The contribution of surprise to the prediction based modulation of fMRI responses

Catarina Amado[a], Petra Hermann[d], Petra Kovács[c,d], Mareike Grotheer[a,b], Zoltán Vidnyánszky[c,d], Gyula Kovács[a,b]

[a] Institute of Psychology, Friedrich-Schiller-University of Jena, 07737 Jena, Germany

[b] DFG Research Unit Person Perception, Friedrich-Schiller-University of Jena, 07743 Jena, Germany

[c] Department of Cognitive Science, Budapest University of Technology and Economics, 1111 Budapest, Hungary

[d] Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences, 1111 Budapest, Hungary

Corresponding Author: Gyula Kovács, Institute of Psychology, Friedrich-Schiller-University of Jena, Leutragraben 1, 07743 Jena, Germany, Tel: +493641/945936, e-mail: gyula.kovacs@uni-jena.de

Number of pages: 23

Number of figures: 4

Number of tables: 2

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

Abstract

In recent years, several functional magnetic resonance imaging (fMRI) studies showed that correct stimulus predictions reduce the neural responses when compared to surprising events (Egner et al., 2010). Further, it has been shown that such fulfilled expectations enhance the magnitude of repetition suppression (RS, i.e. a decreased neuronal response after the repetition of a given stimulus) in face selective visual cortex as well (Summerfield et al., 2008). Current MEG and neuroimaging studies suggest that the underlying mechanisms of expectation effects are independent from these of RS (Grotheer & Kovács, 2015; Todorovic & Lange, 2012). However, it is not clear as of today how perceptual expectations modulate the neural responses: is the difference between correctly predicted and surprising stimuli due to a genuine response reduction for correctly predicted stimuli or is it due to an increased response for surprising stimuli? Therefore, here we used a modified version of the paradigm of Grotheer & Kovács (2015) to induce predictions independently from repetition probability by presenting pairs of faces (female, male or infant) that were either repeated or alternating. Orthogonally to this, predictions were manipulated by the gender of the first face within each pair so that it signaled high, low or equal probability of repetitions. An unpredicted, neutral condition with equal probabilities for alternating and repeated trials was used to identify the role of surprising and enhancing modulations. Similarly, to Grotheer & Kovács (2015), we found significant RS and significant expectation effect in the FFA. Importantly, we observed larger response for surprising events in comparison to the neutral and correctly predicted conditions for alternating trials. Altogether, these results emphasize the role of surprise in prediction effects.

Keywords: Expectation, fMRI adaptation, Prediction, Repetition Suppression, Surprise

1. Introduction

The extensively studied neural repetition suppression (RS) phenomenon (for review see Grill-Spector, Henson, & Martin, 2006) has been recently associated with predictive coding (PC) theories (Friston, 2005) of neural functions (Summerfield et al., 2008). RS describes decreased neuronal response after the repetition of a given stimulus and is used to study the selective properties of neuronal populations (Malach, 2012). Similar to RS, fulfilled expectations also lead to reduced neural activity when compared to incorrect predictions, i.e. surprising events, and this phenomenon has recently been termed as expectation suppression (Grotheer & Kovács, 2015; Todorovic & Lange, 2012). Yet, the relationship between RS and expectation suppression as well as their underlying neural mechanisms are still unclear.

Summerfield et al., (2008) found that the magnitude of RS depends on the probability of stimulus repetitions (Prep): the RS was enhanced in the fusiform face area (FFA; Kanwisher et al., 1997) when faces were presented in blocks in which repetitions were frequent (therefore expected) as compared to when presented in blocks with low repetition probability. Authors suggested that higher-order contextual expectations modulated, via feedback connections, repetition-related processes and interpreted this result in the context of PC models (Friston, 2005; Rao & Ballard, 1999). According to PC, the visual cortex operates under a hierarchical structure where higher areas send predictions about sensory inputs to lower level areas, which then compute the difference between predictions and the actual sensory input (termed as prediction error - ε). To re-estimate and update predictions, ε is forwarded from lower to higher areas of the processing system. Consequently, surprising/incorrectly predicted events generate higher neural activity in comparison with correctly predicted events, maximizing the efficiency of neuronal processing (Friston, 2005, 2010; Friston & Kiebel, 2009). Summerfield et al., 2008 interpreted the enhanced magnitude of RS for expected stimuli as the reduced neuronal activity induced by a smaller ε (following Henson, 2003 claim of a link between RS and ε). This effect of expectation on RS was later replicated for faces (Grotheer et al., 2014; Kovács et al., 2012; 2013;

Larsson & Smith, 2012) and for stimuli of high expertise (Grotheer & Kovács, 2014), while such Prep modulations were not found for object-related RS (Kovács et al., 2013; Mayrhauser et al., 2014) and for unfamiliar characters (Grotheer & Kovács, 2014). It should be noted that all of these above studies used blocks with high and low repetition probabilities to manipulate expectations, for example in blocks with high likelihood of repetition, repeated trials are predicted and alternating trials are surprising while the opposite is true for blocks with low repetition probabilities. Therefore, this mixed design does not allow the independent testing of expectation and repetition effects.

Recently, a MEG study (Todorovic & Lange, 2012) could however, manipulate RS and expectation suppression independently, evoking expectations on a trial-by-trial basis using a preceding cue. Pairs of identical or different tones were presented; the expectations of the subjects were generated by the first tone of each pair, which signaled the likelihood of repetitions with 75% accuracy. The results indicated that expectation suppression and RS have different temporal windows, though an expectancy modulation on repetition effects was also observed. The different mechanisms behind expectation suppression and RS is supported further by Grotheer & Kovács, 2015, where pairs of female/male faces were used as stimuli and their gender was signaling the different repetition probabilities (for example female faces were repeated with high while male faces were repeated with a low probability). This fMRI study showed that RS and expectation suppression are additive, rather than interacting in the FFA and the occipital face area (OFA; Gauthier et al., 2000).

