Cerebral Cortex, 2017; 1–12

doi: 10.1093/cercor/bhx087
Original Article

ORIGINAL ARTICLE

Bayesian Mapping Reveals That Attention Boosts Neural Responses to Predicted and Unpredicted Stimuli

Marta I. Garrido[1,2,3,4], Elise G. Rowe[1,2], Veronika Halász[1,2,3] and Jason B. Mattingley[1,3,5]

[1]Queensland Brain Institute, The University of Queensland, 4072 Brisbane, Australia, [2]Centre for Advanced Imaging, The University of Queensland, 4072 Brisbane, Australia, [3]ARC Centre of Excellence for Integrative Brain Function, The University of Queensland, 4072 Brisbane, Australia, [4]School of Mathematics and Physics, The University of Queensland, 4072 Brisbane, Australia and [5]School of Psychology, The University of Queensland, 4072 Brisbane, Australia

Address correspondence to Marta I. Garrido, Queensland Brain Institute, The University of Queensland, Building 79, St Lucia 4072, Brisbane, Australia. Email: m.garrido@uq.edu.au

Abstract

Predictive coding posits that the human brain continually monitors the environment for regularities and detects inconsistencies. It is unclear, however, what effect attention has on expectation processes, as there have been relatively few studies and the results of these have yielded contradictory findings. Here, we employed Bayesian model comparison to adjudicate between 2 alternative computational models. The "Opposition" model states that attention boosts neural responses equally to predicted and unpredicted stimuli, whereas the "Interaction" model assumes that attentional boosting of neural signals depends on the level of predictability. We designed a novel, audiospatial attention task that orthogonally manipulated attention and prediction by playing oddball sequences in either the attended or unattended ear. We observed sensory prediction error responses, with electroencephalography, across all attentional manipulations. Crucially, posterior probability maps revealed that, overall, the Opposition model better explained scalp and source data, suggesting that attention boosts responses to predicted and unpredicted stimuli equally. Furthermore, Dynamic Causal Modeling showed that these Opposition effects were expressed in plastic changes within the mismatch negativity network. Our findings provide empirical evidence for a computational model of the opposing interplay of attention and expectation in the brain.

Key words: EEG, MMN, modeling, novelty, prediction

Introduction

The way in which we perceive the world around us is thought to be an active inferential process. Rather than passively registering information that arrives at our senses, the brain builds predictive models of what it might encounter next. These theoretical conjectures have been formalized in terms of predictive coding (Rao and Ballard 1999; Friston 2005) and are useful in explaining the ubiquitous phenomenon of larger brain responses to surprising than predictable events (Montague 1999; Garrido et al. 2013) (Opitz et al. 1999; Summerfield and Koechlin 2008). Selective attention is the process of prioritizing information by allocating more cognitive resources to the object of focus, while suppressing information that is irrelevant. Recent extensions of predictive coding have framed attention as the process of enhancing the reliability of prediction errors (Feldman and Friston 2010). This idea has been empirically demonstrated by larger prediction errors for attended than unattended visual objects (Jiang et al. 2013) and sounds (Auksztulewicz and Friston 2015), with the latter going against the longstanding notion of mismatch negativity (MMN) as a preattentive process (Naatanen et al. 2001).

There is a general consensus that expectation dampens neuronal activity and that attention boosts neuronal activity (Summerfield and Koechlin 2008). Thus, superficially at least, attention and prediction appear to have opposing effects. However, the way in which attention interacts with expectation is unclear for 2 reasons. First, there have been very few attempts to manipulate attention and prediction independently, but many instances in which the 2 have been entwined or confounded (Summerfield and Egner 2009), as attention is often manipulated in a probabilistic manner rather than through stimulus filtering or prioritization. Second, the few studies on prediction and attention have yielded a puzzling depiction of what might be happening in the brain. Kok et al. (2012) provided fMRI evidence that attention and prediction have an interactive or synergetic effect by showing greater brain activity in the visual cortex for predicted (than unpredicted) visual stimuli, a finding which was conceptually replicated using electroencephalography (EEG) for auditory stimuli, and expressed in the N1 evoked potential (Hsu et al. 2014). By contrast, (Auksztulewicz and Friston 2015) found that attention increased the typically observed difference between evoked responses to unpredicted versus predicted stimuli, as reflected in an enhanced MMN, and Bekinschtein et al. (2009) found that violation of global rules led to late evoked responses only when participants were aware of such violations.

In this paper, we first formalize 2 theoretical models that have been put forward to explain the interplay between attention and prediction in the brain: the Opposition model and the Interaction model, introduced in Kok et al. (2012). The Interaction model postulates that attention and prediction interact such that neuronal activity is greatest for attended and predicted events. This model is inspired by the idea that attention increases the precision of predictions by weighting prediction errors (Feldman and Friston 2010), and assumes 4 levels of precision, or attention, that depend on the level of prediction. By contrast, the Opposition model posits that attention and prediction have opposing effects on neural activity, such that prediction mitigates and attention boosts neural activity. The predictions of this model are that the neuronal responses will be greatest for attended unpredicted stimuli, and smallest for unattended predicted stimuli. Computationally, this model assumes that neuronal activity is weighted by 2 (instead of 4) levels of attention (attended and unattended). This model is agnostic about the relationship between responses to attended predicted and unattended unpredicted events. Both the Interaction and the Opposition models assume that prediction has 2 levels, such that unpredicted stimuli evoke a larger neuronal response than predicted stimuli. They differ, however, in their treatment of the attention component. Specifically, the Opposition model offers a more parsimonious expression of the effects of attention on neuronal responses (Fig. 1).

Here we tested these models empirically using Bayesian model comparison for scalp and source EEG data, as well as dynamic causal modeling (DCM). We developed a novel auditory task in which participants were presented with independent streams of white noise concurrently in each of the 2 ears, and were instructed to attend to the left channel, the right channel or both channels in separate blocks to detect brief gaps in the noise streams. At the same time, an irrelevant stream of standard and deviant tones was presented in either ear (attended or ignored), providing an orthogonal stimulus set from which to extract neural responses to predicted and unpredicted auditory events.

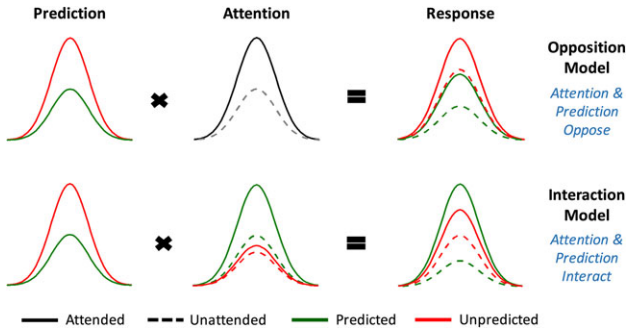

Figure 1. Two competing models for the relationship between attention and prediction. In the Opposition model, predicted (green) and unpredicted (red) neural signals are multiplied by 2 levels of Attention, with attended stimuli (solid lines) receiving a greater boost than unattended stimuli (dashed lines). In the Interaction model (proposed by Kok et al. 2012), predicted and unpredicted signals are multiplied by 4 (instead of 2) levels of attention that depend on the level of the prediction.

Methods

Participants

A total of 21 healthy adults were recruited for the experiment. Data from 2 participants were excluded from further analysis due to poor performance on the behavioral task (accuracy < 50%). The reported analysis was thus performed on data from 19 participants (10 females, aged 19–43, $M = 24.21$, standard deviation = 6.11) with no reported history of neurological or psychiatric disorder and no previous head trauma resulting in unconsciousness. All participants gave written informed consent in accordance with the guidelines of the University of Queensland's ethical committee, and were monetarily compensated for their time.

