not perceived: Visual mismatch negativity during binocular rivalry suppression. *Psychophysiology, 54*(5), 755-763.

Kecskes-Kovacs, K., Sulykos, I. & Czigler, I. (2013). Is it a face of a woman or a man? Visual mismatch negativity is sensitive to gender category. *Frontiers in Human Neuroscience, 7*(532), 1-11.

Kim, C.-H. & Blake, R. (2005). Psychophysical magic: rendering the visible 'invisible'. *Trends in Cognitive Sciences, 9,* 381-388.

Kimura, M. & Takeda, Y. (2013). Task difficulty affects the predictive process indexed by visual mismatch negativity. *Frontiers in Human Neuroscience, 7*(267), 1-13.

Kleiner M, Brainard D, Pelli D, 2007, "What's new in Psychtoolbox-3?" Perception 36 ECVP

This is a provisional file, not the final typeset article

Abstract Supplement.

Kogai, T., Aoyama, A., Amano, K. & Takeda, T. (2011). Visual mismatch response evoked by a perceptually indistinguishable oddball. *NeuroReport, 22,* 535-538.

Kok, P., Rahnev, D., Jehee, J.F., Lau, H.C. & de Lange, F.P. (2012) Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex, 22*(9): 2197-206. doi: 10.1093/cercor/bhr310

Kreegipuul, K., Kuldkepp, N., Sibolt, O., Toom, M., Allik, J. & Näätänen, R. (2013). vMMN for schematic faces: automatic detection of change in emotional expression. *Frontiers in Human Neuroscience, 7*(714), 1-11.

Kremláček, J., Kuba, M., Kubova, Z., Langrova, J., Szanyi, J., Vit, F. & Bednar, M. (2013). Visual mismatch negativity in the dorsal stream is independent of concurrent visual task difficulty. *Frontiers in Human Neuroscience, 7*(411), 1-7.

Kremláček, J. Kreegipuu, K. Tales, A., Astikainen, P., Põldver, Näätänen, R. & Stefanic, G. 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 etection of motion as indexed by event-related potentials and its behavioural correlate. *Frontiers in Human Neuroscience, 7*(476), 1-9.

Lamme, V.A., Zipser, K. & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience, 14*, 1044-1053.

Laureys, S. (2005). The neural correlates of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Science, 9(*12), 556-559

LeDoux, J. E. (1996). The emotional brain: The mysterious underpinnings of emotional life. New

York: Simon & Schuster.

Lorenzo-Lopez, L., Amenedo, E., Pazo-Alvarez, P. & Cadaveira, F. (2004). Pre-attentive detection of motion direction changes in normal aging. *NeuroReport, 15*(17), 2633-2636.

Maekawa, T., Katsuki, S., Kishimoto, J., Onitsuka, T., Ogata, K., Yamasaki, T., Ueno, T., Tobimatsu, S. & Kanba, S. (2013). Altered visual information processing systems in bipolar disorder: evidence from visual MMN and P3. *Frontiers in Human Neuroscience, 7*(403), 1-11.

Meneguzzo, P., Tsakiris, M., Schioth, H.B., Stein, D.J., & Brooks, S.J. (2014). Subliminal versus supraliminal stimuli activate neural responses in anterior cingulate cortex, fusiform gyrus and insula: a meta-analysis of fMRI studies. *BMC Psychology, 2*(52), 1-11.

Movshon, J.A., & Newsome, W.T., (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *Journal of Neuroscience, 16*(23), 7733-7741.

Muller, D., Roeber, U., Winkler, I., Trujilli-Barreto, N., Czigler, I. & Shroger, E. (2012). Impace of lower- vs. upper-hemifield presentation on automatic colour-deviance detection: a visual mistmatch negativity study. *Brain Research, 1472*, 89-98.

Näätänen, R., Tervaniemi, N., Sussman, E., Paavilainen, P., Winkler, I. (2001). Primitive intelligence in the auditory cortex. *Trends in Neuroscience, 24*, 283-288.

O'Shea, R. P. (2015). Refractoriness about adaptation. *Frontiers in Human Neuroscience, 9*(38), 1-3. https://doi.org/10.3389/fnhum.2015.00038

O'Shea, R. P., Kornmeier, J., & Roeber, U. (2013). Predicting visual consciousness electrophysiologically from intermittent binocular rivalry. *PLoS ONE*, *8*(10), e76134. https://doi.org/10.1371/journal.pone.0076134

Panksepp, J. (1998). Affective Neuroscience: The foundations of human and animal emotions. New

This is a provisional file, not the final typeset article

York: Oxford University Press.

Pascal-Leone, A. & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292,* 510-512.

