Investigating Fusiform Face(expectation?) Area

Adam Finnemann

Cognitive Science, University of Aarhus

Jens Chr. Sous Vej 2, 8000 Aarhus, Denmark

Abstract:

A recent study by Egner et al. (2010) has found that activity in the Fusiform Face Area (FFA) is modulated by expectations of faces. This has been used as one of many arguments in favor of predictive coding which takes the idea of an expectation driven brain to its full extent (Clark, 2010).This marks a break with the common textbook view where FFA is regarded as a feature detector specialized in face recognition. In this paper we replicate the fMRI-study by Egner and colleagues to test the different empirical predictions made by predictive coding and feature detection about FFA. In contrast to Egner et al. we find that face expectation didn't modulate activity in FFA. However, a critical discussion of the limitations of our experiment will show that no strong arguments should be drawn on the basis of our study.

Keywords*: Fusiform Face Area, FFA, Predictive Coding, feature detection, fMRI.*

1. Introduction

A standard introduction to the function of the Fusiform Face Area (FFA) is to say that it "recognizes faces". On this view FFA is seen as a feature detection (FD) device that responds selectively to stimuli containing faces[1]. When face recognition is seen as a FD process it becomes primarily a bottom up process where the brain is seen as a passive filter for stimuli to pass through. An opposite view on perception is found in *predictive coding* (PC) which claims that the overarching function of the brain is to make constant and active top-down predictions of sensory input (Clark, 2013).

PC and FD make different predictions about what should drive neural activity. According to FD neural activity should be modulated by features out in world. In contrast PC implies that expectations about the world and error-signals are driving neural activity.

[1] Other research argues that FFA is an area for fine-tuned feature detection since ornithologists and car enthusiast uses it for birds and cars (Anderson, p. 42, 2015). Since this is still within FD this detail doesn't make a difference to our study.

The PC model of the mind proposes a duplex hierarchical architecture of perception (Clark, p. 7, 2013). Where each level in the hierarchy of perceptual processing consists of two types of neural units: the first type encodes representations of causes of stimuli based on predictions received from higher states through backward connections. The second type of units encode error which is the information not explained by predictions. Error-signals are sought to be explained away by passing them on to higher level predictions through feed-forward loops (Clark, p. 7, 2013). Figure 1 visualizes this model. However, it doesn't fully reflect Clark's account where a "predictive estimator" consists of two distinct but laterally interacting representation and error units.

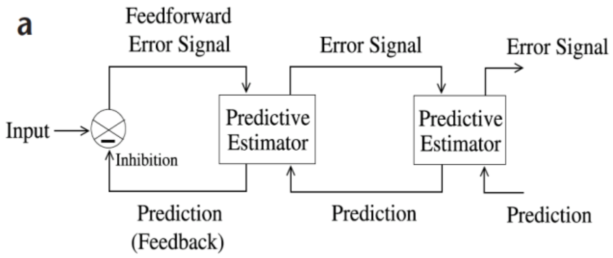

Figure 1: our cognitive architecture according to PC (Rao and Ballard, 1999, p. 80)

If we think of FFA as a predictive estimator we can see that its activity should be governed by error signals and predictions. Egner et al. (2010) have proposed a model relating errors and predictions to neural activity in FFA. Figure 2 shows a visualization of this model.

Figure 2: results of PC model (from Egner et al. (2010) with slight modification)

As the figure shows activity increases linearly with expectations of faces and face surprise. In this model an error signal is assumed to be a surprising/unexpected face (as shown in figure 1.b.). It's worth noting that they assume that the FFA doesn't respond to a surprising absence of faces (Egner et al., p. 11604, 2010).

Figure 1.c shows the sum of the activity in FFA. We see that PC predicts an interacting effect between expectation and stimuli: *going from low to high face expectation should only make a difference in the house stimuli condition*.

In contrast, FD predicts that *expectation shouldn't have an influence on activity*. So no interaction effect is expected here as the activity in FFA only should vary as a function of stimuli. Figure 3 visualizes Egner et al.'s proposed FD model:

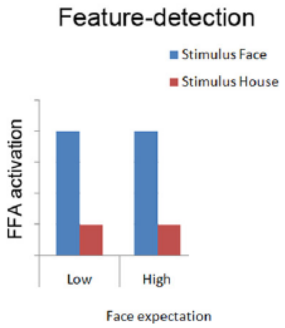

Figure 3: results of FD model (from Egner et al. (2010) with slight modification)

In this project we test the different predictions by replicating an experiment done by Egner et al. (2010), with the difference that they used three levels of face expectation whereas we only have two (figure 2 and 3 as been modified to reflect this difference).

