

# Object-level visual information gets through the bottleneck of crowding.

Jason Fischer & David Whitney

Department of Psychology, University of California, Berkeley, USA

*Corresponding Author:*

Jason Fischer

Ph: +1 530 574 6674

Email: jtf@berkeley.edu

Manuscript pages: 25

Figures: 6

Keywords: face recognition, object recognition, emotion, sparse selection, ensemble, texture

## **Abstract:**

Natural visual scenes are cluttered. In such scenes, many objects in the periphery can be crowded – blocked from identification – simply because of the dense array of clutter. Outside the fovea, crowding constitutes the fundamental limitation on object recognition, and is thought to arise from the limited resolution of the neural mechanisms that select and bind visual features into coherent objects. Thus, it is widely believed that in the visual processing stream, a crowded object is reduced to a collection of dismantled features with no surviving holistic properties. Here, we show that this is not so: an entire face can survive crowding and contribute its holistic attributes to the perceived average of the set, despite being blocked from recognition. Our results show that crowding does not dismantle high-level object representations to their component features.

## **Introduction:**

Much of the visual processing in the brain culminates in object recognition, whereby the vast array of visual features in a scene are segmented and combined into discrete objects that we can interact with. This process happens rapidly and effortlessly, so that our everyday experience is one of a seamless visual environment richly populated with objects. Yet at a given moment, many – perhaps most – of the objects in our peripheral vision are rendered unrecognizable by surrounding visual clutter. This phenomenon, known as visual crowding, constitutes the fundamental limitation on object recognition (Levi 2008; Pelli and Tillman 2008; Whitney and Levi).

Crowding is not a problem of insufficient visual acuity. An object in the periphery that is readily identifiable on its own can become unrecognizable simply by the addition

of one or two neighboring objects (Bouma 1970; Korte 1923). Thus, crowding reflects a limitation on the mechanisms by which visual features are combined into objects and brought to perceptual awareness. Understanding why crowding occurs is a crucial step in understanding visual object recognition.

Of the many theories that have been advanced to explain visual crowding, most have in common the underlying notion of a resolution limit in the object processing stream. That is, they posit that at some stage of object processing, the visual system lacks the necessary resolution to process an object individually when it is surrounded too closely by other objects. Theories have variously placed the resolution bottleneck at the level of elementary feature detection (Flom et al. 1963; Levi and Waugh 1994), feature pooling (Pelli et al. 2004; Pelli and Tillman 2008), attentional selection (He et al. 1996; Intriligator and Cavanagh 2001), and others, but all share the premise that beyond this resolution bottleneck, information about an individual crowded object is lost. The degraded representation of the crowded object leads observers to experience a jumbled, ambiguous percept in its place.

In the present study, we challenged the long-standing belief that crowding destroys object-level information. We presented observers with groups of faces and tested whether the central face in a group, despite being crowded by the surrounding faces, nonetheless exerted an influence on the perceived ensemble information about the set. We asked subjects to report the average expression in a group of faces when the central face was crowded such that it could not be individually identified (Fig. 1c). Despite the fact that subjects were unable to recognize the expression of the central crowded face, they were paradoxically able to incorporate its precise expression into an estimate of the

overall group expression. Our results suggest a different account of crowding than that posited by traditional theories: crowding is not a failure of object processing, but a failure of how individual objects are *perceptually accessed* from the object processing stream.

## **Methods:**

### *Experimental Design*

The experimental stimuli consisted of fifty faces falling on an emotional continuum from neutral to disgusted (Fig. 1a). We generated this emotional continuum by creating a fifty frame morph between a neutral expression and a disgusted expression for the same person, and we defined the difference between two consecutive frames as one emotional unit. We also added noise to the images, as depicted in Figure 1b, in order to control for luminance differences between the faces. To generate the noise in the face images, we first created a luminance histogram (with 256 bins corresponding to the possible grayscale values) for each of the fifty faces in the emotional morph. We averaged the histograms of all fifty images to produce a group histogram (Fig. 1b), and then added noise to each individual image until its histogram matched the group histogram. For example, if image X had fewer pixels of value 133 than the group histogram, then a pixel of value 133 was added to a random location in the image. By randomly moving through all possible grayscale values and iterating this process, we produced fifty images that all contained the same pixels, only arranged differently. On the right in Figure 1b is an example face after the addition of noise to equate its luminance profile with that of the group average. The histogram-based noise approach ensured that the mean luminance of all images was the same, and the mean luminance