Pazo-Alvarez, P., Cadaveira, F., & Amenendo, E. (2003). MMN in the visual modality: A review. *Biological Psychology, 63*(3), 199-236

Pearl, J. (2000). *Causality* (1st edition). Cambridge, United Kingdom. Cambridge University Press.

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision 10*:437-442.

Penny, W., Stephan, K.E., Daunizeua, J., Rosa, M.J., Friston, K.J., Schofield, T.M., Leff, A.P. (2010). Comparing families of Dynamic Causal Models. *PLoS Computational Biology, 6*(3), 1-14.

Plomp, G., Hervais-Adelman, A., Astofli, L. & Michel, C.M. (2015). Early recurrence and ongoing parietal driving during elementary visual processing. *Scientific Reports, 5*(18733).

Prete, G., Marzoli, D. & Tommasi, L. (2015). Upright or inverted, entire or exploded: right-hemispheric superiority in face recognition withstands multiple spatial manipulations. *PeerJ,* 3:e1456; DOI 10.7717/peerj.1456

Rao, R.P. & Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, (2)*1, 79-87

Rensink R.A. (2002). Change detection. Annual Review of Psychology, 53, 245-277.

Rensink, R.A. (2004). Visual sensing without seeing. *Psychological Science, 15*(1), 27-32.

Salzmann, C.D., Murasugi, C.M., Britten, K.H., & Newsome, W.T. (1992). Microstimulation in visual area MT: effects on direction discrimination performance. *Journal of Neuroscience, 12*(6),

845 2331-2355.

846 Sary G., Vogels, R., Kovacs, G., & Orban, G.A. (1995). Responses of monkey inferior temporal
847 neurons to luminance, motion and texture-defined gratings. *Journal of Neurophysiology, 4,*
848 1341-1354

849 Shepman, A., Rodway, P., & Pritchard, H. (2016). Right-laterilised unconscious, but not conscious,
850 processing of affective environmental sounds. *Laterality: Asymmetries of Body, Brain, and*
851 *Cognition, 21*(4-6), 606-632

852 Shtyrov, Y., Goryainova, G., Tugin, S., Ossadtchi, A. & Shestakova, A. (2013). Automatic processing
853 of unattended lexical information in visual oddball presentation: neurophysiological evidence.
854 *Frontiers in Human Neuroscience, 7*(421), 1-10.

855 Sincich, L.C., Park, K.R., Wohlgemuth, M.J., Horton, J.C. (2004). Bypassing V1: A direct geniculate
856 input to area MT. *Nature Neuroscience, 7*(10): 1123-1128. doi: 10.1038/nn1318.

857 Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E., Daunizeau, J. & Friston, K. (2010). Ten
858 simple rules for dynamic causal modeling. *Neuroimage*, 49, 3099-3109.

859 Stefanics, G., Kremlacek, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding
860 view. *Frontiers in Human Neuroscience, 8*(666), 1-19.

861 Stefanics, G., Astikainen, P., Czigler, I., eds. (2015). Visual Mismatch Negativity (vMMN): A
862 Prediction Error Signal in the Visual Modality. Lausanne: Frontiers Media. doi: 10.3389/978-
863 2-88919-560-2

864 Stothart, G. & Kazanina, N. (2013). Oscillatory characteristics of the visual mismatch negativity: what
865 evoked potentials aren't telling us. *Frontiers in Human Neuroscience, 7*(426), 1-9.

866 Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J. & Hirsch, J. (2006) Predictive codes

This is a provisional file, not the final typeset article

for forthcoming perception in the frontal cortex. *Science, 314*(5803): 1311-4.

Summerfield, C. & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Science, 13*(9), 403-9. doi: 10.1016/j.tics.2009.06.003

Takacs, E., Sulykos, I., Czigler, I., Barkaszi, I. & Balazs, L. (2013). Oblique effect in visual mismatch negativity. *Frontiers in Human Neuroscience, 7*(591), 1-13.

Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience, 4,* 219-229.

Tsuchiya, N., Frässle, S., Wilke, M. & Lamme, V. (2016). No-report and report-based paradigms jointly unravel the NCC: Response to Overgaard and Fazekas. *Trends in Cognitive Sciences, 20*(4), 242-243. https://doi.org.10.1016/j.tics/2016.01.006

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V.A.F., (2015). No-report paradigms Extracting the true neural correlates of consciousness. *Trends in Cognitive Science, 19*(12), 757-770.

Tsushima, Y., Sasaki, Y. & Watanabe, T. (2006) Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science*, *314(*5806), 1786-1788

van Rhijn, M., Roeber, U. & O'Shea, R.P. (2013). Can the eye of origin serve as a deviant? Visual mismatch negativity from binocular rivalry. *Frontiers in Human Neuroscience, 7*(190), 1-10.

