Real-world expectations and their affective value modulate object processing

Laurent Caplette[1], Frédéric Gosselin[1], Martial Mermillod[2] et Bruno Wicker[1,3]

[1]Département de psychologie, Université de Montréal, Montréal, Québec, H3C 3J7, Canada
[2]LPNC, CNRS & Université Grenoble Alpes, 38058 Grenoble, France
[3]LNC, CNRS & Aix-Marseille Université, 13331 Marseille, France

Abbreviated title: Expectations and affect modulate object processing

Number of pages: 24
Number of figures: 3
Number of tables: 1
Number of words in the Abstract: 223
Number of words in the Significance Statement: 120
Number of words in the Introduction: 704
Number of words in the Discussion: 1465

Conflicts of interest: The authors declare no competing financial interests.

Acknowledgements: This study was funded by the Fondation Planiol (BW), by the Institut Universitaire de France (MM) and by the Social Sciences and Humanities Research Council of Canada (LC).

Correspondence should be addressed to:
Bruno Wicker
LNC, Aix-Marseille Université
3, Place Victor-Hugo
13331 Marseille cedex 3, France
Email : bruno.wicker@univ-amu.fr
Phone : +33(0)413550900

Abstract

It is well known that expectations influence how we perceive the world. Yet the neural mechanisms underlying this process remain unclear. Studies have focused so far on artificial contingencies between simple neutral cues and events. Real-world expectations are however often generated from complex associations between potentially affective contexts and objects learned over a lifetime. In this study, we used fMRI to investigate how object processing is influenced by neutral and affective context-based expectations. First, we show that the precuneus, the inferotemporal cortex and the frontal cortex are more active during object recognition when expectations have been elicited a priori, irrespectively of their validity or their affective intensity. This result supports previous hypotheses according to which these brain areas integrate contextual expectations with object sensory information. Notably, these brain areas are different from those responsible for simultaneous context-object interactions, dissociating the two processes. Then, we show that early visual areas, on the contrary, are more active during object recognition when no prior expectation has been elicited by a context. Lastly, BOLD activity was shown to be enhanced in early visual areas when objects are less expected, but only when contexts are neutral; the reverse effect is observed when contexts are affective. This result supports recent proposals that affect modulates predictions in the brain. Together, our results help elucidate the neural mechanisms of real-world expectations.

Significance statement

It is well known that expectations shape how we perceive the world. However, the precise mechanisms remain unclear and studies often used stimuli that lack ecological validity. In the present fMRI study, we assessed the effect of real-world expectations initiated by neutral and affective contexts on the neural mechanisms of object recognition. We first show evidence for previous claims that the precuneus and the inferotemporal cortex integrate contextual expectations with sensory information. Our results also suggest that scene-based predictions and instantaneous scene-object interactions are different processes. Finally, we show that the enhanced response usually observed with unexpected objects is reversed when contexts are affective. This result supports a recent proposal concerning the role of affect in the initiation of predictions.

We expect to find hairdryers in bathrooms, tombstones in cemeteries, and baguettes in bakeries, but more rarely tombstones in bathrooms, refrigerators in cemeteries and hairdryers in bakeries. That is, we live in a world where most objects are associated with specific contexts. Throughout a lifetime of experiences, we come to learn these associations, which lead us to form expectations about the objects to be encountered when we navigate the world.

Perception can be understood as the process of integrating such top-down expectations with incoming sensory information. It has been proposed that predictions from high-level areas are transmitted to adjacent lower-level areas and compared with incoming sensory signals, such that only the discrepancy between these two signals – the prediction error – is transmitted up the visual hierarchy (Friston, 2005; see also Mumford, 1992; Ullman, 1995; Rao and Ballard, 1999). In support of this model, expectation of a visual stimulus elicits a specific pattern of activity in the primary visual cortex (Kok et al., 2014, 2017; Hindy et al., 2016) and the perception of an expected stimulus results in reduced neural activity in sensory cortices (Summerfield et al., 2008; den Ouden et al., 2010; Alink et al., 2010; Kok et al., 2012a; Todorovic and de Lange, 2012; see de Lange et al., in press, for a review). Some predictions, however, may require a different mechanism than feedback from adjacent visual areas (Hindy et al., 2016): for instance, the hippocampus has been shown to play a role in the generation of predictions (Hindy et al., 2016; Kok and Turk-Browne, in press), and there is some evidence that parahippocampal (PHC) and retrosplenial (RSC) cortices initiate context-based expectations (Bar, 2003; Bar and Aminoff, 2003; Bar, 2004; Bar et al., 2006; Livne and Bar, 2016; Brandman and Peelen, 2017).

Most studies examining the effect of predictions on perception have used very simple cues such as tones (Summerfield and Koechlin, 2008; den Ouden et al., 2010; Kok et al., 2012a, 2017) or a repetition of the same object (Summerfield et al., 2008; Todorovic and de Lange, 2012). By contrast, expectations about everyday objects usually stem from the surrounding

context. Several previous studies investigated context-object relationships, but they used a simultaneous presentation of the object and the scene (Goh et al., 2004; Jenkins et al., 2010; Kirk, 2008; Rémy et al., 2014), which makes it hard to disentangle scene-object interactions from scene-based predictions (which occur prior to the object's recognition). To our knowledge, this is the first study to explore sequential context-object interactions.