within local regions of the images was very similar. Regions of an image that were slightly darker or lighter than in most other images, such as the bridge of the nose in the pictured face, received an appropriate level of noise to bring them in line with all images. We confirmed psychophysically that the discriminability between two adjacent frames did not vary significantly across the emotional continuum.

We first established a performance baseline for each subject's ability to discriminate the emotional expressions in the periphery by presenting two faces, one to the left and one to the right of fixation, at 16.5 deg. eccentricity (the two central face positions in Fig. 1c, without flanking faces). Although difficult, observers can still readily recognize faces at this eccentricity (Louie et al. 2007; McKone 2004). The faces were presented on a gray background, and each subtended 1.8 x 2.6 deg. of visual angle. The faces appeared for 1 s, followed by a gray background and fixation point awaiting the subject's response. We varied the emotional separation between the left and right faces between 2 and 12 emotional units, and subjects reported whether the left or right face was more disgusted in a two alternative forced choice (2AFC) method of constant stimuli task. For each subject,  $THR_{75}$  was defined as the smallest emotional separation between the faces that could be discriminated 75% of the time.

In the main experiment, we presented two sets of seven faces, with central faces positioned at 16.5 deg. to the left and right of fixation, and six flanking faces surrounding each central face (Fig. 1c). The emotional separation between the two central faces of the sets was fixed at either  $THR_{75}$  (established in the baseline experiment) or zero. The difference between the average expressions of the surrounding faces on the left and right ( $mean_{right} - mean_{left}$ ) varied between -8 and 8 emotional units, and was independent of the

relationship between the two central faces. We computed the average expression of each flanker set as the arithmetic mean of the emotional morph numbers of the six flanking faces in the set (i.e., the difference in flanker means plotted on the x axis in Figure 3 does not include the target faces). In separate runs, subjects either reported which central face was more disgusted, or reported which group of faces as a whole was more disgusted. These two run types were interleaved. Subjects completed at least four runs of each type, with each run comprising 255 trials.

In two control experiments, we repeated the main experiment with the central faces inverted (Fig. 3b) and scrambled (Fig. 3c), while the flanking faces remained upright and intact. To create the scrambled faces, we defined three sections of the face, containing the eyes, nose, and mouth, respectively, and swapped the arrangement of these sections. We also horizontally offset the three segments slightly to disrupt the alignment of the features. This approach left all of the individual facial features intact and upright, while aiming to disrupt the configural arrangement of the face by repositioning the features relative to each other.

In an additional dual-task control experiment, subjects made two judgments on each trial: 1) *Which set, on average, was more disgusted?* and 2) *Were the central faces of the sets upright or inverted?* The central faces were either upright or inverted on randomly interleaved trials. The left and right flanking sets were fixed to have the same average expression on each trial, while the difference in the expressions of the central faces was  $\pm \text{THR}_{75}$ . We selected only the trials in which the subjects misreported the orientation of the central faces for further analysis. Three subjects participated in the dual-task control, two of whom were naïve to the purpose of the experiment.