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S. & Roelfsemal, P.R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science,* 360, 537-542

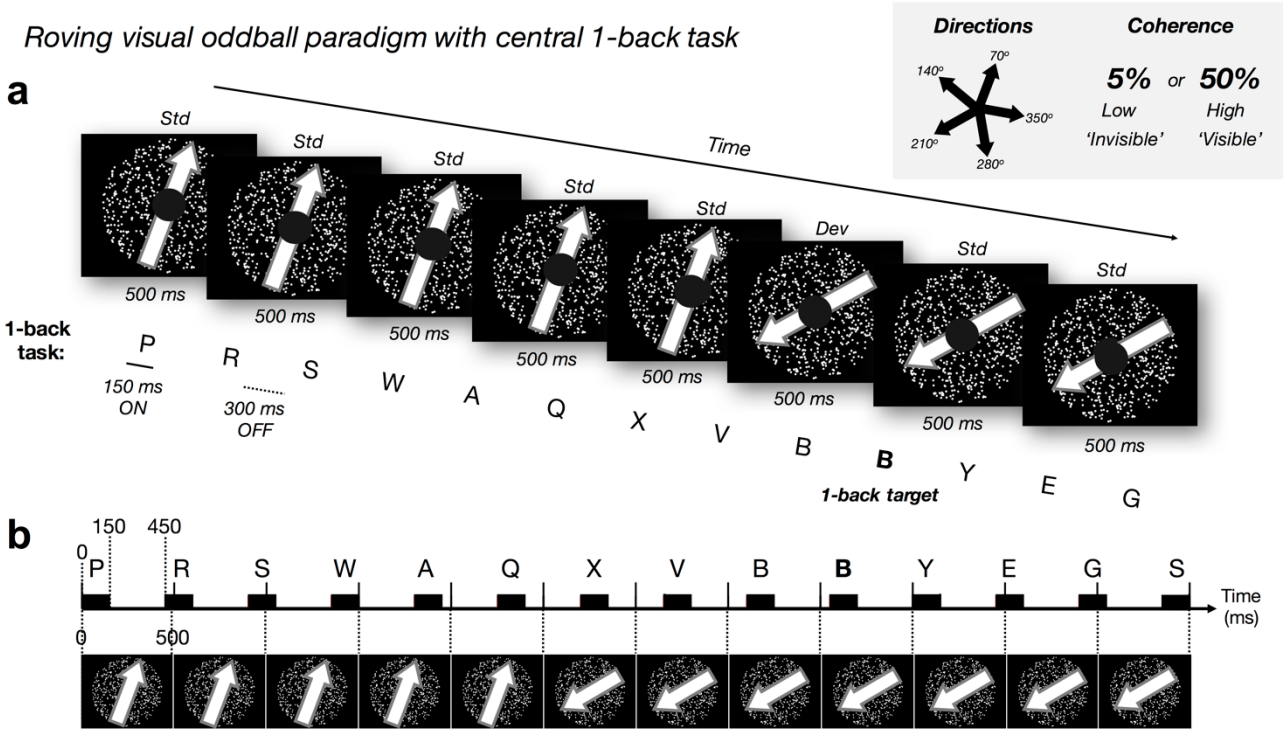
Wager TD, Keller MC, Lacey SC, Jonides J. Increased sensitivity in neuroimaging analyses using robust regression. NeuroImage. 2005;26(1):99–113

Watanabe, T., Nanez, J.E., & Sasaki, Y. (2011). Perceptual learning without perception. *Nature, 413,* 844-848.

Williams, L.M., Liddell, B.J., Kemp, A.H., Bryant, R.A., Meares, R.A., Peduto, A.S., & Gordon, E. (2006). Amygdala-prefrontal dissociation of subliminal and supraliminal fear. *Human Brain Mapping, 27*, 652-661.

This is a provisional file, not the final typeset article

914 **FIGURE LEGENDS**



915

916 **Figure 1 – Experimental paradigm.** (a) Roving visual oddball paradigm with dot motion at

917 five possible directions and two possible coherence levels (low: 5% and high 50%). We
918 manipulated *motion coherence*: high and low, and *motion direction*: standard ('std', frequent:
919 no change) and deviant ('dev', rare/surprising: change of direction). The 'roving design' meant
920 each deviant motion direction became the new standard direction over repetitions. We
921 presented the task-irrelevant random-dot motion (500 ms per trial) in the visual periphery.
922 Every 5 to 8 'standard' trials the global direction of the dots changed (the 'deviant'; randomly
923 selected direction). Every 26 to 30 trials (i.e. 13-15 seconds), we changed both the direction
924 and coherence level of the motion (but discarded EEG events and behavioural responses to
925 these 'double deviants'). Participants focused on a central 1-back letter task (150 ms ON and
926 300 ms OFF: desynchronized with most of the motion trials, aligning once every nine trials),

responding when the same letter was repeated in succession (e.g., bold letter, 'B'). Participants were instructed to ignore the motion stimuli and to be as fast and as accurate in the central task. (b) Schematic illustration of the timecourse for motion trials (every 500 ms) and 1-back letter presentation (black rectangles represent 150 ms letter ON).