In this experiment we end up with a two-by-two factorial design consisting of high and low face expectation together with face and house stimuli. We control expectations by making high face expectation the conditioned response upon green frames, and low face expectation conditioned on blue frames

2. Methods:

2.1 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 Aarhus University.

2.2 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 photos had white or gray backgrounds. All faces had a minimal facial expression 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.

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

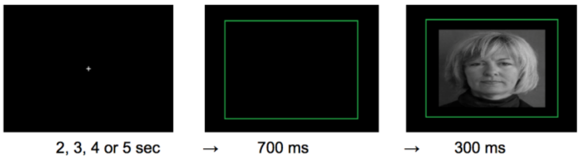

Figure 4: Example of stimuli

After each trial the participant was asked to answer whether stimuli were 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 with middle-finger. 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 color frame with stimuli type. 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 (see figure 5).

Stimuli / Face expectation	**House**	**Face**
High Face expectation	25% 	75% 
Low Face expectation	75% 	25% 

Figure 5: 2x 2 factorial designs. Four conditions: green-house, green-face, blue-house, blue-face.

The participant was instructed of the probabilistic pairing of color frame 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 behavioral data was collected. The behavioral data consisted of participant information (Subject ID, gender and age) as well as reaction time and correct responses.

3. Analysis:

3.1. Behavioral 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 (results), 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.

3.2. 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. The four conditions were a result of our two-by-two factorial design (house/face and green/blue frame).

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 were made our designated voxel for our analysis (the location of the voxel is at the center of the blue cross in the bottom of Figure 6)

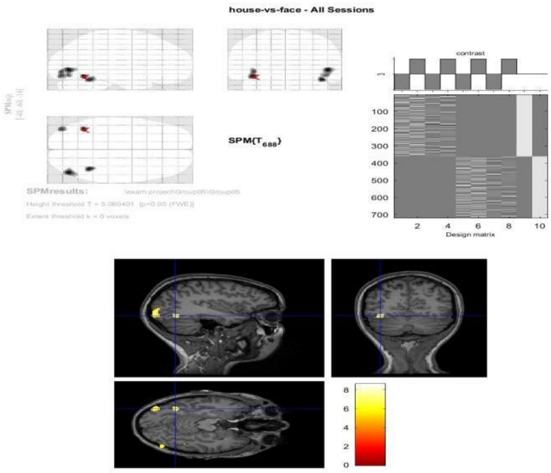

Figure 6: Designated voxel in FFA for analysis located at the center of the blue cross.

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.

4. Results:

- *4.1. Behavioral analysis*

The behavioral 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.

- *4.2. fMRI analysis*

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. The bar diagrams in figure 7 show the mean activity for each condition in the designated voxel. Where 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.

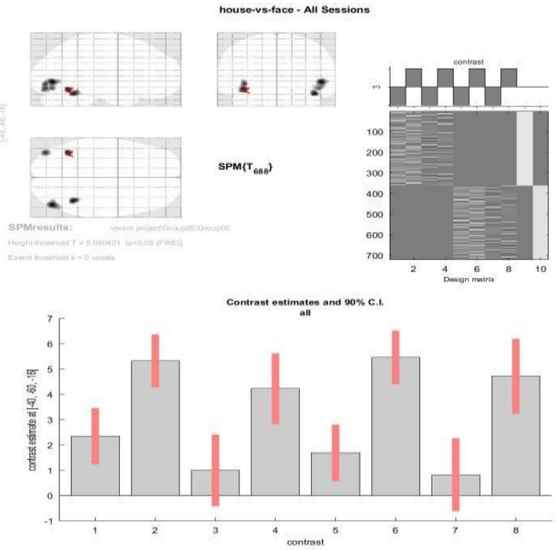

Figure 7: analysis of 4 x 2 conditions. Bar 1 to 4 represents the four conditions for the first trial in the following order: blueframe + house, greenframe + house, greenframe + house and blueframe + face.
Bar 5 to 8 show same conditions for the second trial.