Relatedly, the effect of predictions has not been considered in the setting of an ecological object recognition task. Simple detection tasks (Jiang et al., 2013), delayed discrimination tasks (Kok et al., 2012a, 2014, 2017) or categorization tasks using few alternatives (den Ouden et al., 2010; Kok et al., 2012b) are typically used. Moreover, previous studies on prediction have manipulated predictability by artificial means, either by repeating and alternating stimuli (Summerfield et al., 2008; Todorovic and de Lange, 2012), by having stimuli appearing after different cues with different probabilities during the experiment (den Ouden et al., 2010; Kok et al., 2012a, 2012b, 2014, 2017; Jiang et al., 2013), or by developing arbitrary contingencies shortly before the experiment (Hindy et al., 2016). Associations between contexts and objects formed over a lifetime of experiences may involve mechanisms distinct from these. For instance, real-world expectations are often tinted by some affective value. A visual context can elicit emotional reactions that may influence the recognition of objects in the scene (Lebrecht et al., 2012). In an emotional context (e.g., a cemetery), the affective value may be partially processed before the scene's objects (e.g., a tombstone) and contribute to the object's recognition (Barrett and Bar, 2009). Alternatively, the prediction's affective value might interact with its validity: this is likely to result in a reversal of the prediction error effect in the brain (Miller and Clark, 2018).

In the present study, we aimed to address these shortcomings by investigating how realistic object recognition mechanisms are influenced by task-irrelevant high-level expectations generated by a predictive or non-predictive visual context. The use of everyday

objects and scenes allowed us to use associations between objects and contexts formed over a lifetime of experiences, and to compare affective and neutral expectations.

Materials and Methods

Participants

Seventeen healthy adults (9 female; mean age = 24.8; SD = 4.3) were recruited on the campus of Aix-Marseille Université. Participants did not suffer from any neurological, psychological or psychiatric disorder and were free of medication. The experimental protocol was approved by the ethics board of CPP Sud-Méditerranée 1 and the study was carried in accordance with the approved guidelines. Written informed consent was obtained from all participants after the procedure had been fully explained, and a monetary compensation was provided upon completion of the experiment.

Stimuli

In a first validation study, 35 different subjects were shown thirty-three context names and had to give the names of three objects with a high probability of being present in that context. Then, thirty-three public domain scene color images were selected from the internet as context images (see examples in figure 1a). Context images were selected to ensure, as much as possible, that their three most associated objects did not appear in them (while still being representative of the context category). However, this was not always possible: in 11 instances (6 neutral and 5 affective contexts), one associated object appeared somewhere in the scene, mostly at a small scale in the background (in all these cases, the object image later chosen was of a different exemplar). In a second validation study, an independent sample of 22 subjects identified what they thought the context images represented (to confirm that the image

represented the context), indicated if the context elicited an emotion and, if so, what were the valence (negative to positive, from 0 to 10) and intensity (no emotion to very intense emotion, from 0 to 10). Following this study, 32 visual scenes were selected (one scene was excluded) and split in Affective (e.g., cemetery, beach, luxury hotel) and Neutral (e.g., swimming pool, airport, kitchen) categories at the median of the intensity scores (5.19); valence was not included in the experimental design. On average, neutral contexts had an intensity of 3.19 and a valence of 6.14; affective contexts had an intensity of 6.10 and a valence of 5.53.

Ninety-six color images of objects corresponding to the three most cited names for each context were then selected for the experiment (e.g., swimsuit, diving board and pool ladder for swimming pool; see other examples in figure 1a). For each context, the experimenters also chose three non-associated objects (selected from the ones that had never been associated with the context in the second validation study). Every object was the associated object of only one context and the non-associated object of only one other context; moreover, for each context, each one of the three non-associated objects was associated with a different context. A third and final validation study was conducted to collect quantitative measures of the associations between objects and contexts. Forty-four new subjects indicated on a scale from 0 to 10 how much each object was associated to its predictive and non-predictive contexts (context-object pairs were randomized). Measures were z-scored within each subject and averaged across them.

Finally, we randomized the phases of the mean of the context images in the Fourier domain – separately for each RGB color channel – to obtain 96 different phase-scrambled images.

Data acquisition

Functional imaging data were acquired with an ADVANCE 3 Tesla scanner (Bruker Inc., Ettlingen, Germany) equipped with a 2-channel head-coil. Functional images sensitive to

BOLD contrast were acquired with a T2*-weighted gradient echo EPI sequence (TR 2400 ms, TE 30 ms, matrix 64 x 64 mm, FOV 192 mm, flip angle 81.6°). Thirty-six slices with a slice gap of 0 mm were acquired within the TR; voxels were 3 x 3 x 3 mm. Between 303 and 311 volumes were acquired in each run, excluding the six dummy scans acquired at the beginning of each run for signal stabilization. Additionally, a high resolution (1 x 1 x 1 mm) structural scan was acquired from each participant with a T1-weighted MPRAGE sequence.