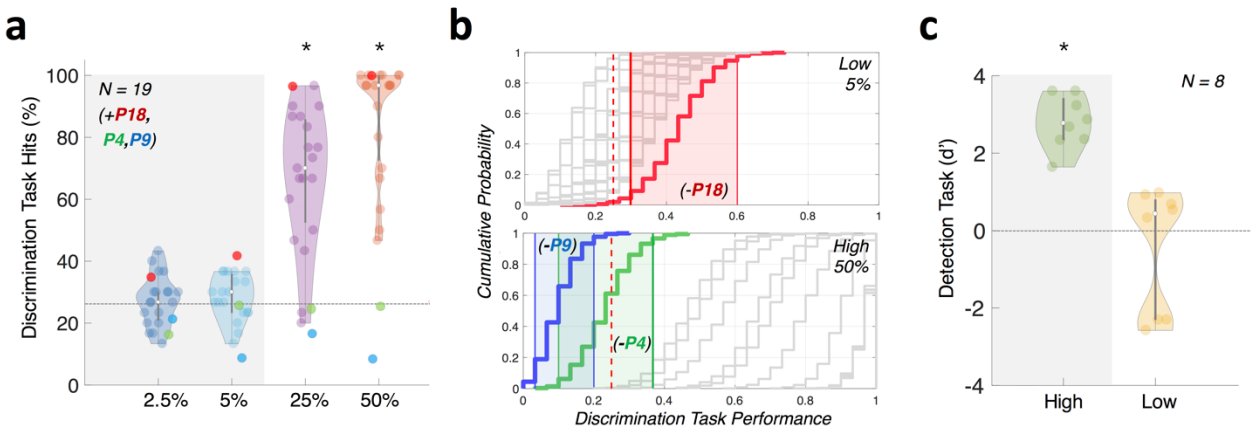


Figure 2 – Results from follow-up psychophysics tasks 1 and 2 confirming motion (in)visibility. **(a)** Direction discriminability follow-up psychophysics task 1 for participant exclusion. Accuracy in this task was used to confirm that the overall direction of motion at 5% coherence was not discernible, whilst 50% coherence motion was perceived and reported. Grey and coloured dots represent included and excluded participants, respectively, for the further EEG analysis, based on our bootstrapping analysis (panel b). **(b)** We confirmed our participant exclusion using bootstrapping of performance (plotted as cumulative histogram; included participants plotted in grey and excluded participant plotted in colour). Participants were excluded if they performed above chance for 5% coherence (left panel, N = 1; P18, red) or below chance for 50% coherence (right panel, N = 2; P4, green and P9, blue). Coloured shading indicates 95% confidence interval for the respective excluded participants. Red dotted line indicates chance-level performance. Removed participants are included in panel (a) only as a reference. **(c)** Motion change detectability (d') for individual participants in psychophysics task

2. We found low-coherence (5%) d' was not significantly greater than zero but d' for high-coherence (50%) was. Note: Error bars show standard deviation.