### *Analysis*

All data analysis was conducted in Matlab R2009a (The Mathworks, Natick, MA). Our primary interest in the main experiment was whether the emotions of the central (crowded) faces influenced the perceived expressions of the two groups in which they fell. To test for such an effect, we first fit psychometric functions to each subject's data. We separately fit one curve to the trials in which the right central face was more disgusted (open circles in Fig. 3) and another curve to the trials in which the left central face was more disgusted (filled circles in Fig. 3). As a reference, we also fit a third curve to the trials in which the right and left central faces had the same emotion (squares in Fig. 3). We used a standard logistic equation of the form  $y = \frac{1}{1 + e^{-a(x-b)}}$  to fit each curve. An influence of the central faces on the perceived average expressions of the sets would manifest as a horizontal displacement of the black and gray curves away from each other (i.e., significantly different  $b$  parameters from the curve fits). To test for a significant displacement between the curves, we used a bootstrapping approach similar to the one outlined by Wichmann and Hill (Wichmann and Hill 2001). We generated a bootstrapped distribution of fits for each curve by resampling the data with replacement 1000 times and fitting a new curve on each iteration. Each bootstrapped sample contained the same number of trials as the original data, but some trials were repeated and some did not appear in a given sample, due to the replacement during sampling. The resampling was performed within each bin along the abscissa, so that each data point in the resampled data represented the same number of trials as in the original data (20 trials per data point). In this way, the resampling procedure provided an estimate of the variability of the

original data at each point along the abscissa, and the distribution of resulting curve fit parameters reflected the confidence intervals around the estimates of those parameters. To compare the black and gray curves, we found the difference between the  $b$  parameter estimates for the two curves on each iteration, and tested whether the resulting distribution of difference scores was significantly different from zero.

### *Eye tracking*

In a set of control runs, we tracked subjects' eye positions using a Tobii X120 eye tracker sampling at 120 Hz. Calibration was performed between each testing run to ensure the accuracy of gaze estimates over the course of the experiment. Gaze recordings were analyzed in Matlab R2009a: we measured the distribution of  $x$  (horizontal) gaze coordinates, and plotted horizontal gaze position as a function of time for visualization (Fig. 4d). The two subjects were experienced psychophysical observers, and their gaze never departed more than  $\pm 2.5$  deg. from fixation (stimuli were at 16.5 deg.)

### *Monte Carlo Simulation*

In a series of control analyses, we simulated subjects' performance in the scenario where the influence of the central faces was due to incomplete or failed crowding. We used the trials from the main experiment in which the central faces were identical (dashed curves in Figure 3), and modified a subset of the trials on each iteration of a Monte Carlo simulation. On each iteration, we randomly sampled a subset of the trials (varying in size from 1% of total trials up to the percentage of total trials in which crowding failed, estimated separately for each subject), and replaced the subject's responses in the



subsampled trials with responses corresponding to a comparison of the two central faces. The estimated proportion of trials on which crowding might have failed is given by  $(2 * (\% \text{ correct} - 50)) / 100$ , since chance is 50%. Subject NS was at chance at discriminating the crowded faces (49.4% correct), suggesting no failed crowding on the whole, but, at the other extreme, subject SM's performance was 63.3% correct, suggesting crowding may have failed on up to 26.6% of trials.

We separately simulated trials in which the left central face was more disgusted than the right, and vice versa, replacing subjects' responses with "left" responses or "right" responses respectively. These altered trials reflect the response that the subject would have made on a trial in which crowding broke down. Taking all of the data, including the altered trials, we fit a psychometric curve using a standard logistic model, as shown in Figure 5a, acquiring a slope (a) parameter and a threshold (b) parameter. These parameters characterize a possible psychometric curve that we might have measured if the subject was simply exploiting failed crowding to perform the task. We plotted the pair of parameters as a point in two-dimensional parameter space in Fig. 5b, with the simulation for the case where the right central face was more disgusted than the left plotted in dark grey, and the reverse case (left more disgusted than right) plotted in light grey. We repeated this subsampling and curve fitting process ~25,000 times for each subject, each time sampling and altering anywhere from 1% of the total trials up to the percentage of total trials in which crowding failed. In this way, for each subject we obtained the two clusters of curve estimates (Fig. 5b), representing the psychometric curves that might result if the subject's responses were based solely on individual access to the central faces on some trials. In a separate simulation, we held the number of altered

trials constant, fixed at the proportion of trials in which crowding might have failed, and varied the weighting on the central faces between 0% and 100%. A larger weighting on the central faces produced a larger effective difference between the average expressions of the two sets. Altered trials were assigned a “left” or “right” response based on the likelihood of such a response given the effective difference between the set means, as determined from the subject’s baseline curve.