Auditory Stimuli

The auditory task developed for the study is depicted in Figure 2. An auditory frequency oddball sequence was played to one ear at 60 dB and overlayed with Gaussian white noise at 40 dB. White noise only was played to the other ear at 40 dB. Two pure tones, standards ($P = 0.85$) and deviants ($P = 0.15$), ($f = 500$ or 550 Hz; counter-balanced between blocks) of 50 ms in duration were played with an inter-stimulus interval of 450 ms. Embedded in the white noise of either ear were 2 types of targets: a total of 30 nonoverlapping randomized periods of no sound (gaps), which could be singular (90 ms gaps only, 15 per block) or doubled (two 90 ms breaks separated by a 30 ms white noise return, 15 per block). The gaps in the white noise of either ear were never within 2.5 s of each other and never occurred at the same time as a tone. Importantly, the presentation timings for the noise gaps and the tones were uncorrelated in order to avoid any systematic effects of bottom-up attention that could have otherwise confounded the ERPs to the tones. All auditory stimuli were created using in-house Matlab scripts, recorded using Audacity Sound Mixer prior to the experiment, and delivered with inner-ear buds (Etymotic, ER3).

Experimental Design

The twelve experimental trial blocks ($T = 3{:}32$ min each) were comprised of a total of 380 tones (with deviants always falling within 4–10 standard tones). Participants were instructed to

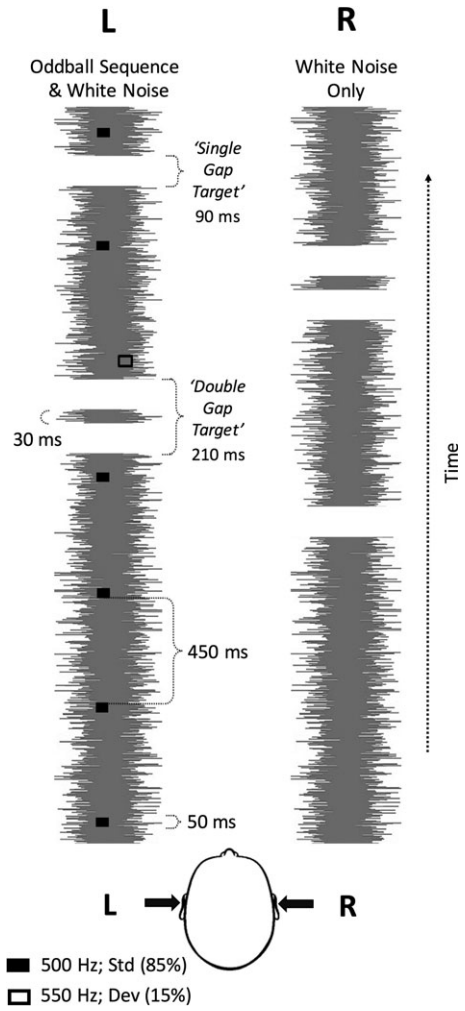

Figure 2. Experimental paradigm. Gaussian white noise embedded with single (90 ms) or double (210 ms) noise gaps (periods of silence, and the targets of this experiment) was played to both ears (different target sequence in each ear). One ear received the oddball sequence of pure tones (50 ms) at either 500 or 550 Hz (counter-balanced between blocks) (ISI = 450 ms, standard $P = 0.85$ black rectangle, deviant $P = 0.15$ hollow rectangle, respectively). Participants were instructed to pay attention to the targets embedded in the white noise in the left, right, or both ears and to ignore the tones. ISI = inter-stimulus interval; L = left ear; R = right ear; Std = standard; Dev = deviant.

listen for and report target gaps within the white noise stream in either the left channel only, the right channel only, or in either channel (divided attention), and to ignore the tones. Each attention condition was repeated 4 times and the order of the blocks was pseudo-randomized such that no participant received the same order. When a target was identified in the attended ear(/s) participants responded with a "1" keypress if the gap was singular and a "2" keypress if the gap was doubled. In one-third of the blocks oddball tones were played in the attended ear, in another third the tones were played in the ignored ear, and in the remaining third, in which participants divided their attention between ears, the tones were presented to either side, counter-balanced between the left and right across separate blocks. Participants performed all blocks in one testing session of 60 min (42:24 min total task duration plus breaks) with an additional 30 min EEG setup period.

Task

Participants were seated in front of a computer screen and wore inner-ear buds for the duration of the experiment. Prior to recordings, participants listened to an example auditory stream of 1-min duration, which demonstrated the single and double gaps in the white noise. Each participant then underwent a brief practice session with auditory stimuli consisting of 9 single and 9 double gaps, and a total of 110 tones. Participants were given feedback about their accuracy in this practice block but not in the experimental blocks. At the beginning of each experimental block, the focus of attention was specified verbally and an arrow (left, right, or both directions) remained on the screen for the duration of the block as a reminder. Participants were asked to make their keypresses in response to target gaps as quickly and as accurately as possible, and to ignore any gaps in the uncued ear (in the focused attention condition). Task performance was assessed based on the percentage of correctly detected target gaps and reaction times. Participants with <50% overall accuracy (proportion correct) were excluded from further analysis.

EEG Data Acquisition and Preprocessing

Continuous EEG data were recorded with a Biosemi Active Two system with 64 Ag/AgCl scalp electrodes arranged according to the international 10-10 system for electrode placement using a nylon head cap. Data were recorded at a sampling rate of 1024 Hz. Preprocessing and data analysis were performed with SMP12 (http://www.fil.ion.ucl.ac.uk/spm/). Data were rereferenced to a common reference, down-sampled to 200 Hz and high-pass filtered above 0.5 Hz. Eye blinks were detected and marked using the VEOG channel before the data were epoched offline with a peristimulus window of −100 to 400 ms. Artefact removal was performed by removing trials marked with an eyeblink and by thresholding all channels at 100 uV. Trial data were robustly averaged before being low-pass filtered below 40 Hz and baseline corrected between −100 and 0 ms. We analysed event-related potentials with respect to the onsets of standard and oddball tones, separately for conditions in which the tones were presented in the attended ear, the unattended ear, or in either ear in the divided attention condition.

Spatiotemporal Image Conversion

Event-related potentials were converted into 3D spatiotemporal volumes per condition and participant. This was achieved by interpolating and dividing the scalp data per time point into a 2D 32 × 32 matrix. We obtained one 2D image for every time bin (from 0 to 400 ms in steps of 5 ms). These images were then stacked according to their peristimulus temporal order, resulting in a 3D spatiotemporal image volume with dimensions of 32 × 32 × 81 per participant. Data were then smoothed at FWHM 12 × 12 × 20 mm^3.

Spatiotemporal Statistical Maps

For each participant, the 3D spatiotemporal image volumes were modeled with a mass univariate general linear model (GLM) as implemented in SPM12. We performed between-subject F-contrasts for (1) the main effect of attention, (2) the main effect of prediction, and (3) the interaction between attention and prediction. Simple effects were estimated using between-subject t-statistic contrasts. The same statistical analyses were performed on the 3D spatial image volume obtained

after source localization (see below). All sensor effects are reported at a threshold of $P < 0.05$ with family-wise error (FWE) correction for multiple comparisons over the whole spatiotemporal volume. For closer inspection of the main effects and interactions obtained at channel Fz (at which predictability effects are typically strongest, Naatanen and Alho (1997)), we implemented a 1D GLM approach using SPM12. We restricted our time window from 0 to 400 ms after stimulus onset and, in a separate analysis, between the typical MMN time window of 100–250 ms (FWE corrected over the time bins considered).

Source Reconstruction

We obtained source estimates on the cortical mesh by reconstructing scalp activity with a Boundary Element Method (BEM) and a standard MNI template for the cortical mesh, in the absence of individual MRIs. This forward model was then inverted with multiple sparse priors (MSP) assumptions for the variance components under group constraints. This allowed for inferences on the most likely cortical regions that generated the sensor-level data. We obtained images from these reconstructions for each of the 6 conditions in every participant. These images were smoothed at FWHM $12 \times 12 \times 12\,mm^3$. We then computed the main effects of attention and prediction, and the interaction (attention × prediction) using conventional SPM analysis. The effect of prediction (*t*-statistic) is displayed at an uncorrected threshold of $P < 0.001$. These weaker significance criteria were used for post-hoc visualization, once the effects had been established under robust criteria at the scalp level, and we only report regions significant at $P < 0.05$ FWE corrected at the cluster level.