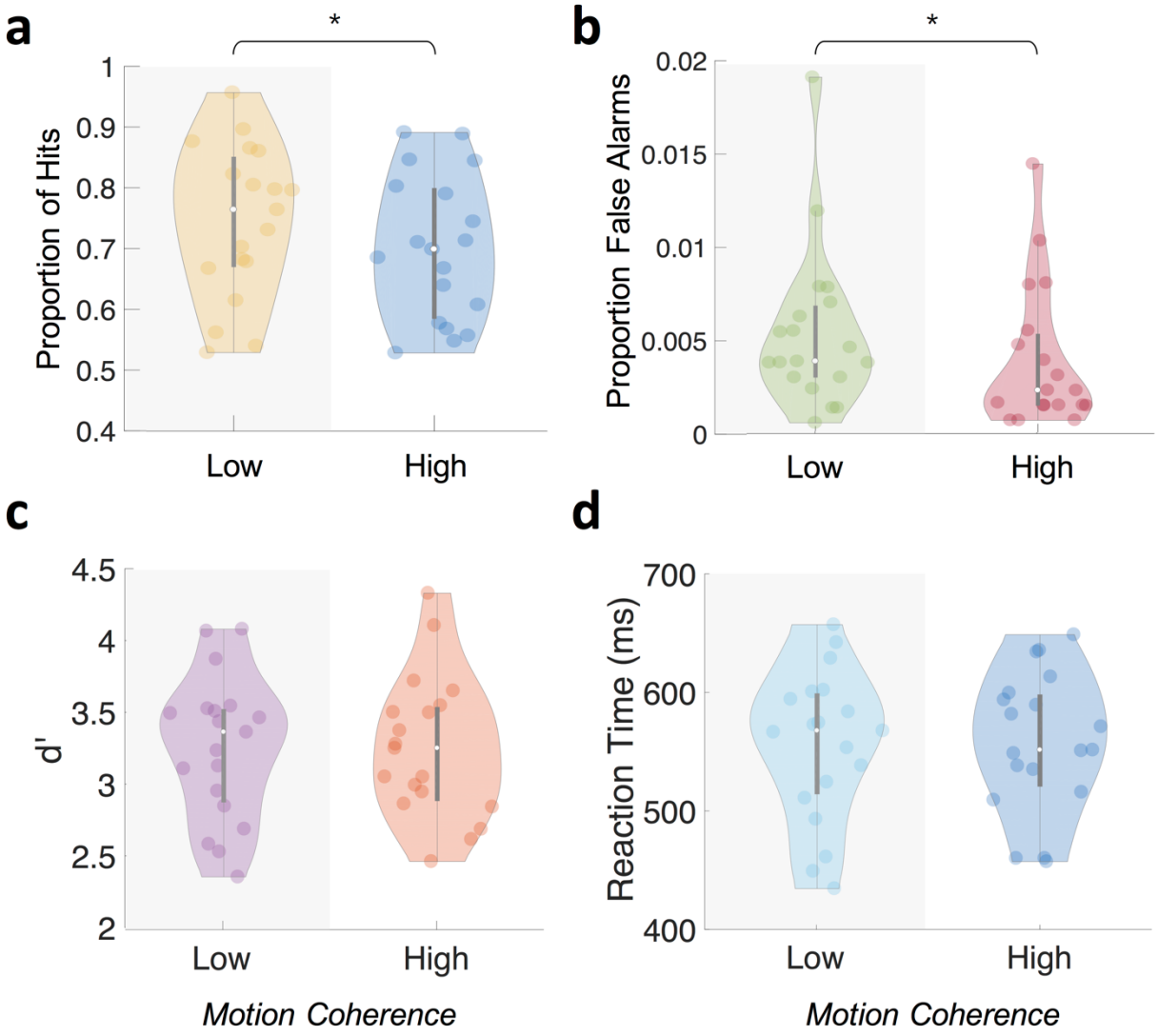


Figure 3 – Behavioural results from central 1-back task during Main Experiment. (a, b) Mean proportion of hits (a) and false alarms (b) was significantly higher ($p < 0.01$ and $p < 0.05$, respectively) when background motion was low-coherence than high-coherence. **(c, d)** Mean d' (c) and reaction times (d) showed no significant difference ($p > 0.05$) between the coherence levels.