However, none of these previous studies could clarify whether the addition of expectation suppression and RS effects is due to a decrease of the response for correctly expected stimuli or an increase of the response to the surprising, unexpected stimuli (Kovács & Vogels (2014) raised this issue and suggested the inclusion of a "neutral" condition with equal probabilities for alternating and repeated trials, in which no expectations are induced (see also Arnal & Giraud, 2012 and Rahnev et al., 2011). Fig.1 illustrates the possible hypotheses regarding

RS and expectation modulations of the neural responses, considering the inclusion of the neutral condition. We reasoned that if the previously observed expectation effects are due to a genuine response reduction, then these trials should lead to lower BOLD signal when compared to the unpredicted, neutral trials as well. However, if the prediction effects are due to the enhanced response in the surprising trials (alternating and repeated) then these should lead to larger BOLD responses when compared to the unpredicted (neutral) as well as to the correctly predicted trials. Thus, a main effect of expectation conditions and a subsequent post-hoc analysis would clarify from which expectation condition the BOLD signal of unpredicted, neutral trials differs most – from the correctly predicted (suggesting the role of expectation in predictions) or from the surprising trials (suggesting the role of surprise in predictions). Here we used the methods, task and paradigm of Grotheer & Kovács (2015) with the additional trials of the neutral, unpredicted condition, to study under which circumstances these top-down (suppressing or enhancing) modulations operate.

Anticipating our results, we found significant RS and expectation effect in the FFA. Further, we observed a significant increase of neuronal responses for the surprising, unexpected events, relative to the neutral and unpredicted events in the alternation trials. The relationship of RS and surprise differed between hemispheres: rFFA revealed a dependence of RS on surprise, whereas lFFA showed the independence of these two processes. Overall our results emphasize the role of surprise in predicted processes.

----Figure 1 about here

2. Material and methods

2.1 Participants

24 healthy subjects participated in the experiment after giving written, informed consent in accordance with the protocols approved by the Ethical Committee of the Friedrich Schiller

University Jena. No participant had any history of neurological or psychiatric illness and all had normal or corrected to normal vision. Due to technical issues, 2 participants were excluded from the analysis and for 1 participant only 2 of the 3 functional runs were acquired. Thus, 22 subjects (8 male; 2 left-handed, mean age (±SD): 23.5 (2.9) years) were involved in the final analysis.

2.2 Stimulation and Procedure

The experimental design (Fig.2) was similar to what has formerly been used to evaluate the relationship between stimulus repetitions and fulfilled expectations (Grotheer & Kovács, 2015), with the exception that two additional trial types were introduced − neutral repetition trials (Neu_Rep) and neutral alternation trials (Neu_Alt). These were corresponding to a previously suggested neutral or unpredicted condition (Rahnev et al., 2011; Kovács & Vogels, 2014) in which the probability of repeated (Rep) and alternated (Alt) trials was identical. The other four conditions were identical to those of Grotheer & Kovács (2015). Briefly, trials were either correctly expected (Exp, high probability) or surprising (Sur, low probability). This conscious expectation was achieved by the fact that the category of the first face (female, male or infant) in each pair signaled with 75% accuracy whether repetitions or alternations were more likely to occur. In other words, the probability of Rep and Alt was contingent with the gender of the first face stimuli. In addition to this and orthogonal to the modulation of expectation, trials could either be repeated or alternating allowing the testing of the independence of expectation and repetition effects. Overall we had six different trial types that were presented randomly within a run: correctly predicted repetition trials (Exp_Rep), correctly predicted alternation trials (Exp_Alt), surprising repetition trials (Sur_Rep), surprising alternation trials (Sur_Alt), neutral repetition trials (Neu_Rep) and neutral alternation trials (Neu_Alt). In all conditions, pairs of female and male faces were either repeating or alternating with an overall probability of 50-50%.

Stimulus presentation was controlled via MATLAB R2014a (The Mathworks, Natick, MA, USA), using Psychtoolbox (Version 3.0.12). 280 grey-scale, digital photos of full-frontal Caucasian adult faces (50% of both genders), similar to the face stimuli of Kovács (2012, 2013) and 140 grey-scale full-frontal digital photos of Caucasian young infant faces aged between 1 and 2 years (collected from the public domain of the world wide web) were placed behind a circular mask and positioned in the centre of the screen on a homogeneous grey background. Stimuli were displayed via an MRI-compatible LCD screen (32' NNL LCD Monitor, NordicNeuroLab, Bergen, Norway; refresh rate: 60Hz) placed at 142 cm from the observer. A total of 3 runs were administered and no stimulus appeared in more than one trial during each run. One run was composed of 240 trials (correctly predicted, surprising and neutral conditions had 120, 40 and 80 trials within a single run, respectively) and took about 13 minutes. Stimuli were presented for 250 ms each, pairwise, separated by a randomly varied inter-stimulus interval (between 400 and 600 ms with 50 ms steps) and followed by a randomized 1 or 2s long inter-trial interval. The first stimulus (S1) could either be identical to (Rep) or different from (Alt) the second stimulus (S2). Stimulus size was 6° in diameter. To avoid low level local feature adaptation the size of either S1 or S2 was randomly reduced by 32% in each trial (Grotheer & Kovács, 2014; Summerfield et al., 2008). The same face category was used for each stimulus pair (i.e. both faces of the pair were always either female, male or infant) and subjects were presented with 33.33% female/male/infant trials (administrated randomly). The stimulus category was used as a cue to signal high-, low- or medium- probabilities of trial type (repetition or alternation) occurrences. For example for a given participant female faces signaled high (75%), infant faces signaled equal (50%) repetition probabilities while male faces signaled high (75%) alternation probability. This way participants could form predictions regarding the likelihood of repetitions and alternations (in this example repetitions are predicted when a female face appeared while alternations are predicted when male faces were presented). Importantly, the third face category (in the above example the infant faces) was used as an unpredictive (Arnal & Giraud, 2012) or neutral cue as repetitions and alternations could follow S1 with equal likelihood in this category. The relationship between face category and repetition probability was counterbalanced across participants in a way that each category signaled high/low/equal probabilities with equal probability across participants (to avoid any possible stimulus effects that could arise due to a different attentional modulation invoked by adult and infant faces - Brosch et al.,(2007)). Prior to the scanning session, participants were informed about the relative repetition/alternation probabilities as well as about their contingencies on stimulus category and adequate task performance was assured. During the scanning session, before the 80th and the 160th trial a message appeared on the screen for 10s to inform participants about the relative repetition/alternation probabilities of the subsequent trials as well as about their contingencies on stimulus category.

