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

E.G. Rowe[1,2,3,4], N. Tsuchiya[3,4,5,6, *], M.I. Garrido[1,2,7,8, *],

[1]Queensland Brain Institute, University of Queensland
[2]Centre for Advanced Imaging, University of Queensland
[3]School of Psychological Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University,
[4]Turner Institute for Brain and Mental Health, Monash University, Melbourne, Victoria, Australia
[5]Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Suita, Osaka 565-0871, Japan
[6]Advanced Telecommunications Research Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan.
[7]ARC Centre of Excellence for Integrative Brain Function,
[8]Melbourne School of Psychological Sciences, The University of Melbourne
*: equal contribution

Correspondence:
E.G. Rowe, Turner Institute for Brain and Mental Health, 770 Blackburn Road, Monash University, Melbourne, Victoria, Australia
Telephone: (+61) 404 058 773
Email: elise.rowe@monash.edu

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

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: This work was funded by a University of Queensland Fellowship (2016000071) and the ARC (Australian Research Council) Centre of Excellence for Integrative Brain Function (ARC Centre Grant CE140100007) to MIG, an ARC Future Fellowship (FT120100619) to NT, as well as ARC Discovery Projects: DP180104128 to MIG and NT and DP180100396 to NT. We thank the volunteers for participating in this study.

ABSTRACT

Detecting changes in the environment is fundamental for survival, as these may indicate potential rewards or threats. According to predictive coding theory, detecting these irregularities relies on both incoming sensory information and our prior beliefs; with incongruity between the two manifesting as a prediction error (PE) response. Many changes occurring in our environment do not pose any threat and may go unnoticed. Such subtle changes can nevertheless be detected by the brain without ever emerging into consciousness. What remains unclear is how such sensory changes are processed in the brain at the network-level. Here, we developed a visual oddball paradigm, where participants engaged in a central letter task during electroencephalographic (EEG) recordings while presented with task-irrelevant (and unreported) high- or low- coherence background random dot motion. Critically, once in a while, the direction of the dots changed. After the EEG session, we behaviourally confirmed that changes in motion direction at high- and low-coherence were visible and invisible, respectively. ERP analyses revealed that changes in motion direction elicited PE regardless of visibility of such changes but with distinct spatiotemporal patterns. To better understand these responses at the network-level, we applied Dynamic Causal Modelling (DCM) to the EEG data. Bayesian Model Averaging analysis of the DCMs showed that both visible and invisible PE relied on a release from adaptation (or repetition suppression) within sensory areas of V1 and MT. Furthermore, while feedforward upregulation was present for invisible PE, results for the visible change PE also included downregulation of feedback connections between right MT to V1. Overall, our findings reveal a complex interplay of modulation in causal network underlying visible and invisible motion changes.

INTRODUCTION

Detecting changes in the environment is fundamental for survival, as these may indicate potential rewards or threats. According to predictive coding theory, we detect these irregularities by relying on both incoming (bottom-up) sensory information and our (top-down) prior beliefs or expectations (Rao & Ballard, 1999; Friston, 2005; Friston & Stephan, 2007; Hohwy, 2013; Clark, 2013). Incongruity between the two manifests as a prediction error (PE) response detectable in event-related potentials (ERPs), representing the difference between neural responses to expected and unexpected events. These PE responses to auditory stimuli have historically received more attention (Näätänen et al., 1978; Näätänen 1992; Garrido et al., 2007), but protocols for visual stimuli have also been developed, with a focus on situations where individuals are aware of a change (Muller et al., 2012; Kimura & Takeda, 2013; Clery et al., 2013; Takacs et al., 2013; Kremlacek et al., 2013; Kuldkepp et al., 2013; Pazo-Alvarez et al., 2004; Maekawa et al., 2013; Stothart & Kazanina, 2013; Astikainen et al., 2013; Gayle et al., 2012; Kreegipuu et al., 2013; Kecskes-Kovacs et al., 2013; Files et al., 2013; Shtyrov et al., 2013). In our everyday lives, however, many changes in the sensory environment that evoke PEs may go unnoticed (Czigler, 2014; Stefanics et al., 2014). And yet, it remains unclear why PE signals to detected changes emerge into consciousness while undetected changes never do.

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 (Kim & Blake, 2005). Employing such tools, researchers have demonstrated PE responses to changes without conscious awareness of such changes. The demonstrated PE responses in EEG include those using backward masking (Czigler et al., 2007; Kogai et al., 2011), binocular rivalry (Jack et al., 2015; 2017; van Rhijn et al., 2013), rapid unmasked presentation (1 ms, d' = 0; 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). DCM, applied to the auditory mismatch negativity (Garrido et al., 2007; Boly et al., 2011) in patients in coma, suggested that conscious awareness of changes requires feedback connectivity from higher to lower cortical areas. In that study, Boly et al., (2011), applied DCM to the auditory mismatch negativity 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). While this study suggests a potential involvement of feedback

modulations in regulating the level of consciousness (Laureys, 2005; Boly et al., 2013), it is unclear whether non-conscious processing in fully awake healthy participants also abolishes top-down feedback connections.

To understand the difference 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 coherence of random dot motion. Next, we used DCM to estimate effective connectivity and examined the involvement of feedback connections for visible and invisible changes.

METHODS

Participants

Twenty eight healthy university students participated in this study (N = 28, 10 females, aged 18-40, *M* = 22.44, *SD* = 4.30). No participants reported any 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 procedure

The entire experiment took under two hours, including the setup of the EEG (Biosemi Active Two system with 64 Ag/AgCl scalp electrodes arranged according to the international 10-10 system). Upon signing the consent form, participants received task instructions and were fitted with EEG electrodes. After the EEG was set up, participants practiced the one-back letter task for 1 min (see below) before being tested in the main task. During the main task, EEG was continuously recorded at a sampling rate of 1024 Hz. The main task was followed by psychophysical tests to assess the discriminability of motion stimuli (see below).