Our goal was to test whether subjects’ actual empirical data from the main experiment differed significantly from what would be expected based on any possible degree of failed crowding alone. To do so, we plotted subjects’ psychometric curve fit parameters from the main experiment (the filled black points in Fig. 5b) on the same plots as the bootstrapped data, and, after fitting a two-dimensional Gaussian distribution to each cluster of bootstrapped points, evaluated how extreme the empirical data points were, relative to the bootstrapped distributions. We examined the standard deviation of the cluster of bootstrapped values along the line passing through its centroid and the empirical point (Fig. 5c), determining the portion of the area under the Gaussian curve that fell beyond the subject’s empirical data.

## **Results:**

The goal of the study was to test whether an intact upright face, though crowded such that its expression could not be individually identified, could nonetheless influence the perceived ensemble expression in a group of faces. Face processing relies on holistic mechanisms that are sensitive to the configuration of facial features (Maurer et al. 2002; Moscovitch et al. 1997; Yin 1969), making faces an ideal stimulus for investigating the

fate of crowded object-level information. The stimuli comprised 50 faces falling on an emotional continuum from neutral to disgusted (Fig. 1a). The difference between two adjacent frames in the continuum was defined as one emotional unit.

### *Single face discrimination baseline*

We first established each subject's baseline threshold for discriminating between the expressions of two faces presented without flankers (see Methods). Figure 2 shows the results from this baseline experiment. For each subject, the threshold emotional separation ( $THR_{75}$ ) was defined as the smallest emotional separation at which the subject correctly discriminated at least 75% of the face pairs. Each subject's individual  $THR_{75}$  was used in the main experiment as the emotional separation between the two crowded central faces.

### *Main Experiment: Discriminating the ensemble expressions of face sets*

In the main experiment, we presented two groups of seven faces each, again centered at 16.5 deg to the left and right of fixation (Fig. 1c; see Methods). In separate, interleaved runs, subjects either compared the expressions of the two central faces, reporting which was more disgusted, or compared the average emotions of the left and right sets as a whole (recent results show that observers can accurately and rapidly judge the average emotion of a set of faces (de Fockert and Wolfenstein 2009; Haberman and Whitney 2007; Sweeny et al. 2009)).

When discriminating the two central faces, each subject's performance dropped significantly as compared to his/her performance in the absence of flankers (depicted in

the shaded bars in Fig.2; all subjects  $p < 0.001$ ). This reduction in discrimination of the target faces with the addition of flankers is due to crowding, and has been demonstrated to result from interactions at the level of holistic face information (Farzin et al. 2009; Louie et al. 2007). Crowding resulted in a drop to chance-level performance in one subject (NS) and slightly above-chance performance in the other three subjects.

Figure 3a shows subjects' performance when viewing the same stimulus configuration, but comparing the average expressions of the two sets as wholes. On the x axis is the emotional separation between the right and left flanker means ( $\text{mean}_{\text{right}} - \text{mean}_{\text{left}}$ ), not including the central faces; positive values indicate that the right set was more disgusted on average, and negative values indicate that the left set was more disgusted. On the y axis is the proportion of trials in which the subject responded that the *right* group of faces was more disgusted. The dashed curve in each plot shows performance when the two central faces were identical, while the black and gray solid curves show performance for trials in which the right or left central face was more disgusted, respectively. The fact that these solid curves are shifted away from each other in every subject indicates an influence of the crowded central faces on subjects' perception of the average group expressions, despite the fact that the central faces were crowded. For example, at the zero point on the abscissa, when the left and right flanking sets had the same average expression, subjects were more likely to perceive the right set as more disgusted when the right central face was more disgusted (the solid black curve), and vice versa when the left central face was more disgusted (the solid gray curve). A bootstrapping analysis, as detailed in the Methods, established that these shifts were significant in the psychometric functions of every subject ( $p < 0.001$  for JF, NS, & SM;

$p = 0.002$  for TH).