To confirm that subjects were paying attention to the stimuli and to guarantee that they were capable of judging the stimulus categories effectively, 18% of the trials were target trials in which subjects had to respond whether the S1 had been a female, male or infant face (Todorovic & Lange, 2012). Therefore, for these target trials 1 sec after S2 presentation, a choice-screen was displayed for 2 seconds showing the different stimulus category options (infant, female and male). The order of the presented options on this choice-screen was counterbalanced across trials. Participants had to indicate their choice by pressing the left, middle or right button. A small color change of the fixation cross functioned as feedback regarding their response (green for correct and red for incorrect answers).

----Figure 2 about here

2.3 Imaging Parameters and Data Analysis

Imaging was performed with a 3-Tesla MR scanner (Siemens MAGNETOM Prisma fit, Erlangen, Germany) located at the Research Centre for Natural Sciences (Hungarian Academy of Sciences) in Budapest, Hungary. A T1-weighted high-resolution 3D anatomical image was acquired using a MP-RAGE sequence. The anatomical data had the following parameters: TR = 2300ms; TE = 3.03ms; 192 slices; 1 mm isotropic voxel size. These images were prearranged based on a sagittal localizer to cover the whole brain. fMRI images (T_2* weighted images) were collected using an EPI sequence (34 slices, 10° tilted relative to axial, TR = 2000ms; TE = 30ms; flip angle = 90°; matrix size = 64 × 64; 3 mm isotropic voxel size).

Details of pre-processing and statistical analysis were described previously (Cziraki, Greenlee, & Kovács, 2010). Briefly, the functional images were realigned, normalized to the MNI-152 space, resampled to 2 x 2 x 2 mm resolution and spatially smoothed with a Gaussian kernel of 8 mm FWHM (SPM12, Welcome Department of Imaging Neuroscience, London, UK). Independent functional localizer run (640 sec long, 20 sec epochs of faces, objects and Fourier randomized versions of faces, interleaved with 20 sec of blank period, 2 Hz stimulus repetition rate; 300 ms exposure; 200 ms blank) served as basis for Regions of Interest (ROIs) selection, which were analyzed using the MARSBAR 0.44 (Brett et al., 2002) toolbox for SPM. The ROIs were selected individually on the single subject level from the thresholded ($p<0.001_{\text{uncorrected}}$) *t* maps of the contrast *faces vs Fourier randomized faces*. The FFA (N=22), average MNI coordinates (±SE) and cluster sizes (±SE) for left and right hemisphere were the following: -40.4 (0.8) - 59.6 (1.3) - 17.7 (0.9) and 54(5); 41.6 (0.9) -57.8 (1.3) -16.8 (0.7) and 54(3). The clusters include voxels higher than $p<0.001_{\text{uncorrected}}$ within a 5mm sphere around the peak voxel. The average locations of these ROIs are presented together with the localizer whole-brain results in Fig.3A.

A time series of the average voxel value within the different ROIs was determined and extracted from our event-related sessions. The convolution of the canonical Hemodynamic Response Function (HRF) of SPM12 with each of the 6 experimental conditions (Exp_Rep, Exp_Alt, Sur_Rep, Sur_Alt, Neu_Rep, Neu_Alt) was used to define predictors for a General Linear Model (GLM) analysis of the data. Target trials were not modelled separately, due to sufficient time between trial and choice screen presentations. Preliminary analyses revealed no main effect of experimental run ($F(2,40)=0.4865$, $p=0.62$, $\eta p^2=0.02$) nor significant interactions between run and trial type ($F(2,40)=1.56$, $p=0.22$, $\eta p^2=0.07$) or conditions ($F(4,80)=1.7$, $p=0.15$, $\eta p^2=0.08$), therefore the results of the three runs were averaged. We performed repeated measures ANOVAs for the FFA with hemisphere (2), expectation condition (3) and trial type (2) as within-subject factors. Post-hoc analyses were performed using Fisher LSD tests.

3. Results

3.1 Behavior

Mean accuracy for gender judgement was 91% (±SD: 8%) across all trial types (Exp_Rep: 93(7)%, Exp_Alt: 94(5)%, Sur_Rep: 84(18)%, Sur_Alt: 88(13)%, Neu_Rep: 92(6)%, Neu_Alt: 90(11)%). The participant's performances did not differ between trial types ($F(1,21)=0.97$, $p=0.34$, $\eta p^2=0.04$). However, participants had a significantly lower performance in trials when their predictions were incorrect (main effect of expectation condition: $F(2,42)=3.8$, $p=0.03$) as compared to trials with correct predictions (Fisher LSD *post hoc* test: $p=0.01$) and a similar trend was seen when compared against neutral trials (Fisher LSD *post hoc* test: $p=0.08$).

On average participants required 1146ms (±SD: 117ms) to determine the gender of the presented faces. Reaction times did not differ significantly between trial types ($F(1,21)=0.105$, $p=0.75$, $\eta p^2=0.01$) or expectation conditions ($F(2,42)=0.82$, $p=0.45$, $\eta p^2=0.04$).

3.2 fMRI

3.2.1 FFA

The ANOVA results of this section will be presented in the following order: 1. Main effect of RS; 2. Main effect of expectation condition; 3. Main effect of hemisphere; 4. Interactions.

We observed a significant repetition suppression (Fig.3B; main effect of trial type: $F(1,21)=26.84$, $p=0.00004$, $\eta p^2=0.56$) with an average signal reduction of 0.04% (equivalent to a relative signal reduction of 14%). We also found a main effect of expectation condition (F(2,42)=5.09, p=0.01, ηp^2=0.2), which was due to a larger BOLD response for incorrect predictions when compared to neutral, unpredicted events (Fisher LSD *post hoc* test: p=0.003). On average the incorrect predictions led to a signal enhancement of 0.04% in comparison with the neutral condition (corresponding to a relative signal increase of 12%), suggesting the role of surprise related response enhancement. Interestingly, a similar tendency was observed when comparing correctly predicted and neutral conditions (Fisher LSD *post hoc* test: p=0.07), indicating somewhat smaller overall responses for the unpredicted, equal probability trials as compared to trials with correct predictions. Unlike in our prior study (Grotheer & Kovács, 2015) we found no difference between correct and incorrect predictions (Fisher LSD *post hoc* test: p=0.2).