Display

The experiment was performed using a Macbook Pro and an external screen with a resolution of 1920 x 1080 pixels with a 60 Hz refresh rate. A viewing distance of 50 cm was maintained for all participants. The experiments were programmed using the Psychtoolbox extension for MATLAB (version 2014b).

Task and visual stimuli during EEG measurement

To achieve the desired level of visibility and to induce visual PE to visible and invisible changes, we manipulated the coherence of random dot motion (Britten et al., 1992). In short, it is well-known that the direction of motion can be consciously discriminated when the level of motion coherence of random dots is

sufficiently high. As the level of coherence approaches zero, participants can no longer consciously see the direction of motion (Watanabe et al., 2011). We exploited this useful feature of motion perception and designed a paradigm to elicit conscious and non-conscious prediction errors from perceptually visible and invisible changes in the motion direction of a cloud of random dots (Figure 1a, demo: https://figshare.com/s/76484519f510ba74891b). These motion stimuli were never task-relevant during the main EEG experiment.

Participants were told that the background motion should be completely ignored. Instead, participants were instructed to pay attention to the central letter 1-back task and press the spacebar as quickly as possible whenever they detected repetition of any letter (Figure 1b). The central 1-back task consisted of a sequence of white letters (randomly selected; roughly 1.8^{o} visual angle in size) centrally presented within a 5° diameter black circle. Each letter was presented for 150 ms with a 300 ms blank interval. One letter was repeated at random every 10 to 15 letters. On average, the number of letter repeat targets per participants across the experiment was 191 (range from 177 to 196, SD = 3.69). We considered participants to successfully detected the target (i.e., 'hit') if the response was made between 0 ms to 1,000 ms after the onset of letter repeat. Prior to the experiment, participants were given a practice block of one minute where feedback on performance was given. The practice block was repeated if their hit rate was below 50%. Two participants failed to correctly detect above 50% of the targets across the entire experiment. We removed these participants from further analyses as we could not confirm they were attending to the central task.

While participants were focusing on the central letters, our main interest was their EEG responses to the task-irrelevant background motion. We manipulated the property of the motion in a 2 x 2 design comparing the *coherence* levels: low and high, and *motion direction*: standard (frequent: no change) and deviant (rare: change of direction) (Figure 1c). We utilised a roving oddball paradigm with dot motion at 5 possible directions (70°, 140°, 210°, 280° and 350°; with 0° representing motion to the right) and 2 possible coherence levels (low: 5% and high 50%). Each block (of the twelve in total) started with one of the five motion directions (randomly chosen). A 30° diameter black circle enclosed the locations of the 750 white dots on a black background. During each 500 ms trial, each dot moved in a straight trajectory at 14°/sec. When any dot went outside of the stimulus window, the dot was redrawn in the opposite location, to keep the density of the dots constant. After 500 ms, we removed all dots and redrew them all at new locations at the same time. After 5 to 8 such redraws, we changed the global motion direction randomly into one of the other 4 directions, which may or may not have coincided with the change of the coherence level. We defined a global direction change (with no change in coherence) as a deviant motion direction event. Every 26 to 30 motion redraws (13-15 seconds), we changed both the direction and coherence, but discarded the EEG response to these double deviants. A block finished after 180 motion direction redraws (90 secs).