### *Inverted and scrambled face controls*

Each subject participated in two additional control experiments to test whether the influence of the crowded central faces on the perceived expressions of the surrounding sets took place at the level of holistic, object-level information rather than at the level of basic visual features. We repeated the main experiment, this time inverting the central faces or scrambling the central faces as depicted at the top of Figure 3. Each is a common method for disrupting configural processing of objects while leaving low-level stimulus features intact (Moscovitch et al. 1997; Yin 1969). If the influence of the crowded central faces that we found in the main experiment was based on low-level stimulus properties such as luminance or orientation rather than face recognition, we would expect to find a similar effect after inverting or scrambling the central faces. In fact, Figure 3b shows that the influence of the central faces on the perceived expressions of the surrounding sets was eliminated when the central faces were inverted (JF:  $p = 0.73$ ; NS:  $p = 0.86$ ; SM:  $p = 0.32$ , TH:  $p = 0.41$ ). Likewise, disrupting the configural arrangement of the facial features by scrambling their positions eliminated the effect of the central faces (Fig. 3c; JF:  $p = 0.24$ ; NS:  $p = 0.66$ ; SM:  $p = 0.22$ , TH:  $p = 0.16$ ).

We performed an additional control in two subjects in which we repeated all three versions of the experiment (upright, inverted, and scrambled central faces), this time separately determining  $THR_{75}$  for inverted and scrambled faces such that they were presented at an emotional separation that was just as discriminable as upright faces when viewed in isolation. Data from the upright central faces replicated the results of the main

experiment, showing a significant influence on the perceived set expressions (Fig. 4a; JF:  $p = 0.001$ , JS:  $p = 0.002$ ). However, despite being readily discriminable when viewed in isolation, the inverted and scrambled central faces still failed to influence the perceived expressions of the surrounding sets (Fig. 4b-c; JF<sub>invert</sub>:  $p = 0.64$ , JF<sub>scramb</sub>:  $p = 0.84$ , JS<sub>invert</sub>:  $p = 0.29$ , JS<sub>scramb</sub>:  $p = 0.15$ ). Thus, the influence of the crowded central faces hinges on them containing upright features that are arranged appropriately, such that configural or holistic face processing is possible. We conducted an additional control experiment in one subject in which both the flankers and central faces were inverted, and again found no influence of the inverted central faces on the perceived set expressions ( $p = 0.22$ ).

During these control runs, we also tracked subjects' eye positions to ensure that eye movements were not responsible for the results. The subjects maintained good fixation throughout all runs (example eye traces are shown in Fig. 4d), and showed the same pattern of results as in the main experiment.

#### *Accounting for incomplete crowding in the main experiment*

A potential concern is that if crowding was incomplete on some trials in the main experiment, individual access to the central faces on those trials might have influenced subjects' responses regarding the average set expressions. While Subject NS was at chance performance when discriminating the crowded central faces (Fig. 2), the other three subjects performed slightly but significantly above chance, indicating that they had individual access to the central faces on some trials. We tested whether these cases in which crowding was incomplete could have been sufficient to explain the influence of the central faces on the perceived set expression that we found.

We simulated the scenario in which subjects averaged the central faces with the surrounding flankers when crowding was incomplete, but did not incorporate the expressions of the central faces on the remaining trials (see Methods for details). We used subjects' performance on the discriminating the crowded central faces to estimate the maximum number of trials in which crowding might have broken down during the set mean discrimination runs. We found that the resulting shifts in subjects' psychometric functions would be much smaller than those that we actually measured in the main experiment (all  $p$  values  $\geq 0.34$ ); that is, crowding did not fail on enough trials to produce an effect of the size that we found. We extended this analysis to include the possibility that subjects weighted the central faces more heavily than the surrounding faces on trials where crowding failed. A Monte Carlo simulation depicted in Figure 5 and detailed in the Methods showed that the outcome of such a strategy is inconsistent with our results. If subjects based their judgments solely on the central faces whenever crowding failed, such a strategy could produce shifts in the psychometric functions of the size that we found in the main experiment, but this strategy would also yield substantially flattened psychometric functions as a result of ignoring the flanking faces on some trials (Fig 5b; recall that in Figure 3, the abscissa represents the relative means of the flanking faces, not including the central faces). All subjects' actual performance fell significantly outside the range of expected performance based on such a weighting strategy (all  $p$  values  $\leq 0.0037$ ). Similarly, over a range of possible weightings of the central faces from 0% to 100%, we found that no weighting would have produced psychometric functions that were simultaneously as steep and shifted as those we measured in the main experiment (actual psychometric functions fell outside the clouds of simulated psychometric