We also observed a main effect of hemisphere (F(1,21)=20.25, p=0.0002, ηp^2=0.49) in the form of larger BOLD responses in the right, when compared to the left FFA. This hemisphere effect interacted with trial type: F(1,21)=5.09, p=0.04, ηp^2=0.2), due to a larger RS effect in the right (Fisher LSD *post hoc* test: p<1e-7) in comparison with the left hemisphere (Fisher LSD *post hoc* test: p=4e-5). Importantly the three-way interaction of trial type × expectation condition × hemisphere was also significant (F(2,42)=3.38, p=0.04, ηp^2=0.14), meaning that the magnitude of RS showed a dependency on expectation condition and hemisphere. This interaction is mainly due to the higher RS for incorrect predictions over the other conditions for the right FFA. Nonetheless, the repetition effect was significant for all expectation conditions and for both the left and right hemispheres (Fisher LSD *post hoc* tests: p<0.001 for all comparisons). Additionally, alternating trials were significantly different for the three expectation conditions for both hemispheres, having the most elevated responses during the surprising events and lower BOLD responses in the neutral and in the correctly predicted conditions (Fig. 3B, Fisher LSD *post hoc* tests: p<0.05 for all comparisons).

----Figure 3 about here

In order to test the robustness of these results we performed the identical analysis on a smaller, spherical ROI with a diameter of 3mm. The results of this and the previously presented analysis were identical (with the exception that the three-way interaction of trial type × expectation condition × hemisphere showed only marginal significance), suggesting that the applied criterion has no major effect on it (Table I).

Table I. Summary of the ANOVA results for a 3 mm sphere ROI.

Main effect: RS (***)	F(1,21)=27.11	p=0.00004	ηp2=0.56
Main effect: expectation condition (**)	F(2,42)=5.79	p=0.006	ηp2=0.22
Main effect: hemisphere (***)	F(1,21)=21.23	p=0.0002	ηp2=0.5
Interaction: hemisphere×RS (*)	F(1,21)=5.84	p=0.03	ηp2=0.22
Interaction: hemisphere×RS×expectation condition (+)	F(2,42)=2.95	p=0.06	ηp2=0.12

3.2.2. Whole-brain analysis

To test whether repetition and expectation effects are encoded by other neurons outside the FFA, we also performed a second-level whole-brain analysis testing for repetition and expectation effects as well as for the interaction of these factors, using a fixed threshold of $p<0.05_{FWE}$ with a cluster size >20 voxels. Testing the main effect of repetition (Alt>Rep) revealed one active cluster in the right fusiform gyrus (MNI [x,y,z]: 36, -52, -14; cluster size: 351; see Fig.4). While not identical with it, this coordinate closely resembles the average coordinate of our rFFA. The opposite contrast (Rep>Alt) led to no significant activations anywhere in the brain. The same threshold yielded two clusters of activations for the Surprising>Expected (Sur>Exp) contrast, revealing higher activations during surprising when compared to correctly predicted trials over the inferior frontal gyrus (MNI [x,y,z]: 48, 24, 10 (BA 45) and 32, 24, -6 cluster sizes: 31 and 30). No significant activations were found for the opposite contrast (Exp>Sur).

To confirm that no region remained unnoticed by the commonly applied but rather rigorous FWE corrected threshold we also analyzed our data at a less conservative threshold ($p<0.0001_{\text{uncorrected}}$; cluster extent of >20 voxels). The Alt>Rep contrast and Sur>Exp showed some additional regions with significant activations (Table II and Fig.4). Furthermore, when compared to surprising trials, correctly predicted ones (Exp>Sur) led to increased activation in the parahippocampal gyrus, hippocampus, claustrum and putamen (Table II and Fig.4). The contrast testing differences between neutral and surprising conditions (Sur>Neu) revealed significant activations in the inferior and middle frontal gyri. Interestingly, these regions were also activated in the Sur>Exp contrast and in fact the location of the two regions is similar for both contrasts (Table II and Fig.4). The whole-brain analysis did not reveal additional active clusters when testing for further expectation effects or for the interaction of RS with expectation conditions.

Table II. Summary of the significant activations based on the whole-brain analysis.

Contrast	Brain region	Coordinates	Cluster size	Threshold
Alt>Rep	Fusiform gyrus	36, −52, −14	351	(p<0.05 FWE)
Alt>Rep	Middle occipital gyrus	−34, −78, 8	89	(p<0.0001)
Sur>Exp	Brodmann area 45	48, 24, 10	31	(p<0.05 FWE)
Sur>Exp	Inferior frontal gyrus	32, 24, −6	30	(p<0.05 FWE)
Sur>Exp	Middle frontal gyrus	46, 20, 42	368	(p<0.0001)
Sur>Exp	Middle frontal gyrus	40, 56, 2	149	(p<0.0001)
Sur>Exp	Inferior frontal gyrus	−34, 20, −6	169	(p<0.0001)
Sur>Exp	Inferior parietal lobe	50, −48, 48	290	(p<0.0001)
Sur>Exp	Brodmann area 9	6, 38, 30	157	(p<0.0001)
	Parahippocampal gyrus	−34, −8, −20	59	(p<0.0001)
Exp>Sur	Hippocampus	−30, −20, −14	40	(p<0.0001)
Exp>Sur	Claustrum	−30, −2, 12	21	(p<0.0001)
Sur>Neu	Putamen	32, −2, 6	209	(p<0.0001)
Sur>Neu	Middle frontal gyrus	40, 8, 36	294	(p<0.0001)
Sur>Neu	Inferior frontal gyrus	50, 24, 12	37	(p<0.0001)
Exp>Sur	Inferior frontal gyrus	34, 26, −6	40	(p<0.0001)

----Figure 4 about here

4. Discussion

Our major result is that surprising events lead to significantly larger activity as compared to unpredicted, neutral events, thereby supporting the hypothesis outlined on Fig.1A and emphasizing the role of surprise in predictive coding processes.