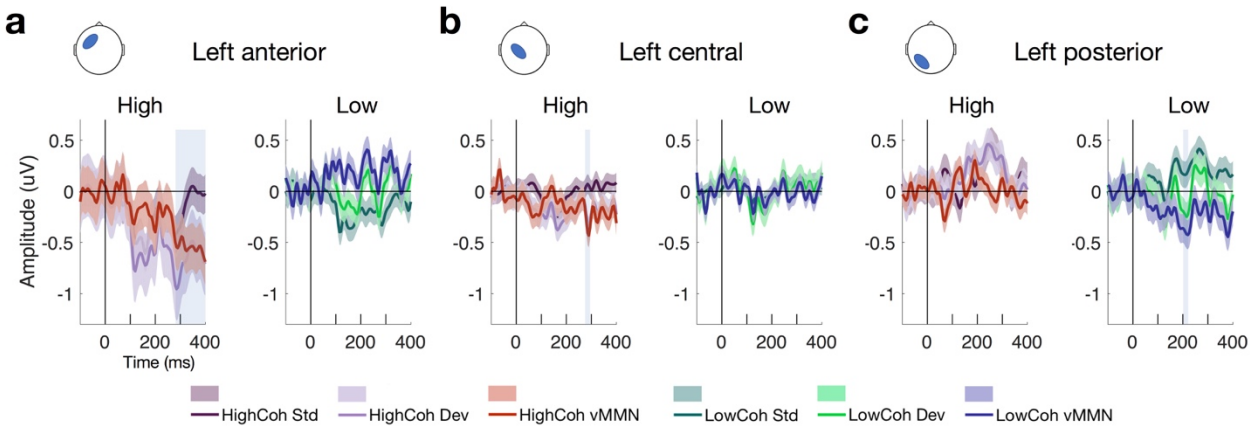
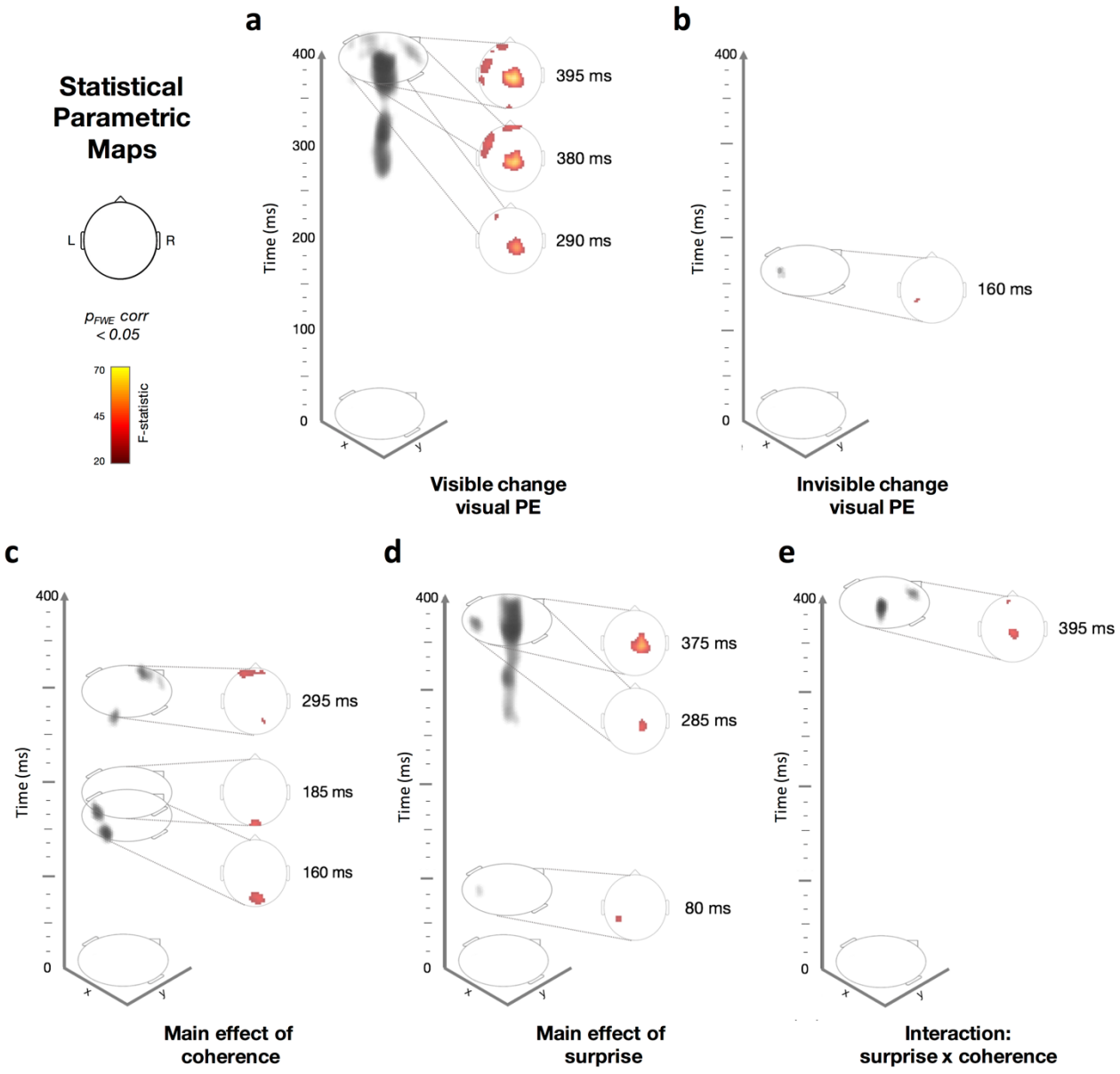


Figure 4 - ERP waveforms showing vMMN. Mean ERPs from electrode clusters with significant effects at (a) left anterior, (b) left central and (c) left posterior clusters, showing standard error across participants (N = 19). Left-hand panels show high-coherence ERPs for standard (dark purple), deviant (light purple) and vMMN difference waves (red; deviants - standards). Right-hand panels show low-coherence ERPs for standard (dark green), deviant (light green) and vMMN difference waves (blue; deviants - standards). Significance after FDR correction for multiple comparisons is denoted by the shaded blue boxes.