functions; all  $p$  values  $\leq 0.0022$ ).

We also considered the possibility of partial crowding – perhaps some features from the target faces consistently leaked through the influence of crowding, but not always to a degree that would allow for correct discrimination of the central faces. It has been shown that some performance variability in crowding tasks can be attributed to idiosyncratic but consistent differences in the difficulty of crowding with different target-distractor configurations (Dakin et al. 2009). If there was any consistency to the kinds of features that might leak through crowding, then the trials that subjects got correct most often in the crowded target discrimination task should have been more likely to show an effect of the central faces in the ensemble discrimination task. Because the stimuli were the same for the crowded face discrimination and ensemble discrimination tasks, we were able to test whether the stimuli that subjects tended to get correct when discriminating the central faces also tended to drive the effect of the central faces in the ensemble discrimination experiment. That is, did variability in the crowding effect correlate with variability in the ensemble discrimination performance, across the same stimuli? For each ensemble discrimination trial, we searched within the crowded face discrimination trials to find matching trials—those trials with the same flanker means (the emotional separation between the central faces was always the same, set at  $THR_{75}$  for each subject). For each trial in the ensemble discrimination task, we computed percent correct for the matching trials in the crowded face discrimination task, and also recorded whether the subject's response when comparing the ensemble expressions was consistent with the relative emotions of the central faces. We correlated these two measures with each other across trials to test whether a trial's difficulty in the target discrimination task was



predictive of whether the subject was likely to show a bias toward using the central face in the ensemble discrimination task. This correlation was not significant in any subject (all  $p$  values  $\geq 0.14$ ); crowding difficulty did not predict whether the central faces influenced the perceived set expressions on a trial-by-trial basis.

The results of the Monte Carlo simulations and partial crowding simulation suggest that partial crowding would not explain the results in Figures 3 and 4. Nevertheless, we sought a more direct test of this possibility by employing a dual task in which we measured crowding and ensemble perception within the same trials.

#### *Dual task experiment*

We conducted a dual-task control experiment to test for an influence of the crowded central faces on the perceived set expressions on a within-trial basis (see Methods). Subjects made two judgments on each trial: 1) *Which set, on average, was more disgusted?* and 2) *Were the central faces of the sets upright or inverted?* We analyzed only the trials in which the subjects misreported the orientation of the central faces for further analysis.

Note that the orientation task provided a stringent test of whether the central faces were successfully crowded. In a separate control experiment, we verified that failure at the upright/inverted task was a good indicator that the expressions of the central faces were crowded as well: when subjects judged both the orientation and the relative expressions of the central faces, performance on the expression discrimination was at chance within trials in which the upright/inverted judgment was incorrect (JF: 47.2% correct, GM: 47.1% correct, TS: 43.1% correct; none significantly different from

chance). That is, when subjects incorrectly identified target face orientation, they were unable to report the target expression above chance.

Figure 6 shows the results of the dual task experiment. On trials in which the subject incorrectly reported upright central faces as inverted, the central faces nonetheless had a significant influence on the perceived set expressions (Fig. 6a; GM:  $z = 2.04$ ,  $p = 0.042$ ; JF:  $z = 3.10$ ,  $p = 0.002$ ; TS:  $z = 3.30$ ,  $p = 0.002$ ). On the other hand, when inverted central faces were misreported as upright, they had no effect on the perceived average set expressions (Fig. 6b; all  $p$  values  $\geq 0.38$ ). Thus, even within the same trials in which subjects could not identify the orientation of the central faces, those faces nonetheless contributed holistic information to the perceived expression of the surrounding sets.