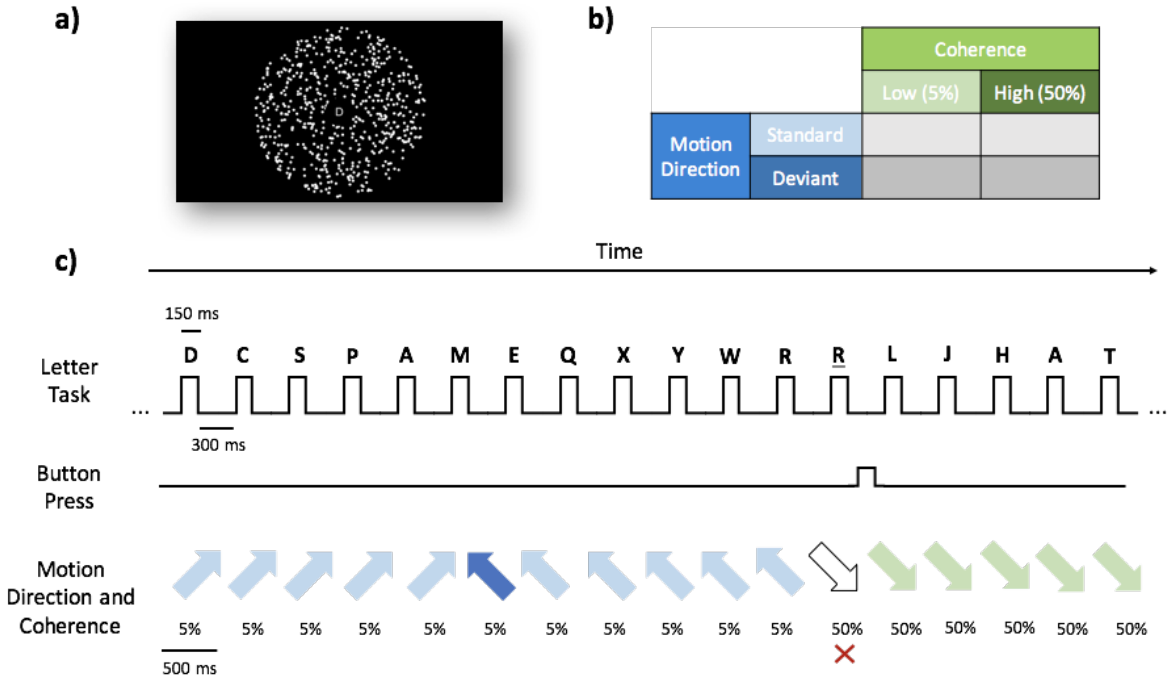

Figure 1 – Experimental paradigm. (a) Display. **(b)** We created a 2 x 2 design comparing the *coherence* levels: low and high, and *motion direction change*: standard (no change) and deviant (change). **(c)** Time course of the EEG experiment. Participants focused on the 1-back letter task (150 ms on, 300 ms off), responding when the same letter was repeated in succession (e.g., underlined letter, 'R'). As task-irrelevant background, we presented random dot motion stimuli, whose global direction changed after 5 to 8 redraws of all dots (every 500ms). Every 26-30 redraw, we changed the coherence level of the motion (5% coherence as blue arrows, 50% coherence as green arrows). We defined the direction changes within a given coherence level as deviant events for that coherence level (the darker arrow) and did not analyze events where both the direction and coherence changed (the empty arrow).

Follow-up direction discrimination task

Directly following the EEG experiment, we ran a motion direction discrimination task to verify the desired levels of visibility of motion at high coherence motion and invisibility at low coherence motion. During these tasks, we did not record EEG. To estimate the discriminability of direction of motion, we ran a four alternative-forced choice (4AFC) direction discrimination task. To conservatively estimate the discriminability, we reduced the direction alternatives to 4 possibilities (80°, 160°, 240° and 320°), which were different to the motion directions used in the main EEG experiment in order to reduce the chance of perceptual learning. In each trial, we presented the motion for 1 second without the central letter task (i.e. the small, black central circle remained but the letters were absent). Again, to conservatively estimate the motion direction discriminability, we did not redraw all the dots after 500 ms as in the EEG experiment but, instead, the dots remained moving for 1 second (except when it came to the boundary, see above). We randomly selected a motion direction in each trial and pseudo-randomly intermixed 4 coherence levels (2.5%, 5%, 25% and 50%) in equal proportions across 120 trials. At the end of every trial, participants reported the perceived direction of motion, from the 4 possible alternatives.

Behavioral results: ***Confirming visibility and invisibility of high and low coherence motion direction changes in each individual***

We used performance in our follow-up psychophysics direction discrimination task to confirm that high and low coherence motion direction changes were visible and invisible, respectively. Based on the performance, we excluded participants from the following EEG analysis based on two criteria: 1) performing above chance at low (5%) coherence motion condition or 2) performing below chance at high (50%) coherence motion condition. Based on our bootstrapping analyses (Supplementary Figure 1), 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. Finally, we ran a follow-up 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; Supplementary Figure 3). In summary, we used the data from N=19 participants for ERP, source, and DCM analysis.

DATA ANALYSIS

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 the false alarm rate. A hit was defined as a correct button press within 1,000 ms since the onset of repeated letter on the screen. A false alarm was defined as a button press made outside of this time window. We examined the rate of the hit and the false alarm rate as well as the reaction times for each hit across participants according to the motion coherence level. We present the d' as a summary of both hit and false alarm.

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 then detected and marked eyeblink artefacts using the VEOG channels with a bad channel maximum rejection threshold of 20%. Following this, data were epoched within a trial time window of 100 ms before to 400 ms after redrawing the random dot motion. Artefact removal then rejected trials marked with an eyeblink and by thresholding all channels at 100uV. To obtain the mean ERP per participant per condition, we robustly averaged the epoched data using a built-in function in SPM12. This robust averaging process weighted each time-point within a trial according to how different it was from the median across trials, that is, it down-weighted outliers within the dataset. Next, as high frequency noise can be introduced during the robust averaging process, we further low-pass filtered the processed data at 40 Hz. Finally, we

baseline corrected the data for each participant and condition by subtracting the (robustly averaged) ERP between -100 to 0 ms. Trials were sorted in terms of the conditions of interest in a 2x2 manner: high versus low coherence, standard (no change) versus deviant (change).

Spatio-temporal Image Conversion

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

Spatio-temporal Statistical Maps

The 3D spatio-temporal image volumes were analysed, on a participant-by-participant basis, with a mass univariate general linear model (GLM) as implemented in SPM12. We estimated the main effects of surprise (i.e., standards vs deviants) and coherence (i.e., high vs low coherence), their interaction, 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 (BEM) head model, and inverting a forward model with multiple sparse priors (MSP) assumptions for the variance components under group constraints (Friston et al., 2008). Since 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 FWHM 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 (DCM)

We used DCM to estimate a generative causal model that most parsimoniously explained the observed ERPs (from 0 to 400 ms peristimulus time window) at the selected source locations with minimal model complexity (Friston, 2003). We used a data-driven approach combined with a priori locations drawn from

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

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

Next, while there could be a huge number of possible DCM modulations based on our 7 identified nodes, we decided to reduce the possible space for modulations based on anatomical information as much as possible. By anatomical criteria, we decided to examine models that always contained recurrent (forwards and backwards) modulation between the input sources of bilateral V1 and MT and intrinsic modulations at these input nodes; we refer to this as the 'minimal' model. Using this reduced number of potential modulation directions, we were left with 7 connections that could be modulated beyond the minimal model: 4 connections between MT and PPC (forwards and backwards in each hemisphere), 2 connections between MT and IT (forwards and backwards in left hemisphere) and 1 connection between PPC and IT (one lateral connection in left hemisphere). We never modulated the lateral connections between the hemispheres. Thus, in total, we tested 129 models (i.e. 2^7 = 128 + a null model with no modulation between any nodes) comprising all combinations of modulation directions (see Figure 5a).