This is a provisional file, not the final typeset article


Statistical Parametric Maps
L
R
p_{FWE} corr < 0.05
70
45
20
F-statistic
a
Time (ms)
400
300
200
100
0
395 ms
380 ms
290 ms
Visible change visual PE
b
160 ms
Invisible change visual PE
c
295 ms
185 ms
160 ms
Main effect of coherence
d
375 ms
285 ms
80 ms
Main effect of surprise
e
395 ms
Interaction: surprise x coherence

966

967 **Figure 5 – Spatiotemporal Statistical Parametric Maps showing thresholded F-**
968 **statistic contrasts from the ERP data from (N=19; p_{FWE} corrected < 0.05).** Scalp-topography
969 orientation with respect to x- and y- axis is shown as a reference for panels (a-e) The colourmap
970 indicates the range of F-statistic values used in panels (a-e), ranging from 20 (black) to 70
971 (yellow), where F=24.18 corresponds to p<0.05 with family-wise error (FWE) correction. On
972 the left side of each panel, we show the 3D representation (grey shading) of the significant
973 clusters across space (x, y) and time (z, from 0 to 400ms after the trial onset). On the right side
974 within each panel, we show a time slice (e.g. Z=395ms, for example) of the F-statistics. Dotted
975 lines connect the time slice with its location within the 3D volume. (a) Visible standards versus
976 (vs) deviants. (b) Invisible standards versus (vs) deviants. (c) The main effect of coherence (high
977 vs low; regardless of surprise) (d) The main effect of surprise (standards vs deviants; regardless
978 of coherence). (e) An interaction between coherence and surprise. (p<0.05).