Experimental Design and Statistical Analysis

The LabVIEW (National Instruments Inc., Austin, TX, USA) software was used to project stimuli during the experiment. Stimuli were projected to a screen positioned in the back of the scanner using a video projector. Subjects could see the video reflected in a mirror (15 x 9 cm) suspended 10 cm in front of their face and subtending visual angles of 42 degrees horizontally and 32 degrees vertically.

Each trial was built as follows: a large cue image (see below) spanning the whole screen during 1 s, a black screen during 1.5 to 4 s (duration randomly selected from a truncated exponential distribution with mean of 2 s), a centered object image on a black background during 133 ms, a black screen during 1.5 to 4 s, and an object name on a black background shown until the subject answered or for a maximum of 1 s (Figure 1b). Subjects answered by pressing one of two buttons on a hand-held response device to indicate if the name corresponded to the object, which occurred on 80% of the trials. A black screen was displayed for an additional 1 s between trials.

On a third of the trials (Predictive condition), the cue image was a scene associated with the object following it (e.g., an airport and a suitcase); on another third (Non-Predictive condition), it was a scene not associated with the object following it (e.g., a church and a tennis racket); on the final third (No-Context condition), it was a scrambled image (always a different

one). Each object was shown once in each of these conditions, for a total of 288 trials. Furthermore, Predictive and Non-Predictive conditions were each split evenly into Affective and Neutral subconditions, following the affective intensity of the context. There was therefore a total of 5 conditions: Predictive Affective (or Pred-Aff for short), Predictive Neutral (Pred-Neut), Non-Predictive Affective (noPred-Aff), Non-Predictive Neutral (noPred-Neut) and No-Context (noCont).

The order of trials was randomized. Randomized trials were divided in 3 fixed functional data acquisition runs of 96 trials. Each functional run lasted between 10 and 12 mins, with short breaks between them. The order of the 3 runs was counterbalanced across subjects.

For preprocessing and statistical analysis, the SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/), running in the MATLAB environment (Mathworks Inc., Natick, MA, USA), was used. T1-weighted structural images were segmented into white matter, gray matter and cerebrospinal fluid, and warped into MNI space. Functional images were

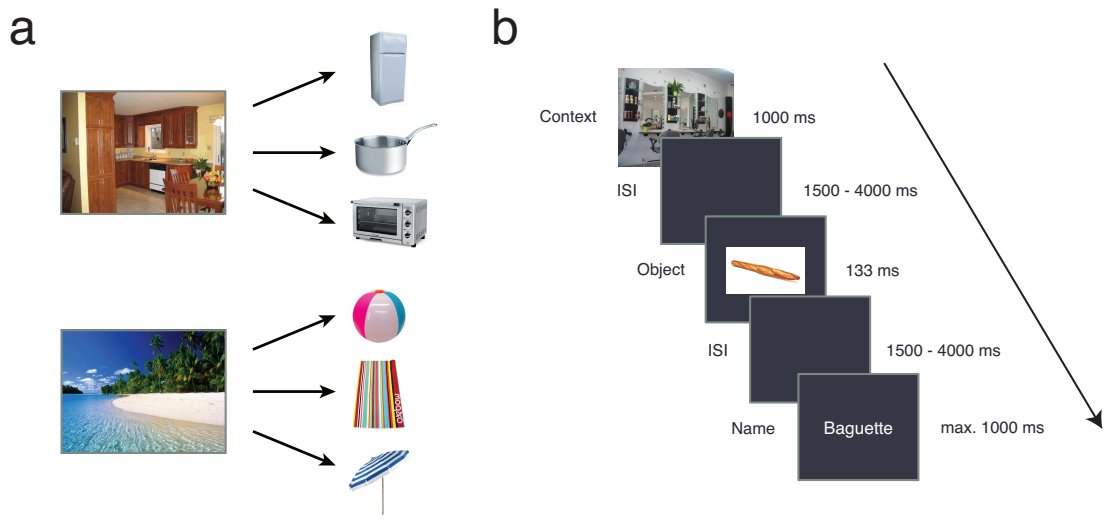

Figure 1. A) Example of one neutral and one affective scene, with their associated objects. See table S1 for a list of all contexts and associated objects. **B)** Example of a trial (Non-Predictive Neutral condition). The object image and name have been enlarged for better viewing.

realigned, unwarped and corrected for geometric distortions using the field map of each participant, slice time corrected, coregistered to the structural image of the corresponding participant, and smoothed using a 6 mm FWHM isotropic Gaussian kernel.

A standard GLM analysis was performed for each subject. Three events were modelled on each trial: contexts (or scrambled images), objects and object names. Object events (the regressors of interest) were modelled for each condition separately (Pred-Aff, Pred-Neut, noPred-Aff, noPred-Neut and noCont); scene events (regressors of no interest) were also modelled separately for each condition; one additional regressor was included for the object names. All these events were modelled as Dirac delta functions (duration of zero) convolved with SPM8's canonical hemodynamic response function. To get rid of potential effects caused by differences in context-object associations, we included an additional parametric regressor which consisted of the context-object associations as determined by our third validation study. This regressor was z-scored separately within predictive contexts and non-predictive contexts but not separately within each subcondition so that differences in context-object associations between affective and neutral contexts were accounted for, but that differences between predictive and non-predictive conditions remained; finally, we convolved it with the hemodynamic response function. The six motion parameters were also included as additional nuisance regressors.