Statistics

Significance sensor space maps for prediction effects are displayed at $P < 0.05$ corrected for multiple comparisons using family-wise error rate. The interaction map is displayed at $P < 0.01$ uncorrected for the purpose of defining a region of interest for follow up Bayesian Model Selection (BMS). Source maps are displayed at $P < 0.001$ uncorrected, but only significant cluster-level $P_{FWE} < 0.05$ are reported.

BMS was employed to make inferences on both scalp and source maps, as well as on DCMs. Note that this framework uses model evidence as a relative (probabilistic) measure for how well one model explains the data relative to another, considered in the model space. Importantly, model evidence seeks the optimal balance between accuracy and model complexity, by favouring the former and penalizing the latter.

Bayesian Model Selection

To compare the 2 models (Opposition and Interaction; see Introduction) of the effects of attention on prediction (standard and deviant tones) we used the BMS methodology described in Rosa et al. (2010), and adapted here for EEG. For this analysis we discarded trials from the divided attention condition and used only the attended and unattended trials from the focused attention conditions (attend left ear only, attend right ear only) for both standard and deviant tones. We created posterior probability maps (PPMs) from individual participant log-model evidences using a random-effects approach (RFX). Here, the winning model was the one with the highest log-evidence (assuming uniform priors over the models) across participants. We performed this analysis at the sensor and source levels by modeling the data with regressors describing the hypothesized relationships amongst the 4 different conditions.

Briefly, covariate regressor weights were applied to every participant and trial under the Opposition model, which predicts reductions in ERP amplitudes across conditions in the following order: (1) attended unpredicted, (2) unattended unpredicted/attended predicted, and (3) unattended predicted. Next, we specified a second model derived from Kok et al. (2012), the Interaction model, which predicts reductions in ERP amplitudes across conditions in the following order: (1) attended predicted, (2) attended unpredicted, (3) unattended unpredicted, and (4) unattended predicted. Voxel-wise whole-brain log-model evidence maps were then created for every participant and model, estimated using the Variational Bayes first-Level Model Specification methodology described in Penny et al. (2005). Source level maps were further smoothed with a 1 mm half width Gaussian kernel. We used the RFX approach to produce PPMs for both models at the group-level. These maps (displayed at a threshold of probability larger than 75% and 50% for scalp and source, respectively) allowed us to compare which model had the higher probability at each voxel in the brain (and at each time point in the scalp level analysis). Further model comparisons for specific regions at the sensor level were undertaken using brain regions selected a priori from the attention by prediction interaction contrast. At the source, these comparisons were made at the peak coordinates of clusters for each model that exceeded 51%.

Dynamic Causal Modeling

Source locations were identified based on multiple sparse priors source reconstruction of the overall mismatch ($P < 0.05$ uncorrected threshold). These regions were: bilateral primary auditory cortices (A1; MNI coordinates: left [−42, −24, 34] and right [44, −22, 38]), bilateral inferior temporal gyri (ITG; MNI coordinates: left [−42, −10, −38] and right [44, 0, −42]) and left inferior frontal gyrus (LIFG; MNI coordinates: [−50, 32, 0]). The choice for the DCM source nodes was undertaken in a data-driven fashion, rather than being based on a priori models drawn from the literature. Previous papers have consistently tested models with A1, STG, and IFG, nodes that we first proposed in 2007 based on a number of fMRI and MEG studies (Garrido et al. 2007). The similarity of the model space proposed here and in previous papers is evident, except for the replacement of STG with ITG. This was motivated by the strong evidence for ITG both in the standard SPM analysis at the source level ($P < 0.05$, FWE-cluster corrected), and the posterior probability maps (probability >80%). It is important to note, however, that while some papers have tested models pertaining to the presence or absence of nodes such as STG and IFG, all assumed that these nodes were correct, given previous literature, rather than refining the model progressively through exploring other candidate nodes (e.g., ITG instead of STG) through DCM optimization or other forms of source reconstruction. Given the strong evidence from our source reconstructed data and the posterior probability maps, as well as the novelty of the paradigm, here we took a data-driven approach rather than adhering to an assumed model specification. Nevertheless, we ran a validation check to rule out the possibility that our source reconstructed nodes were less reliable than the nodes taken a priori from the literature. BMS revealed that the models with the source reconstructed nodes outperformed the models with a priori nodes by 60% probability.

Note that in the absence of individual anatomical landmarks, we used a standard MNI template for the cortical mesh in our source reconstruction, which was then used to identify

candidate nodes for the DCMs. Whether including anatomical information would improve the source reconstruction results at the group level is unclear. This raises an interesting model comparison related to that addressed in Mattout et al. (2007); Henson et al. (2009), who showed that individual MRI does not add to the precision of source estimates compared with an individual deformed template. This was done for MEG data, however, and it is unclear what the impact on EEG might be when using an MNI template without individual deformations. Given that MEG has higher spatial resolution and is more sensitive to approximations in source models, however, it is likely that any potential benefit afforded by individual MRIs would be smaller (not larger) for EEG than for MEG (shown to be negligible). Furthermore, the sensitivity of our group level inference yielded a reconstruction of the expected brain regions underlying the MMN (within the temporal and inferior frontal cortex), even in the absence of a highly realistic head model. Importantly, the locations of the source reconstruction were only used as soft priors in the subsequent DCM analysis, so that source locations could be adjusted individually during the connectivity estimation procedure.

We first optimized the basic connectivity architecture using responses to attended and unattended standards and deviants with no between-trial effects present. This first step considered 2 competing model structures that included bilateral A1 and ITG, but differed in the presence or absence of LIFG. Next, the pattern of changes in extrinsic connectivity was optimized under the fully connected architecture (the winning model) using responses in all 4 conditions for the Opposition and Interaction models. The family of Opposition models used a between-trial effect of [1, 2, 2, 3] for the attended predicted, predicted attended, unpredicted unattended, and attended unpredicted, respectively. The family of Interaction models, on the other hand used [1, 2, 3, 4] for predicted unattended, unpredicted unattended, unpredicted attended, and attended predicted, respectively. The choice of the weights for the between-trial effects was motivated by the theoretical relationship proposed in Figure 1. It is important to note that there is an infinite number of possible combinations of weights that could satisfy the general ordinal relationship between the 4 conditions specified in Figure 1. Here, we assumed a linear relationship, in the absence of theoretical or empirical evidence to assume an otherwise more complex relationship. Specifically, for the Opposition Model, we compared models using [1 2 2 3] or [1 3 3 4], as we had no reason to believe that the attended predicted and the unattended unpredicted conditions would be closer to the unattended predicted condition than to the attended unpredicted condition. Bayesian model comparison revealed that the former outperformed the latter.

Fifteen competing models were tested, each with a different subset of connections—forward (F), backward (B), and recurrent (R)—which also included (subscript i) or excluded intrinsic modulations of A1, and a single null model. Finally, the Opposition and Interaction model-dependent changes in intrinsic connectivity were then grouped by families, under the optimized connectivity architecture. In both DCM estimation steps, models were inverted using a 0–400 ms peristimulus time window.