Predictive coding models assume that ε relies on the discrepancy between observed and predicted sensory states (Friston, 2012), supporting the finding of surprise related enhancement of the activity. However according to theories of PC (Friston 2012; 2009) correct predictions reduce ε, pointing to the involvement of expectation suppression as well.

So far only a handful of studies tried to disentangle these mechanisms, surprise enhancement and expectation suppression, by the application of a third, neutral or unpredicted condition. Recently, Egner et al., (2010) tested subject's perceptual face expectation (low, medium and high) during a task that was orthogonal to the manipulation of expectations. Authors found that surprise events contributed more robustly (about twice as strongly) to FFA BOLD responses when compared to correctly predicted events. Unfortunately, Egner et al., (2010) did not report if their medium condition, where the cue signaled the occurrence of faces with 50% probability was different from either of the other two conditions or not. Our results confirm and extend these results, suggesting enhanced response during incorrectly predicted conditions, unlike the surprise minimization shown in choice behavior (Schwartenbeck et al., 2015).

One important limitation of the current study is, however, that univariate BOLD signal analysis techniques are unable to exclude the role of ε unit activity reduction in determining lower responses in correctly as compared to non-predicted and incorrectly predicted trials. As the BOLD signal reflects the mixture of ε and representational unit activity, due to its low spatial resolution, we can not decide if the observed similar activity for the predicted and neutral, non-predicted conditions is due to similar ε activity within the two conditions or there is also an additional reduction of ε units be the correctly predicted events which is compensated by an enhanced representational unit activity, leading to similar response magnitude or the correctly predicted and non-predicted conditions.

Previously, Kok et al, (2012) found better multivariate pattern (MVPA) classification rates for expected conditions when compared to surprising conditions, suggesting that the reduction of the neural response amplitudes for expected stimuli is associated with an the improvement of the stimulus representation. Therefore, future MVPA studies could clarify further the neural mechanism behind this reduced BOLD response we observed for the non-predicted, neutral and predicted conditions as compared to unpredicted condition. If the classifier performance is equally higher for neutral and expected conditions, the observed expectation suppression is solely due to an increased ε unit activity in the unpredicted trials i.e. due to surprise. On other hand, if the classifier performance is lower for both surprising and neutral events, then one can assume that there is an additional ε unit activity reduction, driven by correct predictions.

Rahnev et al., (2011) also confirms this result when inducing expectations regarding the direction of a moving-dot pattern. Authors found a neural response enhancement for invalidly cued (therefore surprising) events when compared to expected and neutral events (with the expected cued activations falling in between the other two) in the dorsolateral prefrontal cortex and in the intraparietal sulcus. However, unlike in our study, the Rahnev et al., (2011) work revealed no expectation modulation in the sensory areas (in the motion-sensitive medial temporal cortex), but rather exclusively in the bilateral dorsolateral prefrontal cortex.

In contrast with our prior study (Grotheer & Kovács 2015) where a general reduction of the response was found for correctly when compared to incorrectly predicted trials, our current results could find such differences only for the alternating trials. However, this difference of results can easily be explained by the different experimental designs. First, due to the additional (neutral) expectation condition, in the current study, trials where expectations were fulfilled had an average probability of 50% over the 3 conditions only whereas in Grotheer & Kovács (2015) the same probability was 75%. Even though the ratio of Surprising/Expeted trials was identical in the current and in the Grotheer & Kovács (2015) study, the overall number of trials in the expected and surprising conditions was significantly reduced. Second, the introduction of the neutral condition might also serve as distracter from the predictive conditions. PC assumes a constant re-estimation and update of predictions, yet subjects were aware of the equal probability occurrence of repetitions and alternations during the neutral conditions, thus it is possible that no predictions were formed during this specific condition, interrupting the constant up-dating and re-estimation of predictions. Indeed, predictable events usually occur in a non-random fashion, allowing the brain to infer about the probabilistic or deterministic regularity of the different events (Bubic, von Cramon, & Schubotz, 2010). However, previous studies proposed that similar predictive strategies are employed to attempt the extraction of a pattern within random inputs as well (Schubotz & von Cramon, 2002). It should be noted that in the Schubotz & von Cramon (2002) study the involved task required participants to constantly indicate whether the last trials of a sequence were expected (correct) or surprising (violated) based on the sequence order, forcing the creation of predictive strategies during the randomized sequences. While in the current study the gender judgment task was orthogonal to the manipulation of expectation and the events were not presented sequentially (thus our participants were unable to use memory regarding previous neutral events to infer about upcoming random events). Therefore, an overall reduction of the number of trials and the existence of the unpredicted trials might explain the absence of overall BOLD signal differences between E and S trials in the current study.

Contrary to what theories of predictive coding suggest (Arnal & Giraud, 2012) the neutral condition induced the lowest activity in our study, even showing a tendency for leading to lower activity than the correctly predicted condition. This result, however, has been observed in previous studies as well. Rahnev et al., (2011) found lower activations in the dorsolateral prefrontal and inferior parietal cortices for neutral when compared to predicted (incorrectly and correctly) trials. Two mechanisms are able to explain the higher FFA BOLD signal for correctly predicted when compared to unpredicted trials. First, it is possible that it reflects the activity of both representational and ε encoding units simultaneously. In this case, the lower ε can be compensated by the relatively larger representational unit activity during the correctly predicted trials, while only a small ε related activity and a relatively smaller representation unit activity exist in the neutral trials. Second, it is possible that correct predictions do not reduce ε to its minimum. Rather, ε remains larger even in an event that is predicted correctly. This ε, which is smaller in the unpredicted trials, could explain the larger BOLD signal for the correctly predicted trials. Testing these hypotheses, however, will require further specifically aimed studies, preferably with multivariate analysis techniques.