A temporal high-pass filter (cut-off of 128 s) was used to remove low-frequency drifts, and temporal autocorrelation across scans was modelled with an AR(1) process. Contrasts were then computed at the subject level and used for group analyses using one-sample *t*-tests. All voxels inside the brain were analyzed; we maintained the familywise error rate of $p < .05$, two-tailed, at the cluster level (primary threshold of $p < .001$, uncorrected) using random field theory (Friston et al., 1994). The Anatomy (Eickhoff et al., 2005) and WFU-PickAtlas (Maldjian et al., 2003) toolboxes were used to identify activated brain regions based on peak Montreal Neurological Institute (MNI) coordinates.

Results

Behavioral results

Mean accuracy was 97.1% (σ = 2.8%) for the Pred-Neut condition, 97.2% (σ = 2.1%) for the Pred-Aff condition, 96.9% (σ = 2.6%) for the noPred-Neut condition, 95.0% (σ = 2.5%) for the noPred-Aff condition and 95.7% (σ = 2.4%) for the noCont condition. When comparing Pred, noPred and noCont together (ANOVA, n = 17), there was no effect of condition on accuracy ($F(2,16) = 2.17$, $p = .13$, $\eta^2_p = 0.12$). When comparing all conditions except noCont together (ANOVA, n = 17), there was a significant main effect of predictive value ($F(1,16) = 10.36$, $p = .005$, $\eta^2_p = 0.13$), a marginally significant main effect of affective value ($F(1,16) = 4.42$, $p = .052$, $\eta^2_p = 0.08$), and a marginally significant interaction between affective and predictive values ($F(1,16) = 3.97$, $p = .063$, $\eta^2_p = 0.10$).

Mean response time was 640 ms (σ = 121 ms) for the Pred-Neut condition, 650 ms (σ = 120 ms) for the Pred-Aff condition, 638 ms (σ = 110 ms) for the noPred-Neut condition, 636 ms (σ = 127 ms) for the noPred-Aff condition and 632 ms (σ = 116 ms) for the noCont condition. When comparing Pred, noPred and noCont together (ANOVA, n = 17), there was no

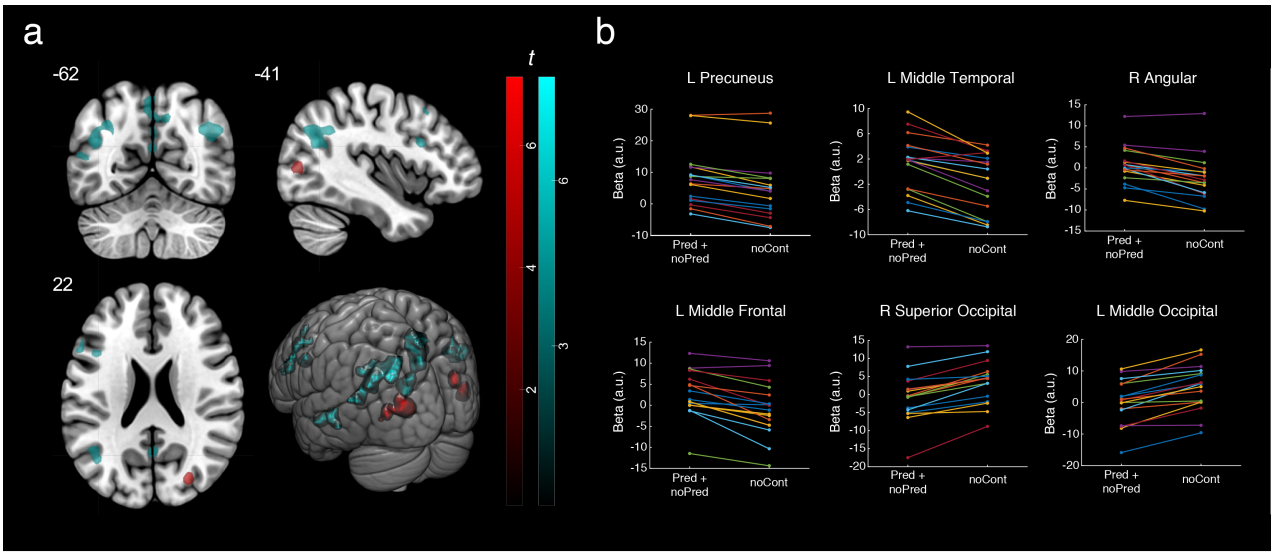

Figure 2. A) Significant clusters for the noCont > (Pred + noPred) (in red) and the (Pred + noPred) > noCont (in cyan) contrasts. **B)** Beta values of individual subjects for noCont and (Pred + noPred) conditions in peak voxels of various significant clusters.

effect of condition on response time ($F(2,16) = 1.53$, $p = .23$, $\eta^2_p = 0.09$). When comparing all conditions except noCont together (ANOVA, $n = 17$), there was no main effect of affective or predictive value and no interaction (Fs(1,16) = 1.67, 0.25 and 0.64 respectively, $p > .20$, $\eta^2_p <$ .03).