Results

Behavioral Findings on Attentional Manipulation

Behavioral results for the target detection task—discriminating single- and double-gaps in concurrent white noise streams in each ear—were grouped into unilateral (focused) or bilateral (divided) attention conditions (30 targets over 8 blocks and 60 targets over 4 blocks, respectively). We excluded any participants who did not achieve mean response accuracy >50%. There was no significant difference in response accuracy (P = 0.14) between the unilateral (M = 71.80%, standard error of mean [SEM] = 5.19%) and the bilateral (M = 68.33%, SEM = 5.13%) conditions. Participants were significantly faster (P = 0.03) to respond in the bilateral (M = 748.16 ms, SEM = 27.67 ms) than the unilateral conditions (M = 779.79 ms, SEM = 34.13 ms), likely reflecting a strategy of responding immediately to any target gap when monitoring both ears under divided attention, as opposed to having to select only relevant gaps in the focused attention conditions (filtering out gaps in the ignored ear).

Attention Amplifies Prediction Errors—Single Channel Analysis

ERPs corresponding to each of the experimental conditions (as well as the MMNs derived from subtracting the standards from the deviants within a condition) were extracted from electrode Fz and compared over time (Fig. 3). The N1 and P2 components were examined as an average across participants and conditions. For this, the lowest time point between 50 and 150 ms and highest point between 150 and 250 ms were determined from the omnibus ERP plot (i.e., the mean ERP across all participants and conditions over time). These time indexes ±25 ms were then used to find the average ERP per condition. Statistical tests of the N1 components found only a main effect of surprise (F[1,72] = 4.9583, P = 0.0291). Similarly, at P2, there was a main effect of surprise (F[1,72] = 17.5898, P = 7.7001e-05), but no further significant main effects or interactions. In addition, results at Fz from 0 to 400 ms using the 1D GLM approach revealed a significant main effect of Attention between 290 and 340 ms ($P_{FWE\_cluster}$ = 0.006), and a significantly larger prediction error for attended relative to unattended conditions at 115–120 ms ($P_{FWE\_cluster}$ = 0.020). We then restricted our analysis to the MMN time window (100–250 ms) and again found a significant main effect of Attention but at an earlier period between 200 and 230 ms ($P_{FWE\_cluster}$ = 0.028). Moreover, there was a significantly larger prediction error for attended versus unattended conditions between 100 and 130 ms ($P_{FWE\_cluster}$ = 0.046). These findings demonstrate that attention amplifies prediction errors.

Larger Responses to Unpredicted Than Predicted Events Regardless of Attention Level—Sensor and Source Space

As shown in Figure 4A, the main effect of Prediction, or surprise (standards vs. deviants), disclosed several significant components comprised of 2 late effects. The first late effect was detected from 200 to 220 ms (peak-level T_{max} = 8.30, cluster-level P_{FWE} < 0.001; at frontocentral channels). The second late component was observed from 290 to 295 ms (peak-level T_{max} = 5.02, cluster-level P_{FWE} = 0.004; at right parieto-occipital channels). We also found simple Prediction effects in all of the attention manipulations (Fig. 4B), that is, attended (peaking at 185 ms), unattended (peaking at 210 ms), and divided (peaking at 195 ms). While there appeared to be qualitatively differences in the strength and extent of the prediction effects across Attention conditions, the interaction between Attention and Prediction did not survive correction for multiple comparisons.

We then used a multiple sparse priors source reconstruction method to investigate the cortical regions that generated the effects at the scalp level. Statistical parametric maps for

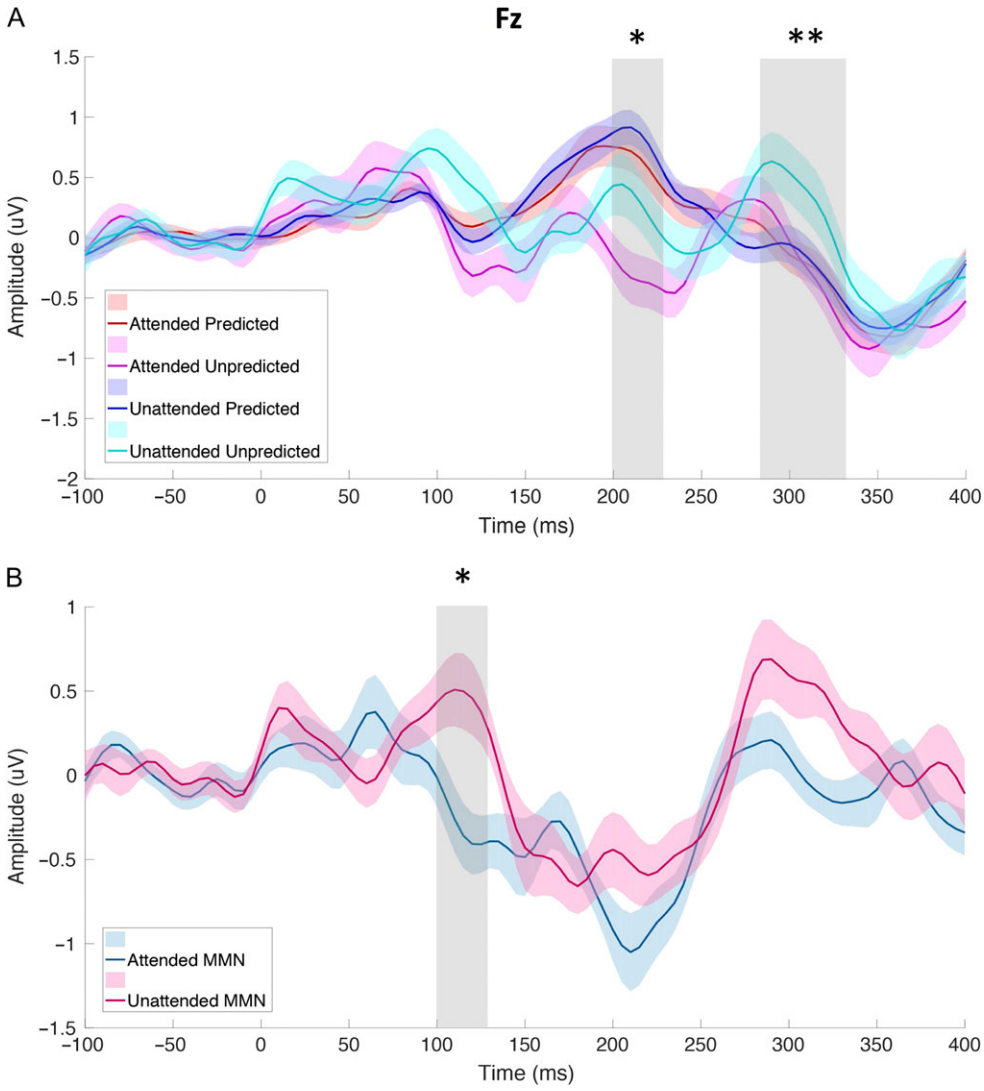

Figure 3. Event-related potentials extracted from electrode Fz for each condition (mean/SEM). (*A*) The ERPs for each of the experimental conditions were extracted from electrode Fz and compared over time. The grey shadings indicate the temporal widows during which a significant main effect of attention was found (**corrected for the whole epoch, *corrected within the a priori MMN time window). (*B*) ERPs for attended and unattended prediction errors (the MMNs; i.e., the difference between unpredicted and predicted) are plotted at electrode Fz. Grey shading indicates the temporal window during which a significant Attention by Prediction interaction was found (*corrected within the a priori MMN time window).

source-reconstructed images revealed 2 significant clusters for the main effect of Prediction in the left ([−42 −10 −38], peak-level T_{max} = 4.14, cluster-level P_{FWE} = 0.019) and right inferior temporal gyri ([44 0 −42], peak-level T_{max} = 3.77, cluster-level P_{FWE} = 0.023) (Fig. 4*C*).

Opposition Wins Over Interaction—Evidence From Posterior Probability Maps

Scalp level

BMS was used to compare the 2 competing models of the relationship between Attention and Prediction (the Opposition or Interaction models; see Fig. 1). Specifically, we were interested in comparing the strength of neural activation under the different manipulations of attention and prediction. We used random effects BMS to create group-level PPMs for each model, derived from the log-model evidence of each participant, that is, the evidence that a given model (Opposition or Interaction) generated the data.