We performed DCM analyses to estimate the parameters of each model, separately for each effect of interest: (1) the visible PE and (2) the invisible PE. For the visible PE, we used the between-trial effect (condition weights) of [0, 1] for the high coherence standard and deviant trials, respectively. For the invisible PE, we used the weights of [0, 1] for the low coherence standard and deviant trials, respectively. DCM analyses started from estimation of the log model evidence for each of our 129 generative models fitted for every participant. The log model evidence quantified how well each model could generate ERPs similar to the observed (pre-processed) ERPs for that participant. Importantly, the iterative free-energy minimization

process used to calculate the log model evidence penalised models with greater complexity to avoid overfitting (Friston et al., 2008). Once estimated, we used a random-effects (RFX) approach to determine the winning model across participants via Bayesian Model Selection (BMS). RFX assumes that a (potentially) different model underpins each participant's responses; making it robust to outliers and best suited for studying perceptual processes whose underlying network structure is likely to be varied across participants (Stephan et al., 2009). To compute a weighted average of the parameter estimates across all models, we employed Bayesian Model Averaging (BMA) (Penny et al., 2010). BMA weights the estimated parameter with the probability of each model associated with that parameter for all participants. In this way, all models contributed to the final connectivity estimate, with the most probable model having the greatest weight and the least probable model contributing the least to the final estimates (Stephan et al., 2010). In a follow up DCM analyses, we compared the visible and invisible PE directly using the betweem-trial effect (condition weights) or [0,1] to contrast the mean ERPs for the visible PE and invisible PE per participant.

RESULTS

We used a random-dot motion oddball paradigm designed to elicit visual PEs to visible and (potentially) invisible changes in motion direction while EEG data were recorded. Participants performed a 1-back task at the centre of the screen and were instructed to ignore the peripheral motion. When a 1-back target was presented, participants were required to make a keyboard response within 1,000 ms of the onset or the trial was considered a miss. Participants accuracy and reaction times in this task were used to determine whether there was a behavioural effect from the task-irrelevant background motion or from surprises induced via a change in direction. We were interested in observing any differences in the spatiotemporal and source reconstructed EEG activity for visible and invisible changed induced PE. Furthering this, we investigated whether there were any network-level differences underlying both types of PE using DCM.

To ensure that, for each participant, high and low coherence motion direction changes were visible and invisible, respectively, we included two psychophysics follow-up tasks (see Supplementary Material). The first of these psychophysics tasks required participants to make a direction discrimination at high or low coherence levels, with 3 participant's being excluded. The second of these psychophysics tasks was performed in a subset of participants, requiring them to report when they 'felt' or 'sensed' a change in direction occurred at both high or low coherence levels. No participants were excluded based on these criteria, but was used as a sanity check.

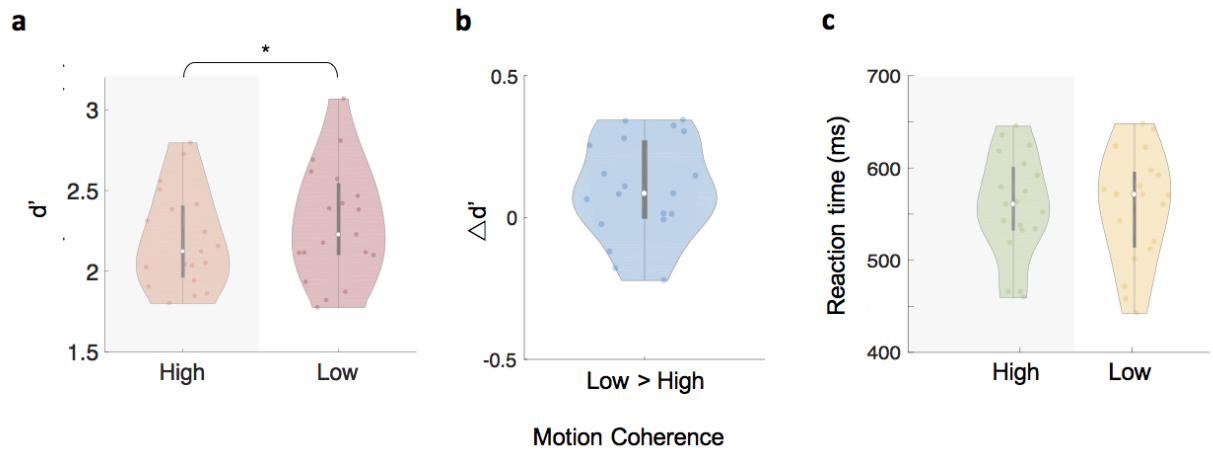

Figure 2 - Behavioural results. (a) Mean d' in the central 1-back task of the main experiment. Hit was defined as the button press within 1,000 ms of a target onset. Significant difference (p<0.05) between the two conditions is indicated by the '*' **(b)** Difference in d' between low and high coherence conditions showing significant increase in performance **(c)** Mean reaction times in the central 1-back task did not show a difference (p>0.05) between the high and low coherence conditions.

Evidence of distraction by visible but not invisible motion directions

After the exclusion based on the follow-up psychophysics task, we analysed the performance of the remaining N = 19 participants in the 1-back task to check if they appropriately focused on the central letter task and ignored the background motion during the prior EEG session. Based on prior studies (e.g., Tsushima et al., 2008), we expected that our low coherence background motion would distract participants more than high coherence motion to degrade the performance of the 1-back task.

Between the high and low coherence conditions, we observed a difference in accuracy but not in reaction times. The proportion of hits during the high coherence condition (M = 70.40%, SD = 11.72%) was significantly lower than in the low coherence condition (M = 74.19%, SD = 12.24%; two-tailed paired t-test $t(18) = -2.47$, $p = 0.002$). Mean number of false alarms across participants was only 1.84 (+/- 3.51, ranging from 0 to 4, except one participant who had 15 false alarms). Combining hits and false alarms, we computed d' measure to obtain the same trend: d' during the high coherence (M = 2.20, SD = 0.30) was significantly lower than in the low coherence condition (M = 2.23, SD = 0.35, t(18) = -2.55, p = 0.02) (Figure 2a, 2b). We found no differences in the reaction times between the high (M = 556 ms, SD = 59 ms) and low (M = 557 ms, SD = 63 ms, t(18) = -0.27, $p = 0.79$) coherence conditions (Figure 2c). Overall, we did not find the results to be consistent with our prior expectation, which we discuss in the Discussion. We interpret these results as evidence that participants were able to maintain their focal attention to the 1-back task and ignored the background motion, whose distraction effect was stronger during the high coherence than the low coherence condition.