fMRI results

To investigate the potential effect of the generation of explicit contextual expectations (occurring only in the Pred and noPred conditions) on brain activity, we contrasted the Pred and noPred conditions with the noCont condition (paired t-test, $n = 17$). Five clusters were significantly more activated in the Pred and noPred conditions than in the noCont condition ($p < .05$, two-tailed, corrected for family-wise error rate (FWER); peak Cohen's $d_z = 1.91$; Figure 2; Table 1): one bilateral cluster in the precuneus, one extending from the left precuneus and middle occipital gyrus to the left angular gyrus, one in the left middle temporal gyrus, one in the left middle and inferior frontal gyri and one in the right angular gyrus. The reverse contrast revealed the specific activation of two clusters in the right superior and

middle occipital gyri and in the left middle occipital gyrus ($p < .05$, two-tailed, FWER-corrected; peak Cohen's $d_z = 1.73$; Figure 2; Table 1).

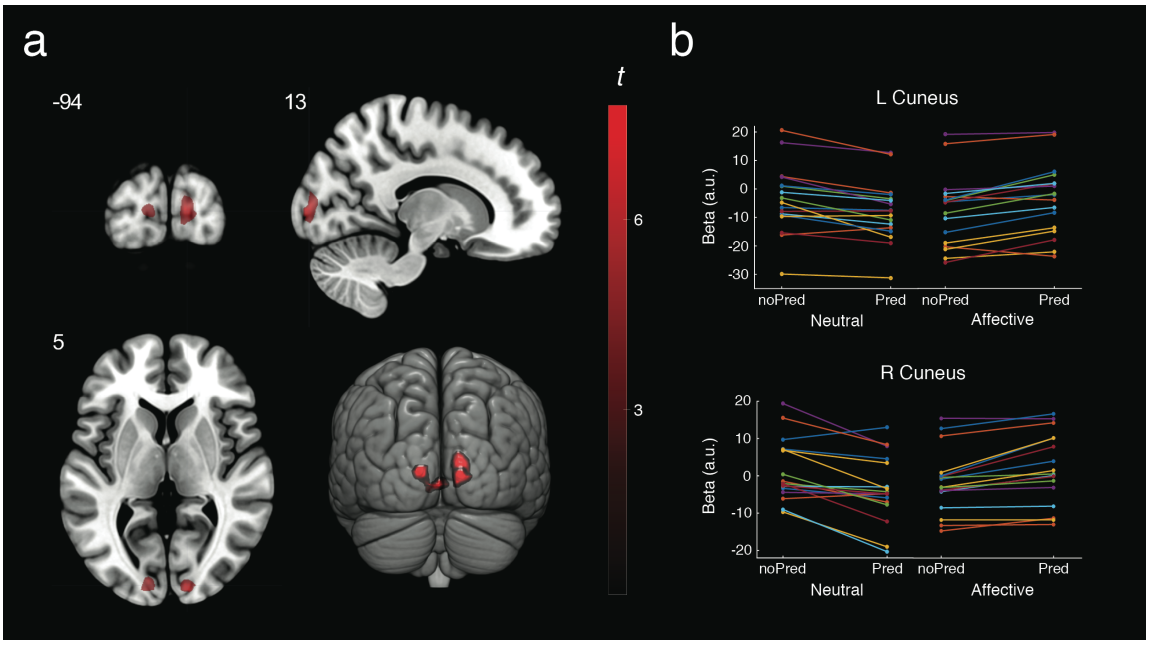

Figure 3. A) Significant clusters for the interaction between affective and predictive values (in red). **B)** Beta values of individual subjects for each condition in peak voxels of significant clusters.

We then investigated whether there was a main effect of predictive value (Pred vs noPred), a main effect of the context's affective value (Aff vs Neut), and an interaction between predictive and affective values on brain areas involved in object recognition (paired t-tests, $n = 17$). There were no significant main effects of predictive and affective values. However, there was a significant interaction between predictive and affective values for two clusters: one in the right cuneus and one overlapping the left cuneus, calcarine gyrus and lingual gyrus ($p < .05$, two-tailed, FWER-corrected; peak Cohen's $d_z = 1.74$; Figure 3; Table 1). We then investigated what simple effects resulted in this interaction: when looking at the simple effects on the peak voxels of each significant cluster, we observed that they were more active in the Pred-Aff condition than in the noPred-Aff condition (left cuneus: $t(16) = 4.80$, $p_{Bonf} = .0008$, $d_z = 1.16$; right cuneus: $t(16) = 4.56$, $p_{Bonf} = .001$, $d_z = 1.11$) and more active in the noPred-Neut than in the Pred-Neut condition (left cuneus: $t(16) = 4.41$, $p_{Bonf} = .002$, $d_z = 1.07$; right cuneus: $t(16) = 4.24$, $p_{Bonf} = .003$, $d_z = 1.03$).