## **Discussion:**

Our results show that crowding does not irreversibly dismantle or destroy the “objectness” of an object. Despite the fact that subjects could not explicitly identify the central faces, the visual system had access to precise information about the expressions of the crowded faces when generating average representations of the sets. Crowding may be an essential bottleneck for object *recognition*, but it does not break down object processing itself.

The present results suggest that the purpose of object processing is not solely for the sake of perceiving individual objects. Indeed, an emerging literature on ensemble coding is revealing that the brain can efficiently compute summary information about groups of visual objects at all levels of complexity (Alvarez and Oliva 2009; Ariely 2001;

Chong and Treisman 2003; Haberman and Whitney 2007; Oliva 2005). These high-level textures may provide a rapid sense of the objects that surround us, constituting the initial perception of a scene before we selectively attend to or perceive the individual constituent objects (Hochstein and Ahissar 2002).

An example where such a tradeoff between the perception of individual objects and the perception of object textures may be particularly apparent is in the phenomenon of “mindsight” (Rensink 2004). Mindsight refers to the ability of observers to accurately detect a change in a scene, without being able to locate where the change occurred. Our results provide a reasonable mechanism for mindsight: if many of the particular objects in the scene are crowded from recognition, an observer may be unable to report which object among many crowded ones has changed. However, since those crowded objects nonetheless contribute to the scene gist, the presence of a change in an individual object could be detected simply by being aware that some ensemble information has changed.

Our study is similar in experimental design to an important study conducted by Parkes et al. (Parkes et al. 2001), but addresses a very different question. They presented arrays of oriented gratings in the periphery, and found that subjects perceived the average orientation of the arrays in a manner consistent with a compulsory averaging of the crowded orientation information, thus uncovering a link between crowding and texture processing. Greenwood et al. (Greenwood et al. 2009) reported similar findings with averaging of crowded positional information. Critically, both of these studies found averaging of *basic visual features* during crowding. While neither study’s results make a direct prediction about the fate of higher-level objects (e.g., faces) during crowding, based on the compulsory integration of basic visual features that occurs at a crowded

location, it is reasonable to think that crowding would render further object processing impossible. This is consistent with the commonly held view that crowding dismantles objects to their component features (Levi 2008), and that those features are then available as a texture (Parkes et al. 2001). Our results, based on the presentation of high-level stimuli rather than basic visual features, show that object-level information in fact still exists in neural representations of the crowded location.

Our results demonstrate that although crowding is the fundamental bottleneck on conscious object recognition (Levi 2008), high-level holistic or configural information about objects is not destroyed or broken down to low level features. The visual system maintains precise high-level object representations even in the crowd.

### **Acknowledgments:**

This work was supported by NEI grant T32EY015387 (J.F.), NIH grant EY018216 (D.W.), and NSF grant 0748689 (D.W.). We thank Jason Haberman for providing the morphed face images used in the main experiment.

### **References:**

- Alvarez GA, and Oliva A.** Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc Natl Acad Sci U S A* 106: 7345-7350, 2009.
- Ariely D.** Seeing sets: Representation by statistical properties. *Psychological Science* 12: 157-162, 2001.
- Bouma H.** Interaction effects in parafoveal letter recognition. *Nature* 226: 177-178, 1970.
- Chong SC, and Treisman A.** Representation of statistical properties. *Vision Res* 43: 393-404, 2003.
- Dakin SC, Bex PJ, Cass JR, and Watt RJ.** Dissociable effects of attention and crowding on orientation averaging. *J Vis* 9: 28 21-16, 2009.
- de Fockert J, and Wolfenstein C.** Rapid extraction of mean identity from sets of faces.

*Q J Exp Psychol (Colchester)* 62: 1716-1722, 2009.