979

This is a provisional file, not the final typeset article

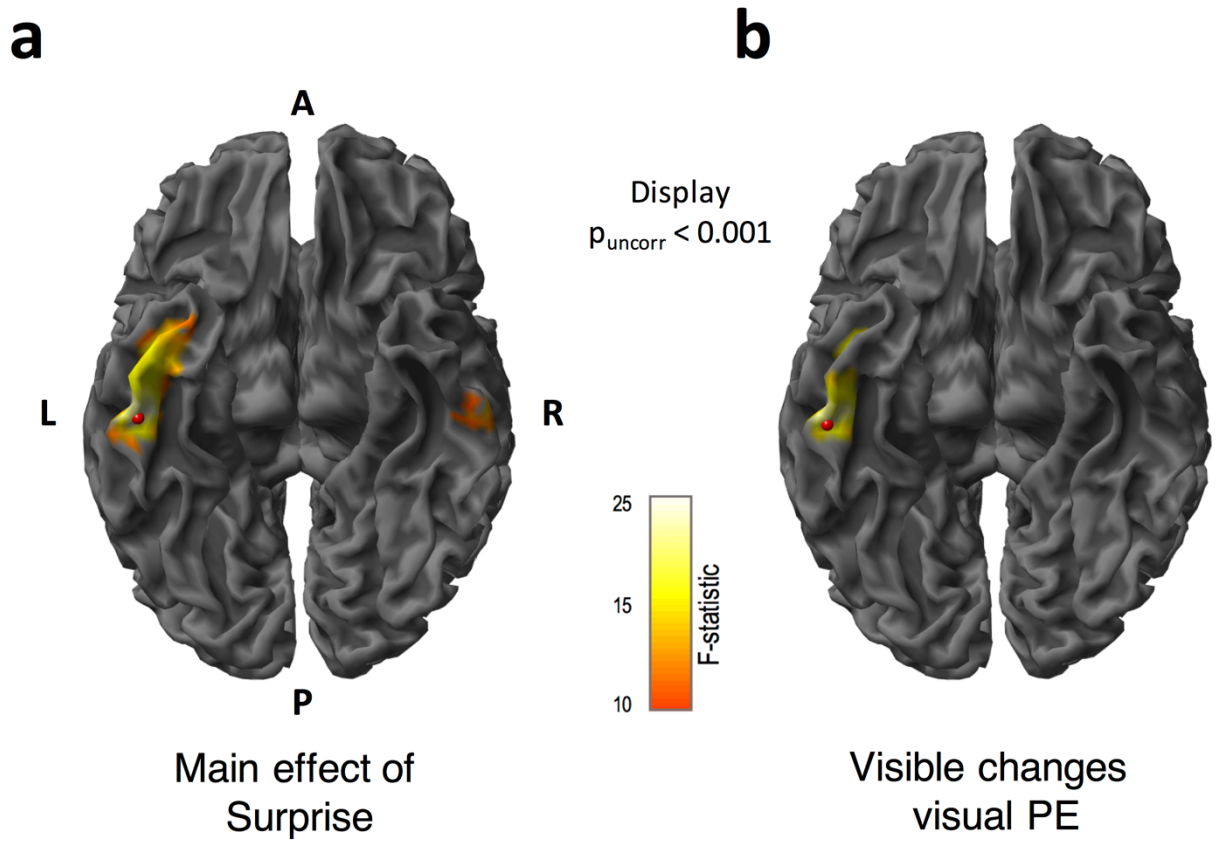


Figure 6 – Source reconstructed F-statistics for the main effect of surprise, and prediction errors (PE) to visible motion direction changes. We obtained source locations using multiple sparse priors (MSP) source reconstruction. These maps are displayed at $p_{uncorr}<0.001$. (a) For the main effect of surprise, we found a significant cluster in the left inferior temporal gyrus. (The cluster in the right inferior temporal gyrus did not survive after correction). (b) Source localisation for visible motion changes revealed a significant cluster in the left inferior temporal gyrus. The F-statistic is displayed for all contrasts within the range of 10 to 25, where F=12.11 corresponds to $p<0.05$ FWE corrected. A = anterior, P = posterior, L = left and R = right. Red beads indicate the location of the peak of F-statistic

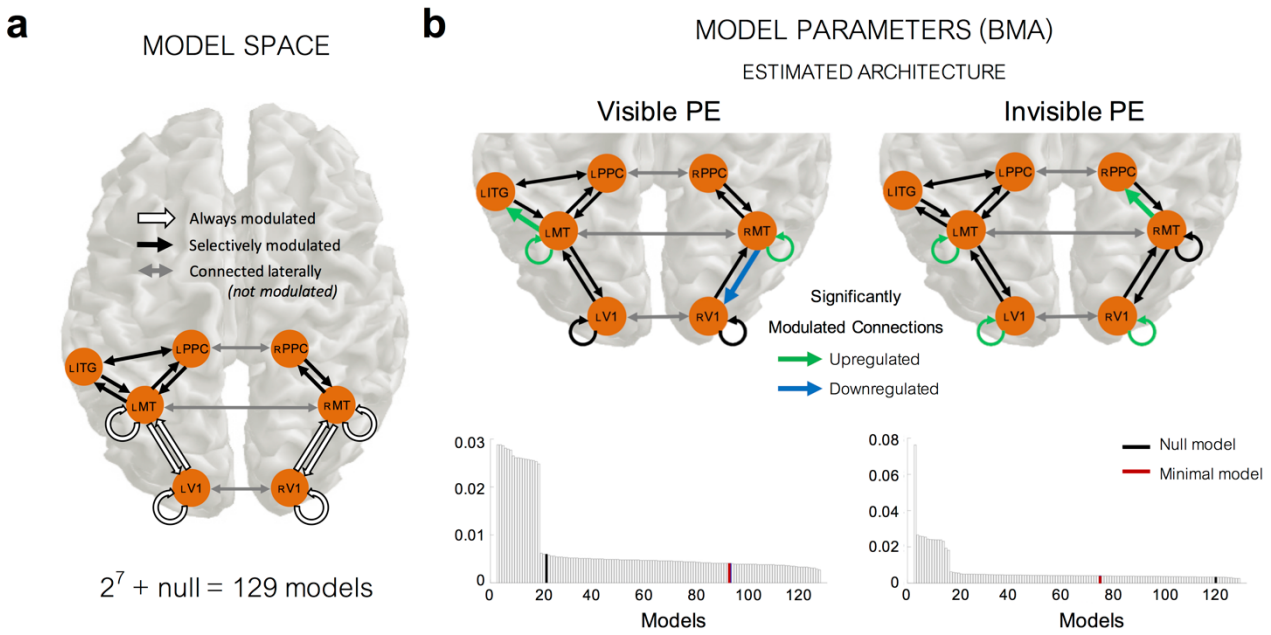


Figure 7 - Dynamic Causal Modelling (DCM) for visible and invisible PE responses. (a) Model space showing the architecture used in all models and modulation of these connections. We used the same architecture in all models, what differed were the allowed modulations of the connection strength between nodes. We systematically varied the modulation direction of all black connections in a binary manner to obtain $2^7 = 128$ models (+ a "null model" which did not have any modulation at or between any nodes). One of the $2^7 = 128$ models also contained "the minimal model", in which the only modulated connections were those between V1 and MT, and intrinsic modulation within V1 and MT. (b) Model evidence results from our random-effects (RFX) Bayesian Model Selection analyses, displayed as the exceedance probability of each model (i.e. the probability that a particular model is more likely than any other model given the group data). Using Bayesian Model Averaging (BMA) we were able to obtain weighted parameter estimates using the model evidence across all models and participants. Significantly modulated connections (False Discovery Rate corrected) are shown in green (upregulated) and blue (downregulated).