Next, we conducted a series of control analyses to ensure that the interaction could not have been the result of undesirable confounds. First, we investigated whether the interaction could have been caused by differences between the objects associated to neutral contexts and those associated to affective contexts by assessing if there was any significant difference in brain activity when they were perceived without a context (noCont condition). There was no significant difference between the conditions ($p_{FWER} > .33$). We also analyzed the image similarities directly: we used the HMAX model (Riesenhuber & Poggio, 1999; Serre et al., 2007), a commonly used model of the early visual cortex, and we computed correlation distances between the responses of the model to each image. We then verified if the between-categories (affective context objects to neutral context objects) distances were larger than the within-categories distances (two sample t-tests): no difference was observed (compared to within-neutral distances: .498 vs .504, $t(3430) = .57$, $n = 1128$ and 2304, $p = .57$; compared to within-affective distances: .498 vs .502, $t(3430) = .38$, $n = 1128$ and 2304, $p = .70$), indicating that we could not find any evidence of a distinction between these two object categories.

Finally, the possibility remained that attention could explain the interaction between affective and predictive values: a similar interaction has indeed been previously reported with attention as a factor instead of affective value (Kok et al., 2012b). A first objection to this claim would be that our behavioral results actually point to an opposite effect: although we observe the same reversed prediction effect for affective contexts that Kok et al. observed for task-relevant stimuli, the lower recognition accuracy in the affective condition suggests that they are not attended more and that attention is not the cause of this interaction. Nonetheless, we decided to conduct an additional behavioral experiment to isolate potential attentional effects better. Twenty-four participants performed a Gabor orientation discrimination task (vertical vs horizontal), in which the Gabor patches (1 cycle per degree) were randomly following either a neutral context image or an affective context image in the same way as in the fMRI experiment

(contexts presented for 1s, 1.5-4 s jitter, patches presented during 133 ms); adaptive procedures were conducted separately in each condition in order to find the contrast sensitivity threshold associated with each condition. Again, no difference was observed ($\log_{10}$(contrast) of -2.10 vs -2.11; p = .94). Since we know that contrast sensitivity is greatly enhanced by attention (see Carrasco, 2006, for a review), it does not seem likely that affective contexts were attracting attention and maintaining it for up to 4s in order for it to alter object processing.

Table 1. Montreal Neurological Institute (MNI) coordinates and T values for significantly activated brain regions.

Brain regions	**Peak MNI coordinates (mm)**			**Nb of voxels**	**Peak T value**
	x	**y**	**z**		
(Pred + noPred) > noCont					
L Middle Occipital	-30	-69	39	232	7.88
L Angular	-42	-63	27		6.63
L Middle Temporal	-54	-21	-9	76	6.92
R Angular	51	-63	36	90	6.16
L Middle Frontal	-48	18	39	94	5.79
L Inferior Frontal	-42	15	24		5.51
L Precuneus	-3	-66	54	224	5.59
R Precuneus	9	-57	42		5.32
noCont > (Pred + noPred)					
R Superior Occipital	27	-78	21	61	7.12
R Middle Occipital	36	-81	12		4.75
L Middle Occipital	-27	-84	6	90	6.76
Predict. x Affect. Interaction					
R Cuneus	15	-96	6	45	7.79
L Cuneus	-12	-90	3	43	6.08
	0	-87	-3		4.55
	-9	-84	-9		3.89

Discussion

Our first aim was to investigate how the generation of expectations about objects from a preceding context might modulate the activity of brain areas involved in object perception. We found significantly more activation in the precuneus, the left middle occipital gyrus, the left middle temporal gyrus, the left frontal cortex and the parietal cortex, when (valid or invalid) contextual expectations were generated prior to object perception, suggesting that these high-level areas are mainly associated with object processing when expectations are generated. These activations specifically represent an interaction between contextual expectations and object bottom-up sensory information: activity related solely to object processing is cancelled out because the objects are the same in both conditions, and activity related solely to the prior presentation of the context is regressed out in the GLM.

To our knowledge, only Summerfield & Koechlin (2008) performed a similar analysis before; however, they used lines as cues and gratings as stimuli, and the cue was directly related to the task (the subjects had to indicate whether the cue and the grating matched). In their study, they observed a significantly greater activation of the middle occipital and fusiform gyri when there was an expectation. We also find a greater activation of the middle occipital gyrus, in addition to many other brain regions. Since expectations in our study are about objects rather than simple grating orientations, regions representing them are likely to be more numerous. The interaction between object and context processing observed in the middle temporal gyrus (a part of the inferotemporal cortex) supports a popular hypothesis according to which top-down contextual predictions would be combined with bottom-up sensory information to facilitate object recognition in the inferotemporal cortex (Bar, 2004). The precuneus and the parietal cortex, which are also activated in this contrast, have previously been linked to episodic memory retrieval and contextual associative processing (Lundstrom et al., 2005; Aminoff et al., 2007; Livne and Bar, 2016; Brandman and Peelen, 2017) which both require the integration of

stored representations with incoming sensory information. Moreover, the precuneus of an observer that views several objects simultaneously is more activated when these objects are contextually related than when they are not (Livne and Bar, 2016); this suggests that the contextual representations elicited by some of these objects are compared to other objects. Recently, activity in the retrosplenial complex, a region comprising the precuneus, has been shown to correlate with supra-additive decoding of objects embedded in scenes, suggesting that the precuneus is responsible for a scene-based facilitation of object representations (Brandman and Peelen, 2017). Interestingly, the interaction we observed between context and object information in the precuneus is also supra-additive (i.e. there is a remaining positive activation after considering the main effects of object and context). We extend previous results by showing that the precuneus integrates object sensory information with valid or invalid scene-based expectations generated prior to object presentation. The inferior and middle frontal gyri were also active during object processing when expectations were generated. These regions have previously been found to respond more to objects in non-congruent scenes than to objects in congruent scenes (Rémy et al., 2014): it is thus likely that they are responsible of integrating contextual information with perceived objects. Other frontal areas have previously been found to both maintain expectations and integrate them with sensory information (Summerfield et al., 2006; Summerfield & Koechlin, 2008).