Expectation-based processes may be important in modulating repetition suppression (for a summary see Kovács & Vogels, 2014; Summerfield et al., 2008). In the current study, we found significant RS for both the right and left FFA. The dependence of the RS on expectation condition, however, showed hemispheric differences: while the observed RS was independent of expectation condition over the left hemisphere it showed an interaction with expectations over the right hemisphere being the largest in the surprising condition. This suggests different neural mechanisms behind the RS for the two hemispheres. This would not be the first time such hemispheric differences are found. For example, Yovel et al., (2008) showed different neural mechanisms in the right and left FFA, finding that only the rFFA was sensitive to symmetry.

It is possible that enhanced responses for the surprising trials reflect the capture of attention additionally to the surprise effect originating from the violation of expectations. However the current experimental design did not enable us to test for possible attentional effects like previous studies: for instance, Jiang et al., (2013) manipulated expectation and attention effects orthogonally. Using multivariate pattern analysis, Jiang and colleagues show that attention enhances the precision of ε in FFA, due to an improvement of the stimulus representation in this area. Overall, the computation of fulfilled or violated expectations requires the attention of the observers (Larsson & Smith, 2012), meaning that attention is a precondition for discriminating expected from unexpected stimuli (for a summary of attention and prediction processes see the review by Summerfield & Egner, 2009).

The whole-brain analysis of the current study is congruent with Grotheer & Kovács (2015); in the sense that no expectation effects were revealed in early visual areas, unlike previous studies (Grotheer & Kovács, 2014; Larsson & Smith, 2012). This supports the idea that our experimental design, which induces expectations explicitly, produces expectation effects on higher-processing levels (Grotheer & Kovács, 2015). Furthermore, the similarity between active clusters for Surprise>Neutral and Surprise>Expected points once again to a distinction between surprising events from the other two (neutral and correctly predicted) conditions.

In summary, we observed elevated BOLD responses in the ventral visual stream for surprising events when compared to neutral ones, emphasizing the role of surprise in prediction based modulation of the BOLD signal. Crucially, the relationship of RS magnitude and expectation effect varies between hemispheres: being dependent in the right hemisphere and additive in the left hemisphere.

Acknowledgements

Supported by a Deutsche Forschungsgemeinschaft Grant (KO 3918/1-2; 2-2). HP and ZV was supported by a grant from the Hungarian Brain Research Program (KTIA_13_NAP-A-I/18).
We thank the anonymous reviewer for pointing out the inability of univariate analysis techniques to separate ε and representational unit activities.

References

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. http://doi.org/10.1016/j.tics.2012.05.003

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. http://doi.org/10.1038/nrn756

Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... Attention capture by infant faces. *Emotion*, *7*(3), 685–689. http://doi.org/10.1037/1528-3542.7.3.685

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, Cognition and the Brain. *Frontiers in Human Neuroscience*, *4*. http://doi.org/10.3389/fnhum.2010.00025

Cziraki, C., Greenlee, M. W., & Kovács, G. (2010). Neural Correlates of High-Level Adaptation-Related Aftereffects. *Journal of Neurophysiology*, *103*(3), 1410–1417. http://doi.org/10.1152/jn.00582.2009

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *The Journal of Neuroscience*, *30*(49), 16601–16608. http://doi.org/10.1523/JNEUROSCI.2770-10.2010

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 815–836. http://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. http://doi.org/10.1038/nrn2787

Friston, K. (2012). Prediction, perception and agency. *International Journal of Psychophysiology*, *83*(2), 248–252. http://doi.org/10.1016/j.ijpsycho.2011.11.014

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1211–1221. http://doi.org/10.1098/rstb.2008.0300

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, *12*(3), 495–504.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. http://doi.org/10.1016/j.tics.2005.11.006

Grotheer, M., Hermann, P., Vidnyánszky, Z., & Kovács, G. (2014). Repetition probability effects for inverted faces. *NeuroImage*, *102, Part 2*, 416–423. http://doi.org/10.1016/j.neuroimage.2014.08.006

Grotheer, M., & Kovács, G. (2014). Repetition Probability Effects Depend on Prior Experiences. *The Journal of Neuroscience*, *34*(19), 6640–6646. http://doi.org/10.1523/JNEUROSCI.5326-13.2014

Grotheer, M., & Kovács, G. (2015). The relationship between stimulus repetitions and fulfilled expectations. *Neuropsychologia*, *67*, 175–182. http://doi.org/10.1016/j.neuropsychologia.2014.12.017

Henson, R. N. A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, *70*(1), 53–81. http://doi.org/10.1016/S0301-0082(03)00086-8

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(47), 18438–18447. http://doi.org/10.1523/JNEUROSCI.3308-13.2013

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. http://doi.org/10.1016/j.neuron.2012.04.034

Kovács, G., Iffland, L., Vidnyánszky, Z., & Greenlee, M. W. (2012). Stimulus repetition probability effects on repetition suppression are position invariant for faces. *NeuroImage*, *60*(4), 2128–2135. http://doi.org/10.1016/j.neuroimage.2012.02.038

Kovács, G., Kaiser, D., Kaliukhovich, D. A., Vidnyánszky, Z., & Vogels, R. (2013). Repetition Probability Does Not Affect fMRI Repetition Suppression for Objects. *The Journal of Neuroscience*, *33*(23), 9805–9812. http://doi.org/10.1523/JNEUROSCI.3423-12.2013

Kovács, G., & Vogels, R. (2014). When does repetition suppression depend on repetition probability? *Frontiers in Human Neuroscience*, *8*. http://doi.org/10.3389/fnhum.2014.00685

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. A. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, *29*(5), 250–256. http://doi.org/10.1016/j.tins.2006.02.008

Larsson, J., & Smith, A. T. (2012). fMRI Repetition Suppression: Neuronal Adaptation or Stimulus Expectation? *Cerebral Cortex*, *22*(3), 567–576. http://doi.org/10.1093/cercor/bhr119

Malach, R. (2012). Targeting the functional properties of cortical neurons using fMR-adaptation. *NeuroImage*, *62*(2), 1163–1169. http://doi.org/10.1016/j.neuroimage.2012.01.002

