The Role of Expectation and Surprise in Fusiform Face Area Activity

Solvej Mathiasen, Mathias Houe Andersen, Adam Finnemann & Line Kruse

Cognitive Science, University of Aarhus

Jens Chr. Sous Vej 2, 8000 Aarhus, Denmark

The Fusiform Face Area (FFA) in the fusiform gyrus has been found to be selectively active when perceiving faces, and interpreted as reflecting a specific processing mechanism for face features. However, this idea has recently been challenged by the paradigm of predictive coding, suggesting that the FFA activity might be better understood as reflecting the expectation or surprise of a face, putting more emphasis on top-down processing. This study tested the two theories using an fMRI design, and found activation only when participants were presented with actual face stimuli, regardless of their expectations of seeing a face, supporting a feature-detection theory. Limitations of the study are discussed as well as the relevance of further testing the theories of predictive coding.

Introduction

For a while it has been commonly agreed that the fusiform gyrus is particularly important for the perception of faces. The first indication of the more specific Fusiform Face Area (FFA), was given by Kanwisher, McDermoot and Chun (1997), who found this region to be significantly more active when perceiving faces than other common objects. Additionally it has been demonstrated, how patients with damage in the occipitotemporal region of the right hemisphere, have acquired prosopagnosia, a selective face recognition deficit (De Renzi et al., 1994). The FFA activity, has commonly been explained by feature detection theories; the idea that recognition is based on independent detection of features, or components, of an image in the world. (Pelli et al., 2006). Thus, in this view the population response of the FFA is primarily a bottom-up analysis, driven by the physical characteristics of a face.

A competing account to feature-detection in this domain is the generative model of predictive coding, in which perceptual inference is explained as a concurrent matching process between top-down- and bottom-up information throughout the visual hierarchy (Figure 1).

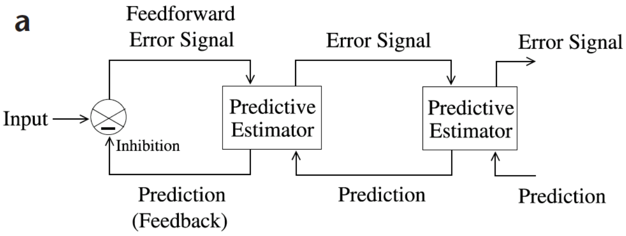

Figure 1. Hierarchical network of predictive coding (Rao and Ballard, 1999).

Egner et al. (2010) demonstrated how the evoked responses to non-face and face-stimuli in the FFA were indistinguishable under high expectations of seeing a face. They interpreted, that the FFA might be better understood as an area of face-expectation rather than an area of face detection, in line with predictive coding theories. Clark (2013) described predictive coding models as having two processing mechanisms that are computationally different, in each state of the visual hierarchy. One consists of representational units, working in a top-down fashion, encoding the probability of a stimulus, i.e. the lower level neural activities, and send predictions to the next lower level via feedback connections. The second consists of error units, coding the difference between the predictions and the bottom-up evidence, and forward prediction error to the higher level via feedforward connections. The backward connections of the representational units then construct predictions with the goal of minimizing error, based on the signals from the error units. In this cortico-cortical feedback system, recognition is thus a task of minimizing prediction error at each level of the hierarchy. Taking this view, the activity in the FFA is explained as reflecting the difference in prediction (face expectation) and prediction error (face surprise) rather than a bottom-up feature detection response (Egner et al., 2010). The central difference between the two theories is thus, that while feature detection mostly emphasizes the role of bottom-up analysis of stimulus features, predictive coding additionally includes a generative model providing top-down predictions and emphasizes the constant interaction between the two processing mechanisms.

The experiment of this study was conducted to test the two theories, feature-detection and predictive coding as accounting for the FFA responses to face stimuli. The experiment was done using a classical conditioning design, in which house and face images constituted the unconditioned stimuli. Preceding the conditioned stimuli, coloured frames were presented, predicting the possibility of either a house or a face stimulus. A feature detection approach would predict the FFA to be activated only during trials in which the unconditioned stimulus is a face, i.e. it responds only to face features. Meanwhile, the frames are expected to have no influence on FFA activity (Figure 2).

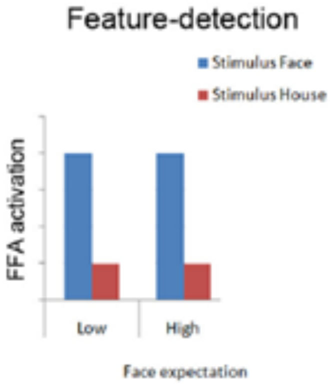

Figure 2. From Egmont, Monti & Summerfield (2010) with slight modification.