When investigating which regions were decoding objects *in* scenes better than objects *and* scenes (in a supra-additive manner), Brandman & Peelen (2017) reported lateral extrastriate loci of activations, including the lateral occipital cortex and the posterior fusiform sulcus. These regions largely differ from the ones we uncovered (most notably the precuneus and the frontal cortex), suggesting that the matching of automatic contextual expectations with sensory evidence recruits different regions than the ones involved in the simultaneous integration of object and background. This result implies that these two processes may be distinct.

The reverse contrast, associated with visual processing of objects when no expectation (neither valid nor invalid) had been generated from a context, yielded bilateral activation of primary visual areas. Activated voxels may be part of areas primarily associated with the processing of sensory information shared by a majority of objects (e.g., intermediate spatial frequencies; Caplette et al., 2014), which is thus reduced when almost any object is expected.

We then investigated whether there was an effect of prediction error or match, i.e. whether some areas were more active at the presentation of the object when the object followed a predictive context or when the object followed a non-predictive context. When neutral and affective contexts were combined, there was no significant difference between predictive and non-predictive conditions; however, there was a significant interaction between predictive and affective values in low-level occipital areas, specifically the left and right cunei. Looking at these clusters, the classical prediction error effect was visible for neutral contexts, i.e. predicted objects elicited a smaller BOLD signal; but, when contexts were affective, this effect was reversed, i.e. predicted objects elicited a larger BOLD signal. Note that previous studies observing a smaller signal for predicted objects have exclusively used affectively neutral cues, making our results compatible with theirs. Furthermore, these brain regions are different from those responding differentially to congruent and incongruent context-object pairs, typically higher-level regions such as the lateral occipital and frontal cortices (e.g., Jenkins et al., 2010; Rémy et al., 2014). This further indicates that scene-object interactions and scene-based expectations are different processes.

These results are not compatible with the proposal that a subject's internal affective state is altering the content of their predictions about object identities (Barrett and Bar, 2009). According to this idea, the affective value of a preceding context (or even a simultaneous context or the object itself; see Barrett and Bar, 2009) would alter the subject's bodily state and bring additional information that could be used by the brain to predict the identity of perceived

objects. Consequently, a similar pattern of results should be visible for neutral and emotional contexts, with only a greater difference in activation between predicted and unpredicted objects for emotional contexts than for neutral contexts (due to the additional emotional information).

Our results are compatible, however, with the general idea that affect interacts with predictive processing (Barrett and Simmons, 2015; Miller and Clark, 2018). One possibility recently put forward by some authors is that, rather than contributing to the content of the predictions, a subject's internal affective state modulates the precision of the predictions (Miller and Clark, 2018). In recent formulations of predictive coding (Feldman and Friston, 2010), the prediction error is weighted by the reliability, or precision, of sensory information. When precision is low, prediction errors are down-weighted and observers rely more on predictions; when it is high, prediction errors are up-weighted and observers rely more on the sensory input. Because this weighting only occurs for neurons representing prediction error and not for neurons representing predictions (Friston, 2009; Kok et al., 2012b), it should lead to an interaction between prediction error and precision (Rao, 2005; Friston, 2009; Kok et al., 2012b), exactly like the one we observed between predictive and affective values.

Kok and colleagues (Kok et al., 2012b) reported a similar reversal of the prediction error effect in the early visual cortex for task relevant stimuli. They argued that this effect was caused by endogenous attention enhancing the precision of the predictions (Rao, 2005; Feldman & Friston, 2010). This cannot be the cause of the effect we observed however, since attention was not manipulated in our study and our stimuli were all similarly task relevant. Furthermore, exogenous attention also similar between our conditions, as revealed by behavioral results obtained in the scanner and in the control contrast sensitivity experiment. This implies that the prediction error reversal in our study was not caused by an increase in attention.

In summary, real-world expectations initiated by contexts, irrespectively of their degree of validity, led to more activation of high-level areas (including parietal and frontal cortices)

during subsequent object recognition; notably, these regions were distinct from those responsible of instantaneous scene-object interactions. Furthermore, the context's affective value interacted with the validity of the prediction it had initiated: classical prediction error effects were only observed with neutral contexts, and a complete reversal of these effects was observed when contexts were emotional. This result is not compatible with the idea that the affective value of a stimulus, and the ensuing internal bodily state of the subject, are contributing to the creation of predictions (Barrett and Bar, 2009); but it is compatible with a modulatory role of affective value over the weight of predictions in perception (Miller and Clark, 2018). In conclusion, our results deepen our understanding of predictive coding in an ecological setting by showing that the mere presence of explicit expectations, and their affective content, modulate object recognition.