As shown in Figure 5, BMS revealed that the Opposition model ("Attention and Prediction oppose") was the more likely (>75% model probability) explanation for the data across most frontocentral channel locations at the majority of time points (70–210 and 290–375 ms). However, the Interaction model ("Attention and Prediction interact") had a higher probability (>75%) of explaining the data between 170 and 230 ms (i.e., within the MMN time window) at central and lateral parietal channel locations. Thus, the relationship between Attention and Prediction differed depending on both the time point and scalp location; although more often than not, Attention and Prediction had opposing effects.

The fact that the Interaction model won within the MMN window and yet we did not find a significant interaction in the classic GLM analysis could perhaps be explained by a Prediction by Attention interaction effect that did not survive

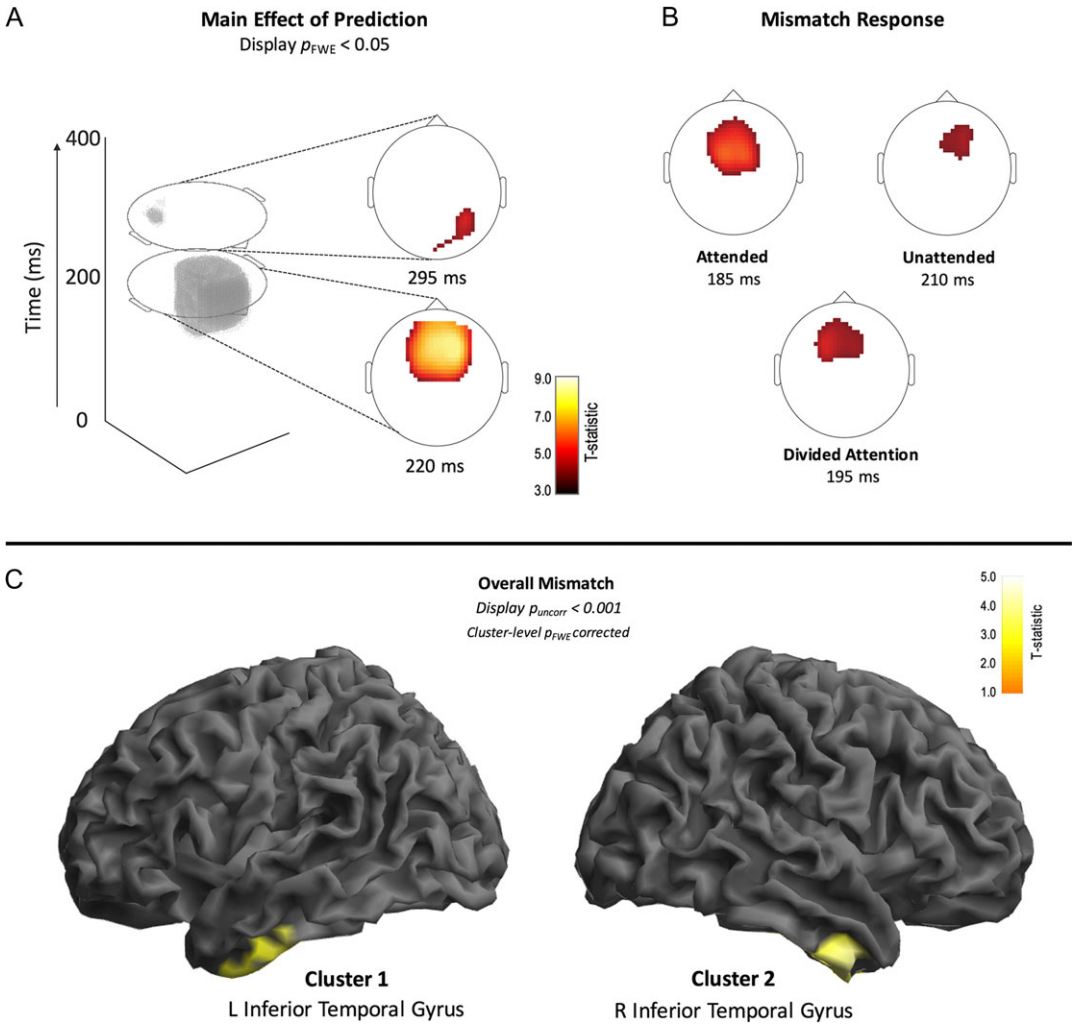

Figure 4. Main and simple effects of prediction at the scalp and source levels. (A) Spatiotemporal statistical analysis revealed significant effects of prediction (predicted vs. unpredicted) over frontocentral areas around 220 ms and over posterior parietal areas at 295 ms (displayed at $P < 0.05$, FWE whole-volume corrected). (B) The effects of prediction across the 3 attentional manipulations revealed a prediction effect in the attended condition at 185 ms, the divided attention condition at 195 ms, and in the unattended condition at 210 ms, all located frontocentrally (displayed at $P < 0.05$, FWE whole-volume corrected). There was no significant interaction (difference in the MMN between the attention conditions). (C) Source reconstruction analysis revealed a main effect of prediction within the left and right inferior temporal gyri. (Displayed at $P < 0.001$ uncorrected and FWE corrected at the cluster-level.)

correction for multiple corrections. We further examined a potential interaction effect, hindered perhaps by a rather conservative multiple comparison correction procedure. Firstly, we used more lenient, uncorrected peak-level statistics to select 2 small interaction clusters at 175 ms (peak-level $F_{max} = 5.79$, peak-level $P_{uncorr} = 0.004$; at central channels) and 360 ms (peak-level $F_{max} = 5.45$, peak-level $P_{uncorr} = 0.006$; at right parietal channels—see Fig. 6). We then took the spatiotemporal coordinates of these clusters and extracted the posterior probability of each model at that particular location. We constructed a 10^3 cube around these coordinates and took the average posterior probability of each model over that volume. Our reasoning was that if an interaction between Attention and Prediction were present in the data, then the Interaction model would have a higher posterior probability compared with the Opposition model at these coordinates. We found that at 175 ms over frontocentral channels there was a negligible difference between the Opposition and Interaction models, with 48% and 52%, respectively (Fig. 6). However, at 360 ms over the right lateral parietal area, the Opposition model probability far exceeded that of the Interaction model, with a value of 80%. Thus, Attention and Prediction appear to have opposing effects later in time.

Source Level

Finally, we applied the same BMS technique employed at the sensor level to our source reconstructed results. BMS revealed that the Opposition model had the higher model probability and larger clusters at the source (Fig. 7). The Opposition model achieved >50% model probability in the left middle temporal gyrus (cluster size; $K_E = 82$) and right inferior temporal gyrus (cluster size; $K_E = 288$). Conversely, the Interaction model achieved > 50% model probability in a smaller cluster in the left middle temporal gyrus (cluster size; $K_E = 32$). We then compared the model probabilities at the center of these clusters and showed that the Opposition model was more probable than the Interaction model in the left middle temporal and right inferior temporal gyri (winning with 82% and 78% probability, respectively). Furthermore, model probabilities extracted