Contrary, predictive coding would expect the frames to work as conditioned stimuli, eliciting activity in the FFA doing all trials where a face is expected or constitute a surprise (Figure 3). Thus, doing trials where the frame is associated with the expectancy of a following face, but is actually followed by a house, FFA activity is expected, as a result of the conditioned stimulus, despite of the unconditioned stimulus being a house. Similarly FFA is expected to respond to trials where the frame represents low face expectation, but is followed by a face image.

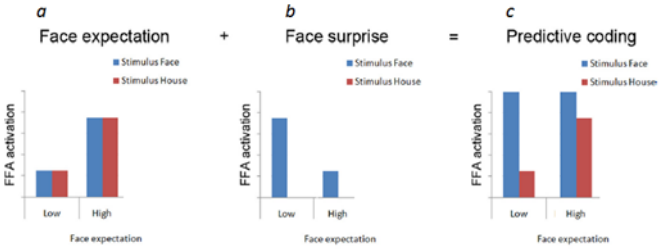

Figure 3. From Egmont, Monti & Summerfield (2010) with slight modification.

Materials and Methods

Participants

The study was conducted twice on one female participant of age 21. She was prescanned to ensure she was neurologically healthy. The study was conducted at University of Aarhus.

Stimuli

The experimental stimuli consisted of pictures of non-known faces on a neutral background and pictures of houses from a frontal view, with limited details (trees, people etc.).

All face images were in black and white, found in Google Images, had a standardized size, and roughly the same composition. The eyes (focal points) were roughly in the same height and all images had white or grey backgrounds. All faces had minimal facial expressions and the gender balance was nearly equal. We assume that all photos were unknown to the participants from the beginning of the experiment.

For house images, all pictures were similarly in black and white and found on Google images. They had a standardized size and were photos of the front facade of the house, all taken from the same perspective. The house images were differently lighted, dependent of the time of the day the picture was taken. This gave different kinds of contrasts in the frame and an uncontrolled background. The size of height was fixed. Since human faces tend to have a more stereotypical form than houses, the width of house images had more variance between the photos compared to the height of face images.

Presentation of stimuli was programmed using PsychoPy.

Procedure

Data was collected in the fMRI scanner in two periods of 12 minutes. Each period contained 143 stimuli trials. House and face stimuli was evenly distributed and presented in random order.

The participant watched a coloured frame (green or blue) alone for 700 ms, and afterwards in combination with stimulus (face or house) for 300 ms. In between stimuli trials was a delay interval, with a fixation cross presented in either 2, 3, 4 or 5 seconds, randomly distributed (see figure 4).

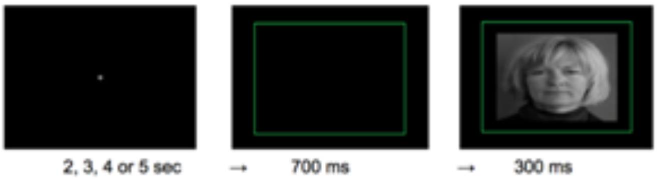

Figure 4: Example of stimuli

After each trial the participant was asked to answer whether stimulus was inverted or not. This was a distraction task, present to keep the participant active and distracted. Normal stimulus was answered with an index-finger tap, inverted stimulus with a middle-finger tap. Only 10% of the trials were inverted, as this parameter was not of direct interest to the experiment.

The goal of the experiment was to form expectations of either house or face stimuli in the participant. These expectations were achieved by pairing colour frame with stimuli type.

Sitmuli/Face Expectation	***House***	***Face***
High face expectation	**25 %** 	**75%**
Low face expectation	***75%*** 	***25%***

Figure 5: 2x2 factorial design. Four conditions: green-house, green-face, blue-house, blue-face.

A green frame was paired with face stimuli 75% of the time and with houses 25% of the time (high face expectation). Conversely, a blue frame was accompanied by houses 75% of the time and faces 25% (low face expectation). That created four conditions as a two-by-two factorial design (figure 5).

The participant was instructed of the probabilistic pairing of frame colour and stimuli, as distinguishing between explicit and implicit expectations was of no interest, and the explicit instruction facilitated greater expectations. As the frames were not predictive of the occurrence of a target stimulus (inverted vs. normal), the subject could gain no performance benefit from using the frames to guide attentional processes.

In the experiment both functional and behavioural data was collected. The behavioural data consisted of participant information (Subject ID, gender and age) as well as reaction time and correct responses.

Analysis

Behavioural analysis

Accuracy of performance was measured to ensure that the participant was paying attention doing the entire experiment. Response accuracy in this task was at ceiling, and error rates were therefore not subjected to inferential statistics.