References

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. J Neurosci 30(8):2960–2966.

Aminoff EM, Gronau N, Bar M (2007) The parahippocampal cortex mediates spatial and nonspatial associations. Cereb Cortex 17(7):1493–1503.

Bar M (2003) A cortical mechanism for triggering top-down facilitation in visual object recognition. J Cogn Neurosci 15(4):600–609.

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5(8):617–629.

Bar M, Aminoff E (2003) Cortical analysis of visual context. Neuron 38(2):347–358.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A 103(2):449–454.

Barrett LF, Bar M (2009) See it with feeling: affective predictions during object perception. Philos Trans R Soc Lond B Biol Sci 364(1521):1325–1334.

Barrett LF, Simmons WK (2015) Interoceptive predictions in the brian. Nat Rev Neurosci 16:419-429.

Brandman T, Peelen MV (2017) Interaction between scene and object processing revealed by human fMRI and MEG decoding. J Neurosci 37(32):7700–7710.

Caplette L, West G, Gomot M, Gosselin F, Wicker B (2014) Affective and contextual values modulate spatial frequency use in object recognition. Front Psychol 5:512.

Carrasco M (2006) Covert attention increases contrast sensitivity: psychophysical, neurophysiological and neuroimaging studies. Prog Brain Res 154A:33–70.

de Lange FP, Heilbron M, Kok P (2018) How do expectations shape perception? Trends Cogn Sci 22(9):764–779.

den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal prediction error modulates cortical coupling. J Neurosci 30(9):3210–3219.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25(4):1325–1335.

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

Friston KJ (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360(1456):815–836.

Friston KJ (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13(7):293–301.

Friston KJ, Worsley KJ, Frackowiak RSJ (1994) Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1:214-220.

Goh JOS, Siong SC, Park D, Gutchess A, Hebrank A, Chee MWL (2004) Cortical areas involved in object, background, and Object-Background Processing Revealed with Functional Magnetic Resonance Adaptation. J Neurosci 24(45):10223–10228.

Gronau N, Neta M, Bar M (2008) Integrated contextual representation for objects' identities and their locations. J Cogn Neurosci 20(3):371–388.

Hindy NC, Ng FY, Turk-Browne NB (2016) Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat Neurosci 19(5):665–667.

Jenkins LJ, Yang YJ, Goh J, Hong YY, Park DC (2010) Cultural differences in the lateral occipital complex while viewing incongruent scenes. Soc Cogn Aff Neurosci 5:236–241.

Jiang J, Summerfield C, Egner T (2013) Attention sharpens the distinction between expected and unexpected percepts in the visual brain. J Neurosci 33(47):18438–18447.

Kirk U (2008). The neural basis of object-context relationships on aesthetic judgment. PLoS ONE 3(11):e3754.

Kok P, Failing MF, de Lange FP (2014) Prior expectations evoke stimulus templates in the primary visual cortex. J Cogn Neurosci 26(7):1546–1554.

Kok P, Jehee JFM, de Lange FP (2012a) Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75(2) :265–270.

Kok P, Mostert P, de Lange FP (2017) Prior expectations induce prestimulus sensory templates. Proc Natl Acad Sci U S A 114(39):10473–10478.

Kok P, Rahnev D, Jehee JFM, Lau HC, de Lange, FP (2012b) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22(9):2197–2206.

Kok P, Turk-Browne NB (2018). Associative prediction of visual shape in the hippocampus. J Neurosci 38(31):6888–6899.

Lebrecht S, Bar M, Barrett LF, Tarr MJ (2012) Micro-valences: perceiving affective valence in everyday objects. Front Psychol 3:107.

Livne T, Bar M (2016) Cortical integration of contextual information across objects. J Cogn Neurosci 28(7):948–958.

Lundstrom BN, Ingvar M, Petersson KM (2005) The role of preuneus and left inferior frontal cortex during source memory episodic retrieval. Neuroimage 27(4):824–834.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19(3):1233–1239.

Miller M, Clark A (2018) Happily entangled: prediction, emotion, and the embodied mind. Synthese 195(6):2559–2575.

Mumford D (1992) On the computational architecture of the neocortex: II. The role of cortico-cortical loops. Biol Cybern 66:241.

Rao RPN (2005) Bayesian inference and attentional modulation in the visual cortex. Neuroreport 16(16):1843–1848.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2(1):79–87.

Rémy F, Vayssière N, Pins D, Boucart M, Fabre-Thorpe M (2014) Incongruent object/context relationships in visual scenes: Where are they processed in the brain? Brain Cogn 84(1):34–43.

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025.

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T (2007) Robust object recognition with cortex-like mechanisms. IEEE Trans on Pattern Anal Mach Intell 29:411–426.

Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006) Predictive codes for forthcoming perception in the frontal cortex. Science 314(5803):1311–1314.

Summerfield C, Koechlin E (2008) A neural representation of prior information during perceptual inference. Neuron 59(2):336–347.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11(9):1004–1006.

Todorovic A, de Lange F O (2012) Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. J Neurosci 32(39):13389–13395.

Ullman S (1995) Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. Cereb Cortex 5(1):1–11.