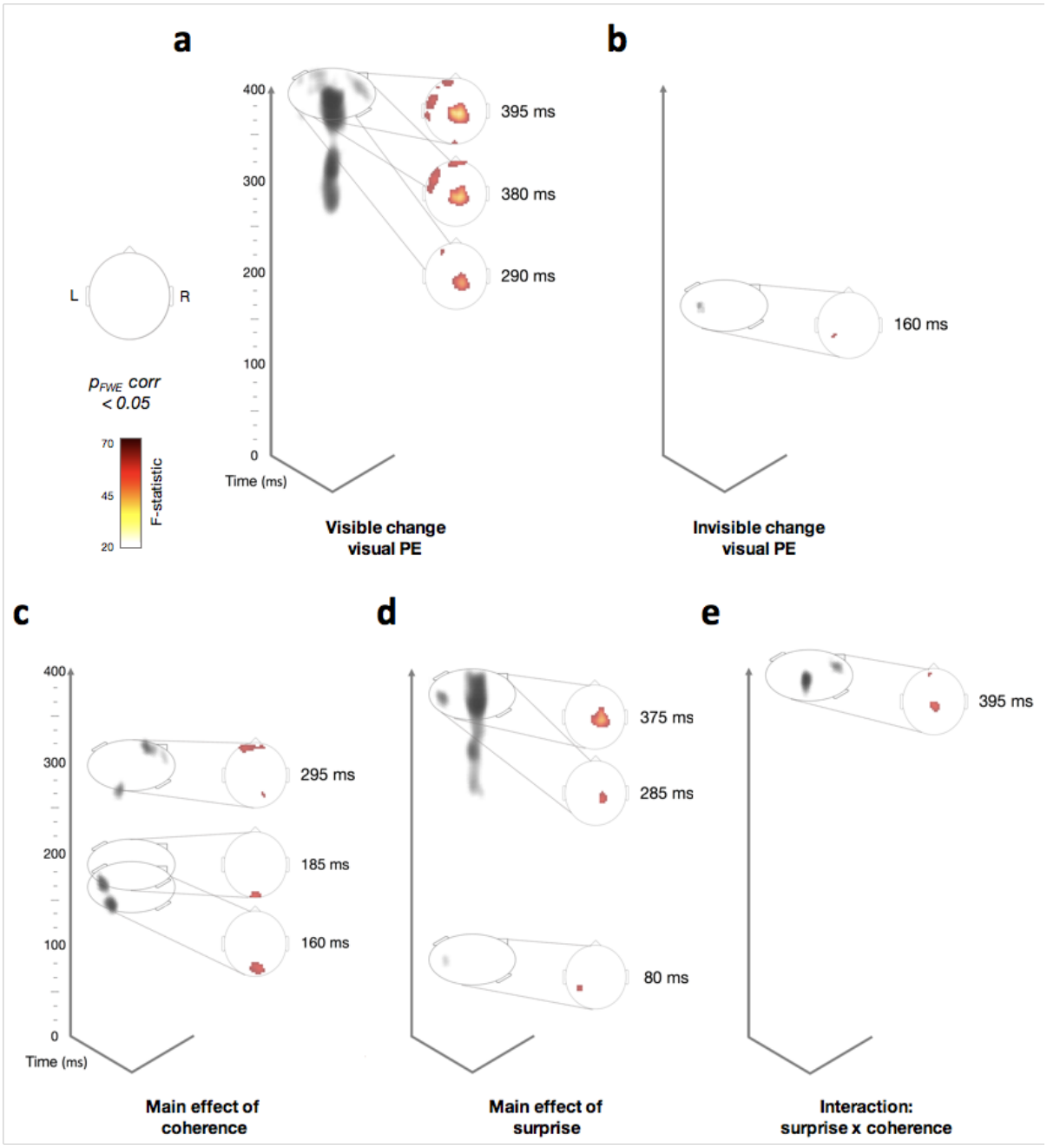

Figure 3 – Spatiotemporal F-statistics on the ERP data from (N=19). (a) PE to visible changes. **(b)** PE to invisible changes. **(c)** The main effect of coherence (high vs low; regardless of surprise) **(d)** The main effect of surprise (standards vs deviants; regardless of coherence). **(e)** An interaction between coherence and surprise. All clusters were obtained using family-wise error correction ($p<0.05$) and the F-statistic is displayed for all contrasts within the range of 20 to 70, where $F=24.18$ corresponds to $p<0.05$ FWE corrected.

Scalp-level ERP analysis

Non-conscious prediction errors occur earlier than conscious prediction errors

Using the 3D images interpolated from ERP data recorded at the scalp, we 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 PE were evoked from both visible and invisible motion direction deviants. PE to visible motion direction changes (Figure 3a) 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 3b); 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 3c), surprise (Figure 3d) and their interaction (Figure 3e). The main effect of coherence (Figure 3c) 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 3d) 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 3e) at central and frontal channels at 395 ms (peak-level F_{max} = 35.93, cluster-level p_{FWE} = 0.002).