Mayrhauser, L., Bergmann, J., Crone, J., & Kronbichler, M. (2014). Neural repetition suppression: evidence for perceptual expectation in object-selective regions. *Frontiers in Human Neuroscience*, *8*. http://doi.org/10.3389/fnhum.2014.00225

Rahnev, D., Lau, H., & Lange, F. P. de. (2011). Prior Expectation Modulates the Interaction between Sensory and Prefrontal Regions in the Human Brain. *The Journal of Neuroscience*, *31*(29), 10741–10748. http://doi.org/10.1523/JNEUROSCI.1478-11.2011

Rao, R. P. N., & 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. http://doi.org/10.1038/4580

Schubotz, R. I., & von Cramon, D. Y. (2002). A Blueprint for Target Motion: fMRI Reveals Perceived Sequential Complexity to Modulate Premotor Cortex. *NeuroImage*, *16*(4), 920–935. http://doi.org/10.1006/nimg.2002.1183

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., Kronbichler, M., & Friston, K. (2015). Evidence for surprise minimization over value maximization in choice behavior. *Scientific Reports*, *5*. http://doi.org/10.1038/srep16575

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. http://doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. http://doi.org/10.1038/nn.2163

Todorovic, A., & Lange, F. P. de. (2012). Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. *The Journal of Neuroscience*, *32*(39), 13389–13395. http://doi.org/10.1523/JNEUROSCI.2227-12.2012

Yovel, G., Tambini, A., & Brandman, T. (2008). The asymmetry of the fusiform face area is a stable individual characteristic that underlies the left-visual-field superiority for faces. *Neuropsychologia*, *46*(13), 3061–3068. http://doi.org/10.1016/j.neuropsychologia.2008.06.017

Figures

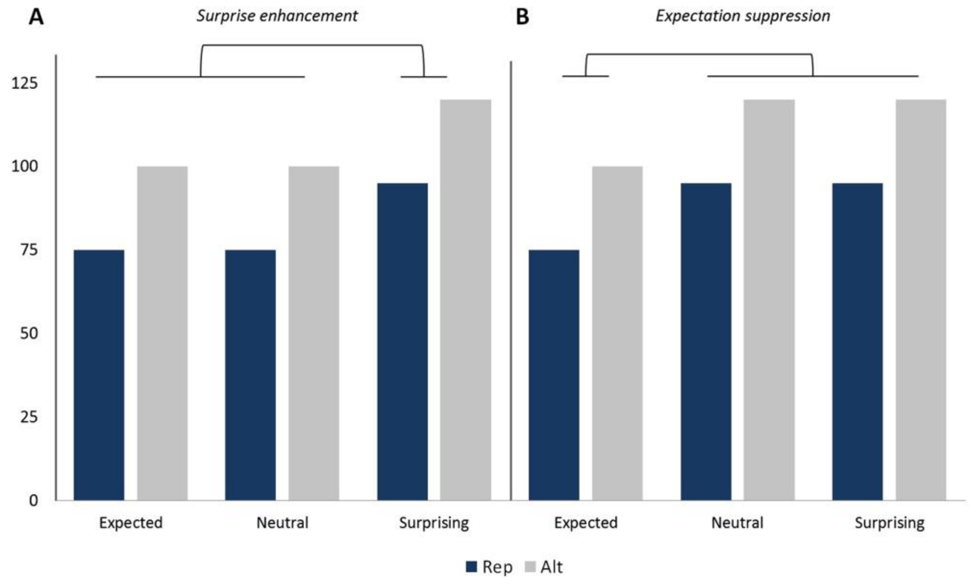

Figure 1. Illustration of how repetitions and expectations might modulate neural responses. Speculative BOLD responses are presented separately for repeated (Rep) and alternating (Alt) stimuli under correctly predicted, neutral and surprising conditions. (A) Surprising trials (Rep and Alt) lead to elevated BOLD responses when compared to neutral and correctly predicted ones (Rep and Alt), signaling the role of surprise enhancement. (B) Correctly expected trials (Rep and Alt) lead to lower responses when compared to neutral and surprising trials (Rep and Alt), suggesting the role of expectation suppression. Please note the existence of a response difference between fulfilled and violated expectations in the current scheme, as suggested previously (Arnal & Giraud, 2012) and shown by Grotheer & Kovács, 2015.