Additionally, response times were measured, to confirm that reaction time was not affected by the manipulation of perceptual expectations. Reaction times were analyzed in a general linear model as a function of the type of frame, condition and the interaction between them.

fMRI analysis

Image acquisition:

The fMRI scanner used 240 fMRI volumes with 39 slices per volume. TR: 2 s.
Time specification: seconds.

Image analysis:

The preprocessing of the images took place in SPM12. Images were realigned to fit the first image in order to nullify movement during scans. Secondly, the realigned images were resliced so that the voxels on each image matches the voxels of the image from the first scan. The mean image from the previous step was coregistered with the structural scan. Segmentation processing segmented the coregistered image into different types of tissue, for instance white and grey matter, bone and air. Based on the segmentation the functional images were spatially normalized. Lastly, the images were smoothed. Here the activity of each voxel is calculated as a mean of its surrounding weighted by a Gaussian kernel.

Statistical analysis:

Our model consisted of four conditions, which coded for the onsets of stimuli and duration (1 sec) for each condition. Firstly, a contrast between face and house conditions was made to identify the Fusiform Face Area (FFA) using a familywise-error-corrected p-value of 0.05. This successfully localized a cluster of active voxels where we would expect FFA to be. In this cluster the voxel with maximal difference between face and house conditions made our designated voxel for our analysis (the location of the voxel is at the center of the blue cross in Figure 6)

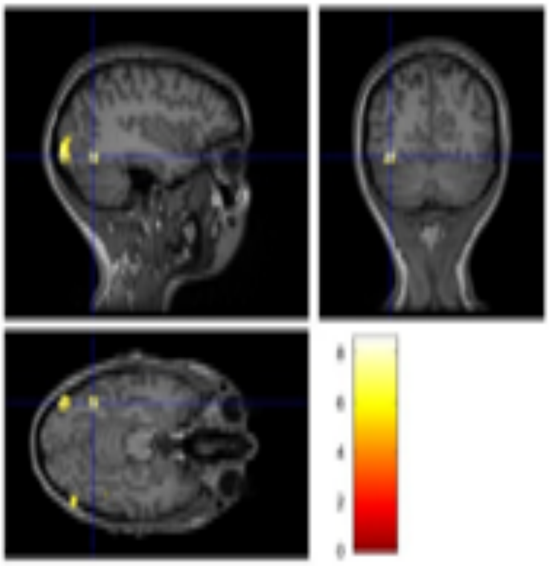

Figure 6: Designated voxel in FFA for analysis located at the center of the blue cross.

Secondly, the mean effect of each condition in the designated voxel was found using an F-test in SPM. Lastly, an overall contrast between expected an unexpected faces was made in the designed voxel using an uncorrected p-value of 0.05.

Target trials were coded as events of no interest in the fMRI analysis.

Results

Behavioural analysis

The behavioural analysis revealed no effect of the type of frame or condition on reaction times, F=0.71(3, 281), p=.55, indicating that manipulation of perceptual expectations did not affect the task performance of the participant.

fMRI anlalysis

Results revealed an effect of the house vs. face stimuli in the FFA, in which face images produced significantly greater activation than house stimuli, with a familywise-error-corrected p-value of 0.05.
A single voxel with the greatest difference between house and face conditions were chosen to be the designated voxel for the rest of the analysis.

Figure 7 shows the mean activity for each condition in the designated voxel. Bar 1 to 4 belong to the participant's first trial, and bar 5 to 8 are results of each condition for the second trail.

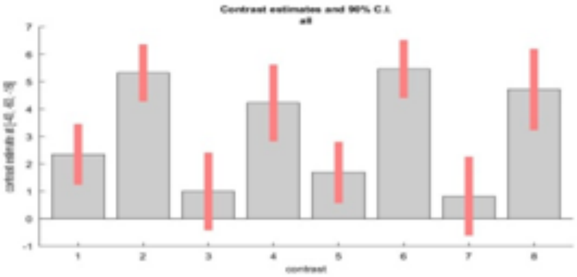

Figure 7: analysis of 4 x 2 conditions. Bar 1 to 4 represents the four conditions for the first trial in the following order: blueframe + house, greenframe + house, greenframe + house and blueframe + face. Bar 5 to 8 show same conditions for the second trial.

The general theme in the result reflects high activation whenever a face stimulus is presented (bar 2, 4, 6 and 8). Conditions showing houses produced significantly less activation (bar 1, 3, 5 and 7). Within stimuli conditions we don't see a significant difference depending on expectation.