Source-level analysis

Left ITG as a source for conscious PE

We applied Multiple Sparse Priors (MSP) source reconstruction to estimate the cortical regions involved in generating PE to visible and invisible motion direction changes. In Figure 3 (the scalp-level ERP), we found that the main effects of surprise, coherence, and the interaction spanned the whole epoch, thus we decided to use the whole epoch data (0-400 ms) rather than to temporally constrain the data for source reconstruction (see Data Analysis for details). Figure 4 shows the source-level results for the main effect of surprise, PE to visible changes at an uncorrected threshold of p<0.001. We did not find any significant sources for invisible PE, main effects of coherence or interaction when we corrected for multiple

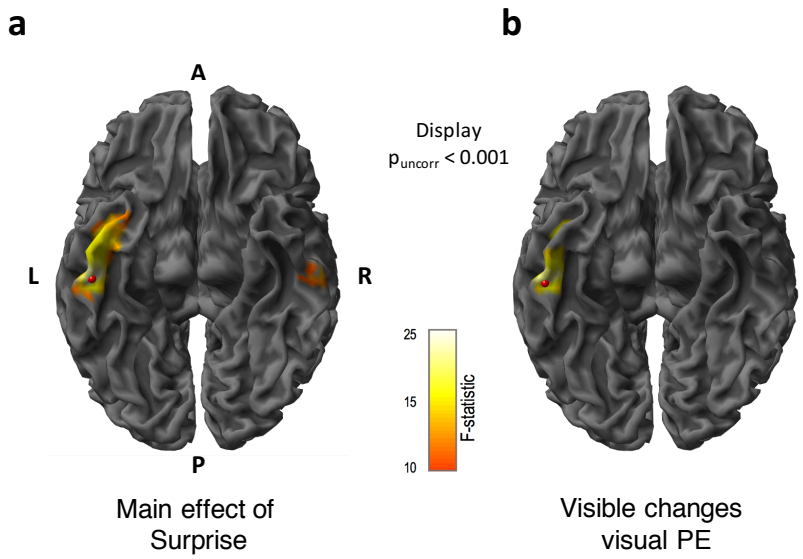

Figure 4 – Source reconstructed F-statistics for the main effect of surprise, and prediction errors (PE) to visible motion direction changes. We obtained source locations using multiple sparse priors (MSP) source reconstruction. These maps are displayed at p_{uncorr}<0.001. **(a)** For the main effect 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.

comparisons. Only the main effect of surprise and PE to visible changes were significant at the source-level after correction. For the main effect of surprise, we found one significant cluster in the left ITG ([-52 -26 -30, peak-level F_{max} = 22.19, cluster-level p_{FWE} = 0.003, Figure 4a). For the PE to visible changes, we found a similar cluster in the left ITG ([-48, -12, -32], peak-level F_{max} = 17.91, cluster-level p_{FWE} = 0.039, Figure 4b). The effect of surprise for the invisible conditions revealed a cluster on the right hemisphere (Figure 4a) which did not survive correction.

Dynamic Causal Modelling (DCM)

We used DCM to examine how the source location identified in the previous step, interacted with cortical locations known to specialise in motion processing, and how the strengths of these interactions are modulated by the visibility of motion changes. Specifically, we identified one region at the family-wise error corrected (*P*<0.05) threshold (between-subject F-tests) (Anatomy Toolbox; Eickhoff et al., 2005): the left inferior temporal gyri (**ITG**; MNI coordinate: [-48, -12, -32], Figure 4). We included the sensory input nodes of bilateral primary visual cortices (**V1**; MNI coordinates: left [-14, -100, 7] and right [17, -97, 9]) and the bilateral middle temporal visual cortices (**MT+/V5**; MNI coordinates: left [-48, -69, 7] and right [50, -66, 11]) as these regions are essential for visual motion processing (Born & Bradley, 2005; Plomp et al., 2015). We also included the bilateral posterior parietal cortices (**PPC**; MNI coordinates: left [-46, -46, 54] and right [52, -42, 50]) as these regions are known for higher-level visual motion processing (Ilg et al., 2004). Based on

our 7 identified nodes, we tested 129 models comprising all combinations of modulation directions (beyond the fixed 'minimal model', Figure 5a; see Methods for more information).

Optimised DCM modulation parameters explaining differential ERPs for visible and invisible PE

Among all considered 129 models, we examined which model(s) best explained the ERPs associated with visible and invisible PE. We examined how each connection was modulated by the visibility of motion direction changes, and in the absence of a single winning model, we pooled the parameter estimates from all models weighted by the evidence of each model, a procedure known as Bayesian Model Averaging (BMA). Figure 5a shows the defined model space, including: the fixed model architecture (white arrows) and how each model was allowed to vary in terms of the direction of modulation (black arrows, see Methods for more information). Once we had estimated the log model evidence for every model and participant, we used RFX BMS to determine the winning model at the group-level. Figure 5b shows, for both the visible and invisible PE, the exceedance probability of all models at the group-level. The exceedance probability is the probability that a particular model is more likely than any other model given the group data.

For the visible PE, there was no single clear winning model amongst our 129 models. This is unsurprising when using RFX analysis as effects can be diluted if a large number of models are compared in a small sample size. In this situation, it is recommended not to single out the best architecture alone, but rather to consider all models taking into account the strength of the evidence for each model, a procedure known as a Bayesian Model Averaging (BMA) (Friston et al., 2016; Stephan et al., 2010). BMA computes the connectivity parameters for each participant and averages them by weighting the expected probability of each model (i.e. probability that this particular model generated the data for a randomly selected subject). The parameter outputs from BMA can then be used to make inferences at the group-level. According to the BMA estimates with FDR correction for multiple comparisons (q = 0.05), we found significant positive modulation of the intrinsic self-connections of left MT (mean parameter outputs = +0.2004, two-tailed t-test against 0: *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 to determine the model parameters weighted according to each model's expected probability. We found that, similarly to the visible PE DCM results, RFX BMA showed significant