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

5. Discussion

The two theories which are under investigation make radically different predictions about the high expectation and house condition (corresponding to bar 3 and 7). According to PC this condition should cause a level of activity similar to the activity seen in bar 2, 4, 6 and 8. In contrast to this FD predicts activity similar to bar 1 and 5.

Figure 7 shows clearly that bar 3 and 7 are more similar to the activity in the low expectation and house condition (bar 1 and 5) which is exactly what FD predicts.

In other words high expectations of faces didn't produce high neural activity. Thus this finding is supports the conclusion that FFA functions as a feature detector of faces and not as a face predictor.

Our subsequent analysis strengthens this conclusion. In the contrast between low and high expectation condition we didn't find significant unique activity only present in the high face expectation conditions in our designated voxel in FFA.

This finding is at odds with the original findings of Egner et al. who found high activation in all high face expectations conditions (Egner et al., 2010). That we haven't been able to replicate their findings presents prima facie a scientific problem. However before concluding this, our results should first undergo a critical discussion.

The best explanation of the difference between our and the original finding is that we only ran our experiment on a single participant. Thus we have no means to average random idiosyncratic effects away and our result therefore doesn't have any statistical power. The seriousness of this problem won't be reflected in the amount of space I will devote to it since it has a rather trivial solution: more participants. Instead I will focus my discussion on more challenging problems.

In our interpretations and conclusions we assume that the conditioning of high face expectation upon green frames actually took place. We tried to ensure this by verbally instructing the participant about the probabilistic relations between frames and stimuli, furthermore these relations should be implicitly learned during the experiment. It's possible that the association between green frames and high face expectation didn't take place.
A difference to the original experiment is that our participant only underwent 143 trials in comparison to the original experiment where 600 trials were used. So one possible problem is that 143 trials aren't enough to create a strong enough association between frames and stimuli in order for strong predictions to emerge in FFA. If this learning didn't take place it can be hypothesized that our activity only reflects error-signals of not yet predictable faces. A counter argument to this "not-enough-learning- problem" is that we don't see any learning at all across trials. If the function of FFA depended on implicitly learned expectation (as PC predicts) we would expect that some learning would have taken place in the second trial. We could therefore expect bar 6 and 7 to bigger than bar 3 and 4 relatively to the other bars in their respective trails. It should be noted that comparing across trials isn't straight forward since changes in both the fMRI scanner and participant take place over time. Nevertheless the activity across the two trials is *very* similar which indicate that learning don't modulate activity in FFA.

A possible confounding factor lies in attention. Since expectations modulate attention a competing explanation of our finding is that our results are formed by attentional processing (Posner et al., 1980). We have two arguments to support that this isn't the case. The task used was orthogonal to the four conditions thus attentional strategies would have an equal effect on each condition. Furthermore the behavioral analysis also showed no significant difference in RTs across conditions. According to Posner et al. (1980) increased attentional processing should be reflected in increased reaction times.

To sum up, since we only have 1 participant it's unwise to draw conclusions with regard to PC versus FD as such. Instead we can say that if we increased the number of participant and the result remained the same, then we would have an argument against the PC model of FFA.

As Andy Clark notes the project of exploring direct neural evidence for PC, as we have tried, is still in its infancy (Clark, p. 11, 2013). Further direct evidence in support of PC could be given by utilizing the different predictions made by PC and FD with respect to time of activity. A consequence of FD is that activity in FFA should occur only *after* face stimuli, whereas PC allows activity in FFA already at the onset of the initial frame reflecting predictions, while activity after face stimuli should vary as a function of error magnitude. Thus using the same paradigm as ours but using equipment with higher temporal resolution it would have been possible to add further support to either PC or FD.

References

Anderson, John R. *Cognitive Psychology And Its Implications*. 1st ed. New York: Worth publ., 2015. Print.

Egner, T., J. M. Monti, and C. Summerfield. "Expectation And Surprise Determine Neural Population Responses In The Ventral Visual Stream". *Journal of Neuroscience* 30.49 (2010): 16601-16608. Web.

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

Rao, Rajesh P. N., and Dana H. Ballard. *Nature Neuroscience* 2.1 (1999): 79-87. Web. 22 May 2017.

"Whatever Next? Predictive Brains, Situated Agents, And The Future Of Cognitive Science". *Behavioral and Brain Sciences* 36.03 (2013): 181-204. Web.