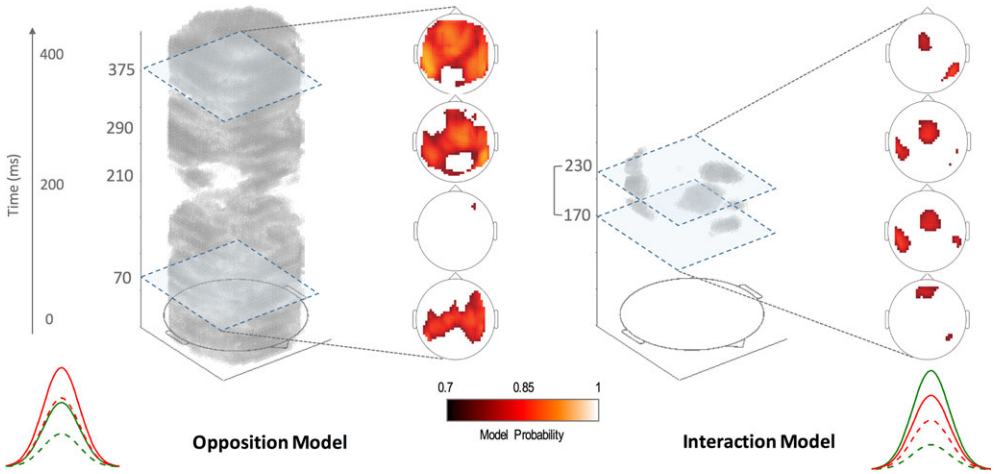

Figure 5. Scalp Posterior Probability Maps of the Opposition and Interaction models over space and time. Maps display the posterior probability for both models, thresholded at probability >75% over space and time. Scalp maps show the 4 time points with the largest significant clusters. The Opposition model wins (Attention and Prediction oppose) across most frontocentral channels at the majority of time points (70–210 and 290–375 ms). The Interaction model wins (Attention and Prediction interact) at the frontocentral and lateral parietal regions of the scalp (channel locations) between 170 and 230 ms.

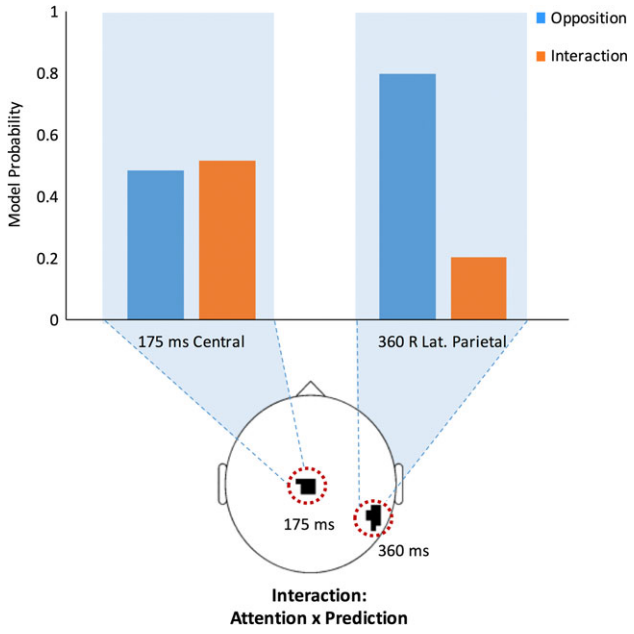

Figure 6. Bayesian Model Comparison within the spatiotemporal clusters extracted from the Prediction by Attention interaction. We extracted model probabilities using the coordinates (scalp location and time points) of 2 clusters from the Interaction results (based on the liberal threshold of $P < 0.001$ uncorrected). If interaction effects were present, the Interaction model would be more likely to win over the Opposition model at these coordinates. At 175 ms (within the MMN time window) and over central electrodes, there was a very slight advantage for the Interaction over the Opposition model. However, at 360 ms over right lateral parietal channels, the Opposition model probability far exceeded that of the Interaction model, with a probability of 80%.

from the peak of the Interaction model cluster showed only a slight advantage for the Interaction over the Opposition model (with 57% probability for the Interaction model) in the left middle temporal gyrus. Such a small difference between the probability of the Interaction model over the Opposition model at this cluster suggests we should be cautious in drawing any strong conclusions about its functional anatomy.

Dynamic Causal Modeling

The prior location of the cortical sources included in our DCMs was based on MSP source reconstruction of ERPs corresponding to the 4 conditions (attended standards, attended deviants, unattended standards, and unattended deviants) of the Overall Mismatch. Statistical parametric maps were inspected at a more liberal threshold of $P < 0.05$ (uncorrected) to identify candidate neural sources of the effects observed on ERP amplitude for the DCM analysis (Auksztulewicz and Friston 2015). Following the selection of candidate sources, model structure was optimized by comparing 2 alternative connectivity models using data from each of the experimental conditions, with or without bilateral connections between the left inferior temporal and inferior frontal gyri, with no between trial effects present. Results indicated that the best model included recurrent connection amongst all regions, that is, inputs to LA1 and RA1, with LA1 connected to LITG, and LITG connected to LIFG, as well as connections linking RA1 and RITG, and lateral connections between LITG and RITG. The selected model was then used to further optimize condition-specific changes in the extrinsic connectivity by comparing the types of extrinsic connections present. Next, 15 competing models were tested (Fig. 8A, B), each with a different subset of condition-specific modulations of connections, according to the Opposition and Interaction models on forward (F), backward (B), and recurrent (R) connections (with, i, and without intrinsic modulations of A1), as well as a null model precluding any modulations (N). These models were fitted to each participant's data to explain observed differences in ERP amplitude. Random-effects BMS revealed that the Opposition model with modulation of forward connections outperformed all other models (Fig. 8C).

Discussion

In this study, we adjudicated between 2 alternative computational models of the effect that spatial attention has on expectations. Using Bayesian model comparison of scalp PPMs we found that, except for an early time window (within the typical MMN), the Opposition model won over the Interaction model. This suggests that, for the most part, attention provides an equivalent boost to neuronal responses to predicted and

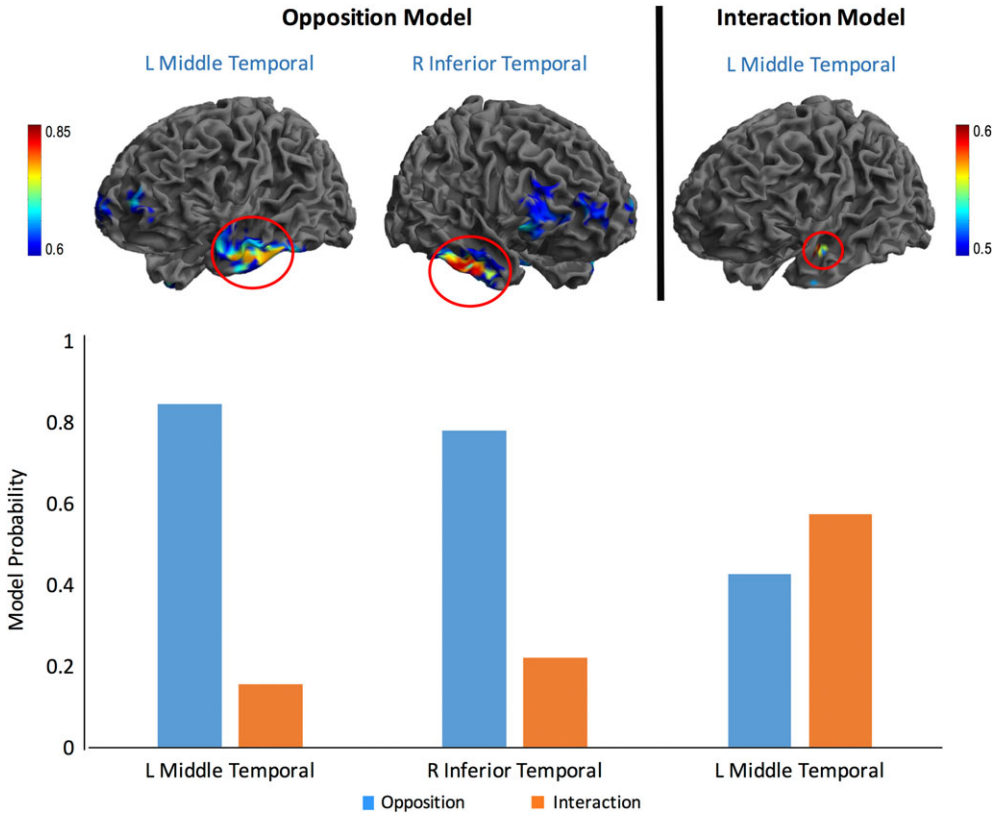

Figure 7. Source posterior probability maps of the Opposition and Interaction models (top) and model probabilities for the 3 major clusters of the 2 models (bottom). BMS was used for model inference at the group-level for the source reconstructed images. Here, the Opposition model achieved >50% model probability in the left middle temporal (cluster size; $K_E = 82$) and right inferior temporal (cluster size; $K_E = 288$) gyri. The Interaction model achieved >50% in a small cluster in the left middle temporal gyrus (cluster size; $K_E = 32$). Overall, the Opposition model achieved higher probability over a larger number of voxels. Extraction of model probabilities from the peak of the Opposition clusters showed that this model won with 82% probability in the left middle temporal gyrus and with 78% probability in the right inferior temporal gyrus. Model probabilities extracted from the peak of the Interaction cluster showed minimal differences between either model at this location, with 57% posterior probability for the Interaction model. Note the differences in the color map scales between the Opposition and Interaction models.