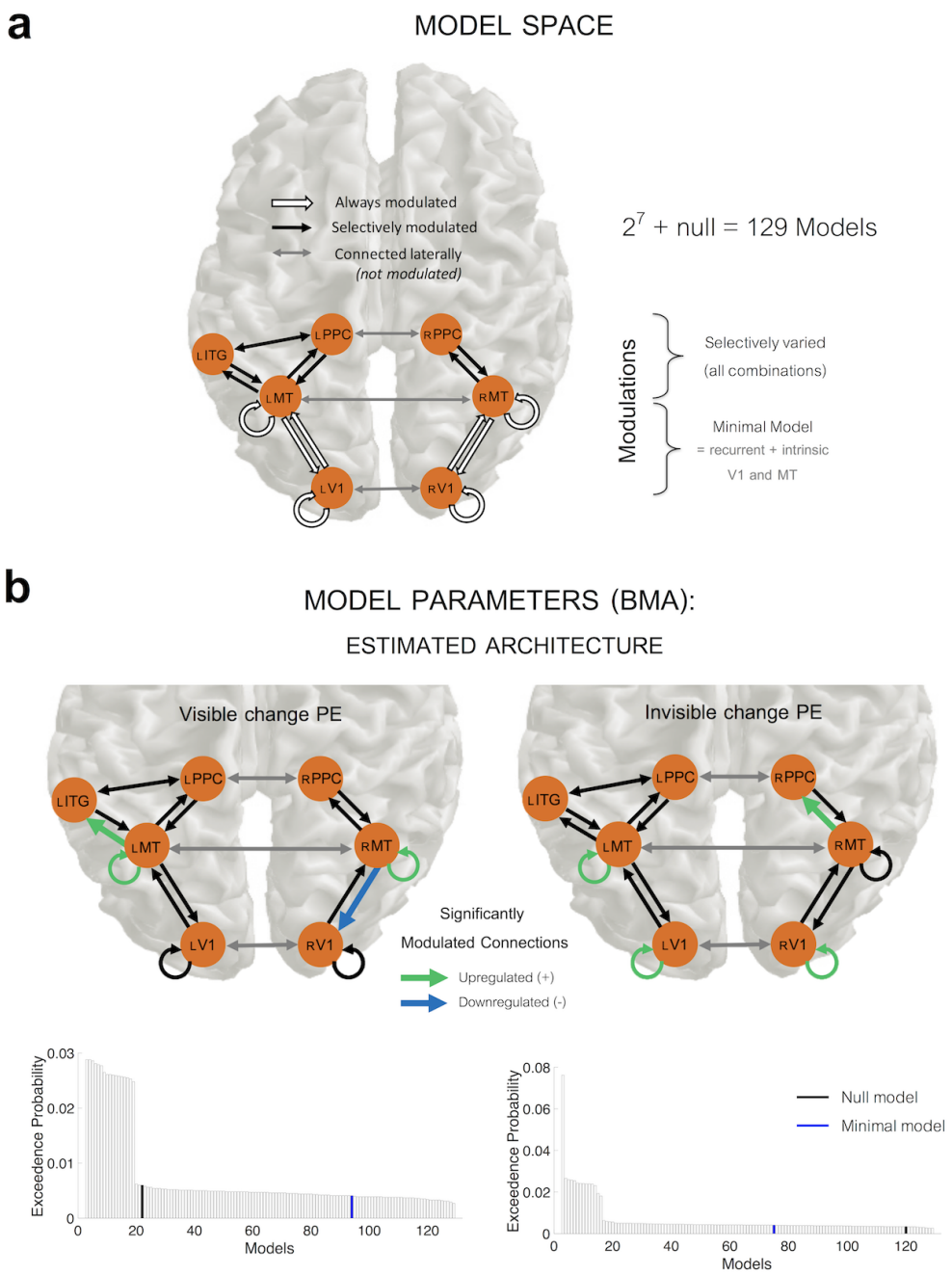

Figure 5 - Dynamic Causal Modelling (DCM) for visible and invisible PE responses. (a) Model space showing the architecture used in all models and modulation of these connections. We used the same model architecture in all models, that is, the same nodes and same connections were "present" in all models. What differed across models were the allowed modulations of the connection strength between nodes. To reduce the model space, we allowed all models to modulate the strength of recurrent (forwards and backwards) connections between V1 and MT+ bilaterally and at intrinsic (within area) connections of bilateral V1 and MT+. We did not allow any modulation of lateral connections between hemispheres. We systematically varied the modulation direction of all black connections for 2^7 = 128 models (+ a null model which did not have any modulation at or between any nodes). "Null model", here, refers to a model, where all connections are present but not allowed to modulate, and in addition, it lacks the modulation between V1 and MT as well as intrinsic modulation within V1 and MT. One of the 2^7 = 128 models also contained "the minimal model", where no

connections were modulated except those between V1 and MT, as well as intrinsic modulation within V1 and MT. (b) Model evidence results from our random-effects (RFX) Bayesian Model Selection (BMS) 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).

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. We applied RFX BMA to determine the model parameters in the absence of one single winning model and found that, after correction for multiple comparisons, there were no significantly modulated connections between any of the nodes.

DISCUSSION

In this study, we aimed to elicit prediction error (PE) related neural activity via changes in stimulus statistics that were either consciously noticeable or not. As planned, we successfully manipulated the awareness of the stimulus changes with supra- and sub-threshold motion coherence of a dynamic random-dot display and confirmed the visibility of motion stimuli with a follow-up psychophysics task. Our ERP analyses (Figure 3) confirmed robust visible PE responses, which were widespread in space and time and invisible PE responses, which were much more confined in space and time. Our source level analysis extended the ERP analysis to locate the source of the visible PE response in the left ITG. In addition to the left ITG, we considered cortical nodes that are known to be critically involved in motion processing (V1, MT, and PPC) and performed DCM analyses, hoping to reveal the network underpinning the generation of visible and invisible PE responses.