Discussion

In the current experimental design, a feature detection theory would expect activation in the FFA only doing trials where a face image is presented. Thus, it would predict bar 2, 4, 6 and 8 in the result graph to be significantly higher than the remaining bars. Predictive coding theories would expect FFA activity in trials with either high expectation of a face or a surprise of a face, regardless of the actual stimulus presented. Thus, it would predict bar 2 and 6 (high face expectation + face stimulus), and bar 3 and 7 (high face expectation + house stimulus) to be high, because of a high face expectation, whereas it expects bar 4 and 8 (low face expectation + face stimulus) to be high because of a face surprise. Bar 1 and 5 is not expected to elicit significant FFA activation.

The FFA activation found in this study seemed related only to face stimuli, as the probabilistic frames driving the expectations of either a house or a face stimulus had no effect on the FFA responses. Thus, only face features elicited an FFA response, supporting a feature detection theory of a bottom-up driven visual processing.

However, this could be a result of an inadequate experimental design in relation to establishing a conditioned response following the conditioned stimuli. Due to the short length of the experiment, associations between the frames and the following stimulus, might not have had an adequate foundation to develop properly, such that frames were irrelevant to the participant's expectations of the following stimulus. In this case, the effect of expectations would be impossible to measure using this experimental design. The similar study done by Egner et al. (2010) consisted of 600 trials, which might have been a more suitable duration, in providing the foundation of establishing an association between the frames and the following stimulus, and could explain part of the difference in results found between their experiment and the current.

The behavioural analysis suggests that the perceptual expectations did not effect reactions times, as expectations are typically thought to mediate attention, resulting in higher reaction time (Posner et al., 1980). Further, the task of the participant was orthogonal across conditions, meaning that even if the participant did use any attentional strategies in her performance, it would have no influence on the BOLD signal, as no difference between conditions was found. Thus, it is very unlikely that the results are a function of attentional effects.

To further study the two theories and their adequacy in explaining the selective activity in the FFA, it might be relevant to investigate the timescale of the FFA responses. Taking a feature-detection approach, one would expect the activity in this experiment to occur only *after* presentation of the face stimulus, whereas holding a predictive coding approach, we would expect activation to occur *before* the presentation of a face. If the activity is related to the expectation of seeing a face, presenting a green frame should be enough to elicit the FFA response. A method of higher temporal resolution, could thus contribute further to the understanding of the mechanisms underlying the FFA activity.

The current discussion between feature-detection and predictive coding theories, resembles the previous shift seen in the understanding of the dopamine system. In 1998, Schultz found that the activity of dopaminergic neurons, which had previously been thought of as reward responses, did not depend on the presence or absence of a reward, but was elicited whenever there was a difference between the predicted reward and the actual reward (Schultz, 1998). These results led Schultz to characterize dopamine activity as a prediction error signal. The similarity between this shift and the discussion of feature detection and predictive coding as explanations of FFA activity, suggest that predictive coding may be relevant in understanding many regions and functions of the brain (Rao and Ballard, 1999). Additionally, predictive coding provides a useful framework for understanding several context-dependent phenomena such as repetition suppression, in which the evoked response of a stimulus is reduced by repetition of the stimulus (Summerfield et al., 2008). Predictive coding mechanisms are highly context-sensitive, and the evoked response is expected to change according to the context, e.g. repetition, in which the predictability is increased, and prediction error is reduced, thus reducing the evoked response signal (Clark, 2013).

Empirical evidence of predictive coding is still minimal, as well as no study has yet attempted to investigate the theory at the neural level. This study indicated feature detection in FFA. However, predictive coding provides explanations of several phenomena, which feature detection cannot, and the consequences of a shift from feature detection to predictive coding theories, is worth a thought. When seeking to identify and explain the patterns of neural activity it would become essential to strictly control for ones expectations and beliefs (Clark, 2013). Additionally, Clark argues, it means that the same computational resources are constructing believing and perceiving. Thus, predictive coding mechanisms, if true, would make it highly difficult to distinguish between perception and cognition.

References

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences,* 1-73.

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994) Prosopagnosia can be associated with damage confined to the right hemisphere – An MRI and PET study and a review of the literature. *Neurophyschologia, 32,* 893-902.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *The Journal of Neuroscience, 2010, 30(49),* 16601-16608.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience, 3,* 191-197.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G., R. (2014). Cognitive Neuroscience. The biology of the Mind. 4e.

Kanwisher, N., McDermott, J., & Chun M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience 1 June 1997,* 17 (11) 4302-4311.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision research 46, 28*, 4646-4674.

Posner M, I., Snyder, R, R., & Davidson, B, J. (1980) Attention and the detection of signals. *J Exp Psychol 109,* 160-174.

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,* 79-87.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80(1),* 1-27.

Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision Res 48,* 1391-1408.

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.