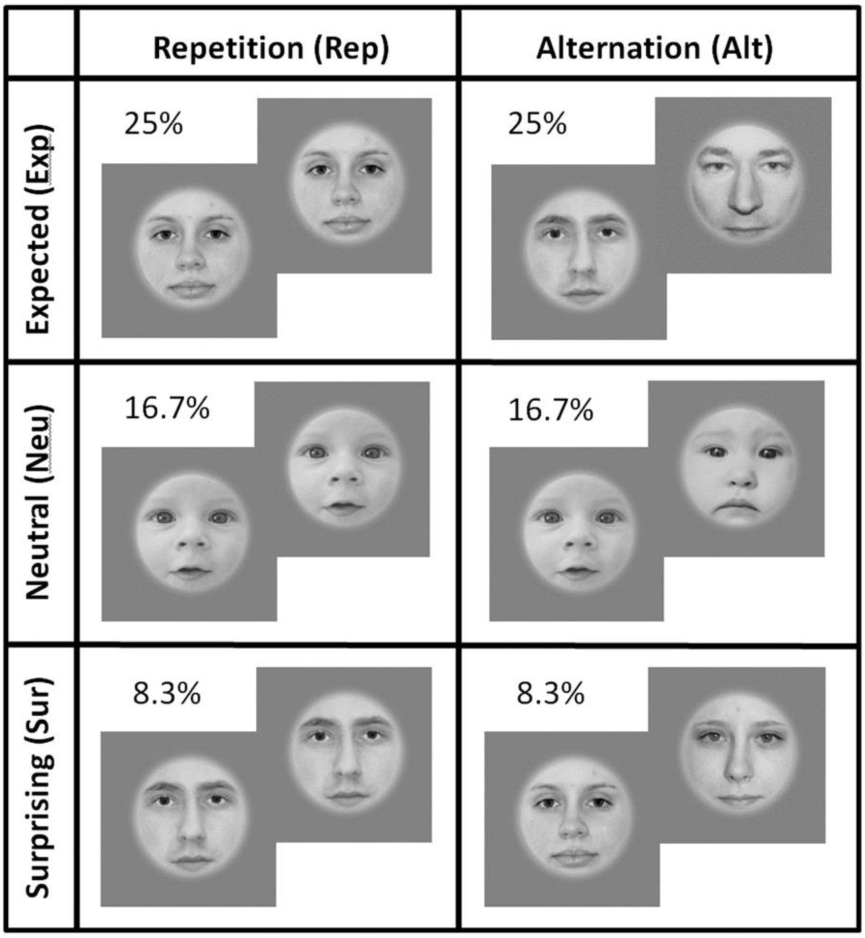

Figure 2. Experimental design and conditions. Each stimulus category (either infant, male or female faces) was used as a cue to signal different repetition/alternation probabilities (high, low or equal) randomly for every participant. Here we present a situation where the face category signaling high repetition probability was female (Exp_Rep), while male faces signaled high probability of alternations (Exp_Alt). Conversely, male/female faces signaled low probability of repetitions/alternations (Sur_Rep/Sur_Alt). Infant faces indicated equal repetition and alternation probabilities (Neu_Rep and Neu_Alt).

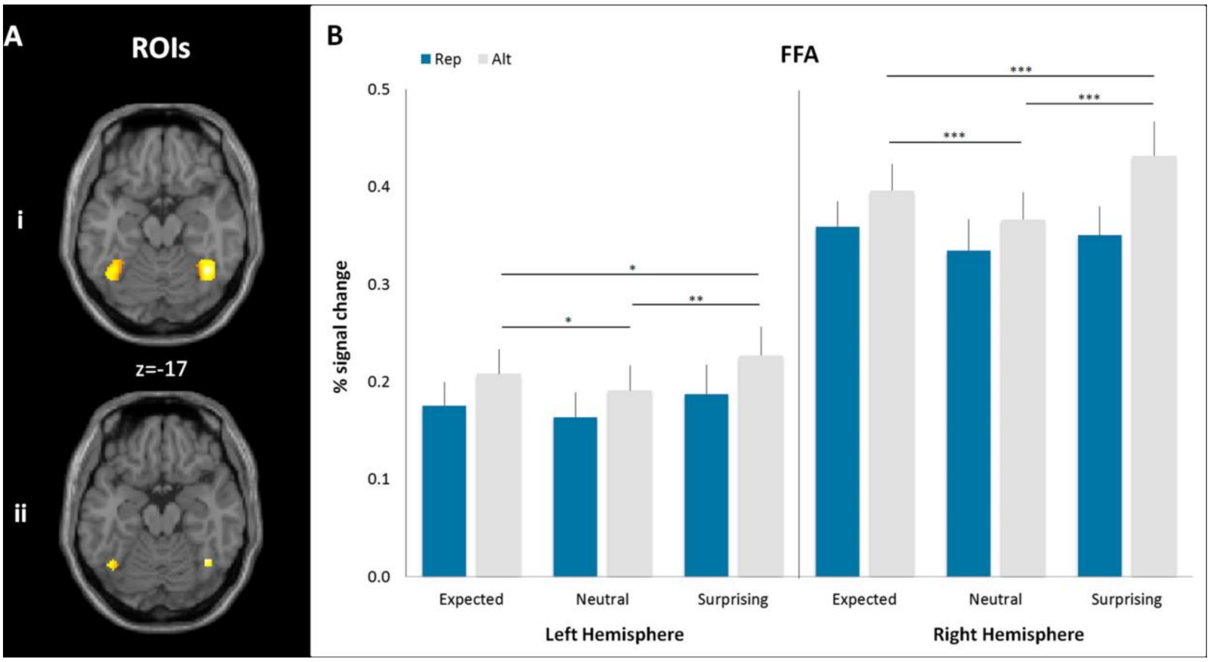

Figure 3. The location of the left and right FFA and the respective activations in the form of percent-signal change. A. Results of the functional localizer used to determine the location of FFA. i) Average activations contrasting faces and Fourier noise images ($p<0.05_{FWE}$ with a cluster extent of > 50 voxels). ii) A 4mm sphere around the average peak coordinates of FFA. B. Effects of expectation conditions (E, N and S) and repetitions (Alt/Rep). Percent-signal changes (±SE) of the FFA (left and right hemispheres) are presented separately for trial types and expectation conditions. The RS effects were significant ($p<0.001$) for all conditions in left and right hemispheres, therefore not marked separately. *$p<0.05$, **$p<0.01$, ***$p<0.001$. Rep - repetition trial, Alt - alternation trial

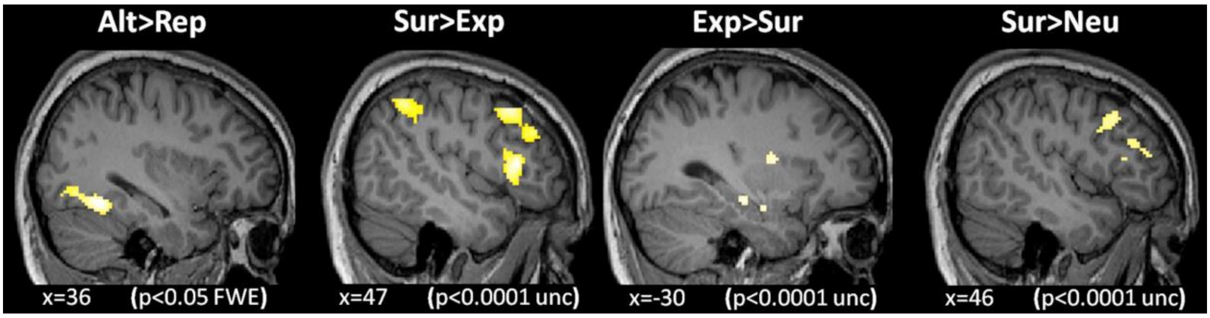

Figure 4. Results of the whole-brain analysis. Significant activations for the main effects of trial (Alt>Rep) and for the main effects of expectation conditions (Sur>Exp; Exp>Sur and Sur>Neu) are presented (cluster extent of >20 voxels).