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. At the scalp, 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. Our finding is consistent with some studies that elicited non-conscious visual PE. In an MEG study using vertical gratings and backward masking of rapidly presented deviants, Kogai et al. (2011) found similar latencies in the non-conscious PE from 143-154 ms in striate regions. Further, Czigler et al., (2007), backward masked coloured checkerboards at different stimulus onset asynchronies (SOAs) and found

SOAs at 27 ms elicited PE between 124-132 ms at occipital channels. Other studies, however, report later PE to invisible oddballs. For example, Jack and colleagues (2017) reported slightly later "non-conscious" PE (to deviant stimuli presented monocularly to the non-dominant eye) peaked around 250 ms (as did the conscious PE), but the authors could not strictly confirm the invisibility of all of their deviant stimuli. The longer latencies for the non-conscious PE have been echoed by other studies that also could not guarantee an absence of conscious awareness of deviant stimuli (van Rhijn et al., 2013; Jack et al., 2015). We suggest these discrepancies in the latencies arise from both stimulus differences and in the methods used to render stimuli non-conscious, with the primary factor for extended non-conscious PE durations being the inability to confirm the absence of conscious awareness of deviants.

But why does the invisible PE occur earlier than visible PE? Here we offer two possible explanations. The first is related to the effects of attention. Previous visual PE studies that have used consciously *perceivable motion direction* changes found an earlier PE at 142-198 ms when attention was absent (Kremlacek et al., 2006) but a later PE component that emerged when attention was directed to the deviants (Kuldkepp et al., 2013). In our paradigm, participants may have attended to the visible motion resulting in a delayed PE. Conversely, invisible motion may not have deployed attention, which resulted in an earlier PE. However, this explanation does not speak to the neural mechanisms. An alternative explanation lies on a putative difference in neural adaptation for visible and invisible stimuli. It is plausible that adaptation for the visible stimuli suppresses the earlier PE component. This adaptation effect may be weaker for invisible stimuli, although this interpretation does not fully concur with our DCM results showing adaptation in V1 (but not right MT) for invisible stimuli (see below and Figure 5b). Additionally, the later PE component for visible stimuli may be more related to the awareness of prediction violation. Further studies that orthogonally manipulate adaptation and prediction might be able to disentangle these possibilities (Kok et al., 2012; Summerfield & Egner, 2009; Summerfield et al., 2006).

Spatial features of the prediction error (PE) reponses with visible motion changes

In terms of spatial characteristics of visible PE, our source-level analyses (which pools responses over time) provide evidence that visible PE, as well as the main effect of coherence (regardless of the level of surprise), are generated by cortical sources within the left ITG. According to our literature search, we did not encountered much evidence to suggest a critical role of left ITG for awareness of motion direction changes, either as correlated neural activities with neuroimaging in healthy participants or as perceptual deficits due to focal and lateralized brain lesions in patients. Literature on the lateralisation of conscious awareness of specific stimuli to the left hemisphere is scarce; reported only in a handful of studies of emotional processing (Brazdil et al. 2001; Gazzaniga, 2000; Kimura et al., 2004; Meneguzzo et al., 2014; Prete et al., 2015; Shepman et al., 2016; Williams et al., 2006). In these studies, subliminally presented face stimuli activated right amygdala (via a subcortical route) in response to fear but left amygdala for supraliminal fear (Prete et al., 2015; Williams et al., 2006), and using masked affective-stimuli activated the

right hemisphere in response to negative stimuli in both the visual (Kimura et al., 2004) and auditory domains (Shepman et al., 2016). However, unlike these studies, we used visual motion stimuli and examined the difference between PE rather than responses to visual stimuli more generally. One possible explanation for the observed lateralisation of conscious perception of motion changes to left hemisphere comes from work in split-brain patients showing the left hemisphere is more adept at monitoring probabilities to infer causal relationships based on series of events over time (Gazzaniga & Smylie, 1984; Roser et al., 2005) and at creating internal models to predict future events (Wolford et al., 2000). This suggests that when stimuli are consciously perceivable, the left ITG generates and updates predictions and PE. However, due to the lack of literature, specifically on visual PE lateralisation, further work is needed to fully understand the role of left ITG source for conscious PE processing.

Network-level modulations of causal connections investigated by DCM

Using DCM, we intended to extend earlier visual PE studies by investigating the network-level properties underlying the generation of PE to visible and invisible changes. The primary question of our DCM analysis was whether there was evidence for top-down modulations for PE to visible or invisible change. According to the predictive coding framework, when an unexpected stimulus occurs, the PE signal is propagated from lower to higher brain areas, resulting in upregulation of forward connectivity. This, in turn, is followed by the revised prediction from high- to low-level brain areas, resulting in increased feedback connectivity. Previous studies are consistent with this theory that consciously 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 5b). 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 are caused (in part) automatically via a release from adaptation (i.e. repetition suppression) upon deviant motion onset (i.e. change in stimulus statistics) regardless of whether this change is consciously perceived. Findings of the generators of visual PE to motion changes being located in motion centres or the dorsal pathway itself have been suggested previously by other studies of visible motion direction changes (Kremlacek et al., 2006; Pazo-Lavarez et al., 2004). We add to these findings by showing that this still holds when the changes are invisible.

CONCLUSION

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

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.

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.

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 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., 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.

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.

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.

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, 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

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.

Keskes-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.

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. & Naatanen, R. (2013). vMMN for schematic faces: automatic detection of change in emotional expression. *Frontiers in Human Neuroscience, 7*(714), 1-11.

Kremlacek, 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.

Kuldkepp, N., Kreegipuu, K., Raidvee, A., Naatanen, 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

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.

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

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

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. (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), 2331-2355.

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

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

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

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

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

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

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J. & Hirsch, J. (2006) Predictive codes 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., Wilke, M., Frassle, 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., Seitz, A.R. & Watanabe, T. (2008) Task-irrelevant learning occurs only when the irrelevant feature is weak. *Current Biology, 18*(12), R156.

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.

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.