unpredicted stimuli. Similarly, at the source level we found stronger evidence for the Opposition model underlying a frontotemporal network. We investigated this further with DCMs that employed trial-dependent plastic changes according to either the Opposition or the Interaction model. In agreement with the model-based scalp and source analysis, we found that the family of Opposition models better explained the data. Classic SPM analysis of spatiotemporal maps revealed an effect of prediction across and within all attentional manipulations, which peaked within the typical MMN time window and at frontocentral channels. This effect was statistically greater in the attended compared with the unattended conditions at the single channel level, where MMN is typically seen, suggesting that attention amplifies prediction errors. At the whole spatiotemporal map level, however, this interaction effect did not survive correction for multiple comparisons over the whole space-time, despite the appearance of somewhat larger clusters for the attended than the unattended condition,

Our finding of a prediction error effect in all attention conditions (attended, unattended, and divided) is in agreement with a vast body of work suggesting that the MMN is elicited regardless of attention, and hence is "pre-attentive" in nature (Naatanen et al. 2001). This is in contradistinction to Auksztulewicz and Friston (2015), who did not find an effect of prediction in the absence of attention (although this might have been due to a lack of power, as very few trials were included). Again, our finding of a prediction error effect regardless of attention is opposite to Todorovic et al. (2015), who found that while beta synchrony decreased with expectation in the unattended condition, no difference was found in the attended condition. The latter is seemingly at odds with the idea that attention amplifies prediction errors as previously shown (Jiang et al. 2013; Auksztulewicz and Friston 2015), and as revealed in the current study. A number of factors could explain such conflicting results. Perhaps most importantly, very different paradigms and measures were employed across the relevant experiments. Both our study and that of Auksztulewicz and Friston (2015) investigated the effects of attention and prediction on evoked responses in an oddball paradigm, whereas Todorovic et al. (2015) focused on endogenous oscillatory activity. Moreover, both Auksztulewicz and Friston (2015) and Todorovic et al. (2015) manipulated temporal attention, whereas here we manipulated spatial attention. Finally, in our experiment attention and prediction were manipulated within the same spatial location (left or right ears), but were drawn toward independent auditory "objects" (noise for the attention task, and tones for the concurrent oddball stream). By contrast, the aforementioned studies (and that of Kok et al. (2012)) manipulated attention and prediction within the same (visual or auditory) object. It is possible that our attention manipulation, based on spatial selectivity, had a small effect on the tones (in the attended condition), given that these were task-irrelevant and that they

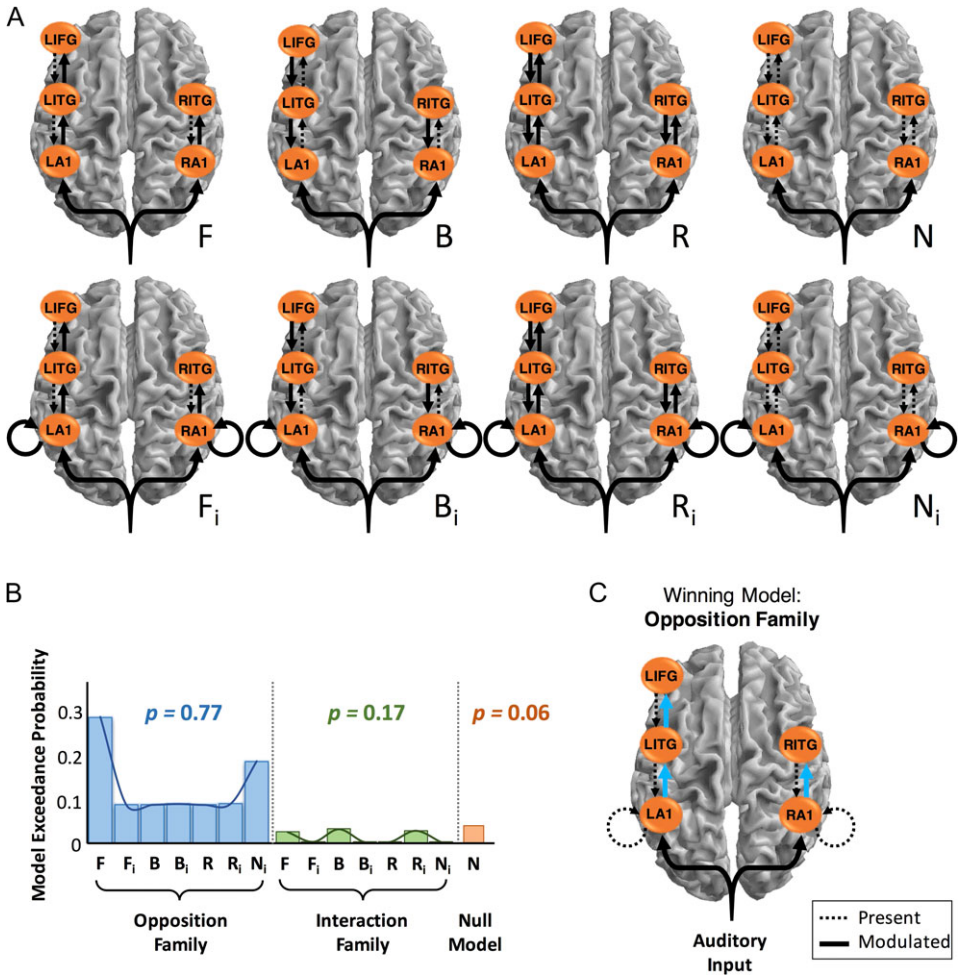

Figure 8. Dynamic Causal Modeling hypotheses testing for plastic changes according to the Opposition and Interaction families of models. (*A*) Eight model architectures were considered, which tested for trial-specific modulation in forward (F), backward (B), and recurrent (R, i.e., both forward and backward) connections, as well as a null model precluding any modulations (N) in extrinsic connections. These models were considered with (bottom row) and without (top row) intrinsic modulations of A1 (subscript *i*). The nodes in the model included bilateral primary auditory cortices (LA1 and RA1), bilateral inferior temporal gyri (LITG and RITG) and left inferior frontal gyrus (LIFG). (*B*) Extrinsic connectivity was optimized using responses in all 4 conditions under the Opposition or Interaction model. A total of 15 competing models were tested, each with a different subset of condition-specific modulation of connections, according to the Opposition (blue) and Interaction (green) models on F, F_i, B, B_i, R, R_i, and N_i. N did not include any modulations (orange). Summed model exceedance probabilities across each family show the winning family as the Opposition Family (left; blue). (*C*) The winning model architecture had recurrent connections between all regions, intrinsic modulation of A1, lateral connections between bilateral ITG, and included condition-dependent modulations according to the Opposition model in the forward connections (blue lines).

never occurred at the same time as the task-relevant noise gaps. However, we believe that this is improbable for 2 reasons. First, the onset of the noise gaps was unpredictable and hence participants had to constantly monitor the stream of sounds on the task-relevant side of space. Second, it is unlikely that participants learned that the noise gaps never coincided with the tones, and could therefore momentarily disengage attention from the noise task. Having said that, the possibility remains that by having the participants focus on the noise streams instead of the tones, our attention manipulation might not have influenced the neural representations of the tones as much as it would have, had we asked the participants to focus on the tones. Future work should test whether manipulating attention and prediction for common versus independent stimuli alters the extent to which they interact.

In this work we directly compared 2 competing models of the effects of attention on expectations—the Interaction and Opposition models—put forward in Kok et al. (2012). The data in that study were consistent with the Interaction model when considering regions of the visual cortex (V1, V2, and V3). Here, however, we took a different approach by implementing the models computationally and directly testing them against our data. By using Bayesian model comparison of statistical maps of EEG activity, and DCMs for ERPs, we were able to quantify how likely each of these 2 models was at every point of space and time at the scalp level, at each voxel in source space, and in the trial-dependent plastic changes within a cortical network. The Opposition model was unambiguously favored in our data at every level, that is, scalp, source, and network. At the network level we found that the plastic changes according to the Opposition model were more pronounced in forward connections. This is consistent with the idea that attention boosts, or heavily weights, prediction errors, which are then conveyed upward in the cortical hierarchy. Such prediction errors signal the need to update an internal perceptual model of the world, in turn prompting learning. At first glance it may appear that boosting of prediction errors is more consistent with the Interaction model. It is important to note, however,

that the corollary of the Interaction model is that attention reverses prediction, such that larger responses will be observed for predicted compared with unpredicted stimuli. In this sense, attention changes the sign of the prediction error instead of boosting it. On the contrary, boosting of prediction errors could in principle be accommodated by the Opposition model as it predicts a larger difference between unpredicted and predicted responses in the attended versus unattended condition. Having said this, our instantiation of the Opposition model is agnostic to such a relationship and was not modeled explicitly here. The Opposition model simply assumes that unpredicted responses will always be larger than predicted responses, regardless of attention, and that attention will boost these responses. It may also appear surprising that our attention manipulation did not modulate backward connections, which are thought to convey updated predictions (Friston 2005). In our paradigm, however, the predictions did not require constant updating, unlike in some other paradigms in which the rule constantly changes, such as in roving MMN (Garrido et al. 2008) or in reversal learning (Ghahremani et al. 2010). In such scenarios, it is possible that attention would modulate prediction updating via feedback processing (Desimone and Duncan 1995; Spratling 2008). We should, however, be cautious when interpreting the findings from our best individual model. While we have good evidence for an advantage of the Opposition family (77%) over the Interaction family (17%), and thus can assert that attention and prediction have opposing effects on plastic changes, we are less confident about where exactly in the network these effects might be expressed, given the relatively small advantage for the forward model over the remaining models tested.

While the better performance of the Opposition over the Interaction model is generally at odds with the findings by Hsu et al. (2014) and Kok et al. (2012), there was a narrow window of agreement in which the Interaction model was better at explaining the data at the scalp level, perhaps tellingly within the MMN time frame. This is an interesting finding, as it seems to point to a tonic Opposition effect between Attention and Prediction, and a phasic Interaction effect. Again, there are differences in both the type of paradigm and the neuronal measures between our study, which used EEG, and the experiment of Kok et al. (2012), which used fMRI. Although Attention was manipulated spatially in both studies, in our study it was directed towards a different (instead of the same) object. Moreover, our Prediction manipulation was learnt from the sequence of stimuli, rather than instructed (as in Kok et al. (2012)).

In conclusion, our findings provide empirical evidence for a computational model of the opposing interplay of attention and expectations in the brain. These opposing effects are manifested in neuronal activity and in plastic changes within a frontotemporal network engaged in sensory prediction errors. We demonstrate that attention boosts neuronal responses to predicted and unpredicted stimuli, and replicate the finding that attention boosts prediction errors, in keeping with the predictive coding framework (Rao and Ballard 1999; Friston 2005). Finally, we demonstrate that prediction errors are elicited regardless of one's state of attention, providing further support to the idea of a preattentive nature of change detection systems in the brain (Naatanen et al. 2001).

Funding

Australian Research Council (ARC) Discovery Early Career Researcher Award (DE130101393) and a University of Queensland Fellowship (2016000071) to M.I.G., an ARC Australian Laureate Fellowship (FL110100103) to J.B.M., the ARC Centre of Excellence for Integrative Brain Function (ARC Centre Grant CE140100007) to M.I.G. and J.B.M., and an ARC Special Research Initiative—Science of Learning Research Centre (SR120300015) to J.B.M.

Notes

We thank the volunteers for participating in this study and Maria Joao Rosa for discussions. *Conflict of Interest*: The authors declare no competing financial interests.

References

Auksztulewicz R, Friston K. 2015. Attentional enhancement of auditory mismatch responses: a DCM/MEG study. Cereb Cortex. 25:4273–4283.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L. 2009. Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci USA. 106:1672–1677.

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. Annu Rev Neurosci. 18:193–222.

Feldman H, Friston KJ. 2010. Attention, uncertainty, and free-energy. Front Hum Neurosci. 4:215.

Friston K. 2005. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci. 360:815–836.

Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM. 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage. 42:936–944.

Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Friston KJ. 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. Neuroimage. 36:571–580.

Garrido MI, Sahani M, Dolan RJ. 2013. Outlier responses reflect sensitivity to statistical structure in the human brain. PLoS Comput Biol. 9:e1002999.

Ghahremani DG, Monterosso J, Jentsch JD, Bilder RM, Poldrack RA. 2010. Neural components underlying behavioral flexibility in human reversal learning. Cereb Cortex. 20:1843–1852.

Henson RN, Mattout J, Phillips C, Friston KJ. 2009. Selecting forward models for MEG source-reconstruction using model-evidence. Neuroimage. 46:168–176.

Hsu YF, Hamalainen JA, Waszak F. 2014. Both attention and prediction are necessary for adaptive neuronal tuning in sensory processing. Front Hum Neurosci. 8:152.

Jiang J, Summerfield C, Egner T. 2013. Attention sharpens the distinction between expected and unexpected percepts in the visual brain. J Neurosci. 33:18438–18447.

Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP. 2012. Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex. 22:2197–2206.

Mattout J, Henson RN, Friston KJ. 2007. Canonical source reconstruction for MEG. Comput Intell Neurosci. 2007: Article ID 67613.

Montague PR. 1999. Reinforcement learning: an introduction. Trends Cogn Sci. 3:360–360.

Naatanen R, Alho K. 1997. Higher-order processes in auditory-change detection. Trends Cogn Sci. 1:44–45.

Naatanen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. 2001. "Primitive intelligence" in the auditory cortex. Trends Neurosci. 24:283–288.

Opitz B, Mecklinger A, Friederici AD, von Cramon DY. 1999. The functional neuroanatomy of novelty processing: integrating ERP and fMRI results. Cereb Cortex. 9:379–391.

Penny WD, Trujillo-Barreto NJ, Friston KJ. 2005. Bayesian fMRI time series analysis with spatial priors. Neuroimage. 24:350–362.

Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci. 2:79–87.

Rosa MJ, Bestmann S, Harrison L, Penny W. 2010. Bayesian model selection maps for group studies. Neuroimage. 49: 217–224.

Spratling MW. 2008. Reconciling predictive coding and biased competition models of cortical function. Front Comput Neurosci. 2:4.

Summerfield C, Egner T. 2009. Expectation (and attention) in visual cognition. Trends Cogn Sci. 13:403–409.

Summerfield C, Koechlin E. 2008. A neural representation of prior information during perceptual inference. Neuron. 59: 336–347.

Todorovic A, Schoffelen JM, van Ede F, Maris E, de Lange FP. 2015. Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band. PloS One. 10: e0120288.