**Farzin F, Rivera SM, and Whitney D.** Holistic crowding of Mooney faces. *Journal of Vision* 9: 1-15, 2009.

**Flom MC, Heath GG, and Takahashi E.** Contour Interaction and Visual Resolution: Contralateral Effects. *Science* 142: 979-980, 1963.

**Greenwood JA, Bex PJ, and Dakin SC.** Positional averaging explains crowding with letter-like stimuli. *Proc Natl Acad Sci U S A* 106: 13130-13135, 2009.

**Haberman J, and Whitney D.** Rapid extraction of mean emotion and gender from sets of faces. *Curr Biol* 17: R751-753, 2007.

**He S, Cavanagh P, and Intriligator J.** Attentional resolution and the locus of visual awareness. *Nature* 383: 334-337, 1996.

**Hochstein S, and Ahissar M.** View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36: 791-804, 2002.

**Intriligator J, and Cavanagh P.** The spatial resolution of visual attention. *Cogn Psychol* 43: 171-216, 2001.

**Korte W.** Über die Gestaltauffassung im indirekten Sehen. *Zeitschrift für Psychologie* 93: 17-82, 1923.

**Levi DM.** Crowding--an essential bottleneck for object recognition: a mini-review. *Vision Res* 48: 635-654, 2008.

**Levi DM, and Waugh SJ.** Spatial scale shifts in peripheral vernier acuity. *Vision Res* 34: 2215-2238, 1994.

**Louie EG, Bressler DW, and Whitney D.** Holistic crowding: selective interference between configural representations of faces in crowded scenes. *J Vis* 7: 24 21-11, 2007.

**Maurer D, Grand RL, and Mondloch CJ.** The many faces of configural processing. *Trends Cogn Sci* 6: 255-260, 2002.

**McKone E.** Isolating the special component of face recognition: peripheral identification and a Mooney face. *J Exp Psychol Learn Mem Cogn* 30: 181-197, 2004.

**Moscovitch M, Winocur G, and Behrmann M.** What Is Special about Face Recognition? Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition. *J Cogn Neurosci* 9: 555-604, 1997.

**Oliva A.** Gist of the scene. In: *the Encyclopedia of Neurobiology of Attention*, edited by Itti L, Rees G, and Tsotsos JK. San Diego, CA: Elsevier, 2005, p. 251-256.

**Parkes L, Lund J, Angelucci A, Solomon JA, and Morgan M.** Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci* 4: 739-744, 2001.

**Pelli DG, Palomares M, and Majaj NJ.** Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J Vis* 4: 1136-1169, 2004.

**Pelli DG, and Tillman KA.** The uncrowded window of object recognition. *Nat Neurosci* 11: 1129-1135, 2008.

**Rensink RA.** Visual sensing without seeing. *Psychol Sci* 15: 27-32, 2004.

**Sweeny TD, Grabowecky M, Paller K, and Suzuki S.** Within-hemifield perceptual averaging of facial expressions predicted by neural averaging. *Journal of Vision* 9: 1-11, 2009.

**Whitney D, and Levi DM.** Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn Sci* 15: 160-168.

**Wichmann FA, and Hill NJ.** The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys* 63: 1314-1329, 2001.

**Yin RK.** Looking at Upside-down Faces. *Journal of Experimental Psychology* 81: 141-&, 1969.

### **Figure Legends:**

Figure 1: Experimental Stimuli. **a)** Faces ranged from neutral (0) to disgusted (49) in a fifty frame morph; one emotional unit was defined as the difference between two consecutive frames. **b)** The average luminance profile across all fifty faces in the main stimulus set. Each face had a unique noise distribution added in order to equate its pixel contents to the group histogram (see Methods). The example image shows the noise added to frame 36 of the morph. **c)** In the main experiment, we presented two groups of faces centered at 16.5 deg. to the left and right of fixation (faces are shown here at a reduced eccentricity for visualization). In separate runs, subjects either judged which central face was more disgusted, or which set, on average, was more disgusted. In this example, the left and right flanking sets contain the same six faces (average expression of 35 emotional units), while the left central face is 12 units more disgusted than the right.

Figure 2: Single face emotional discrimination. For each subject, performance at discriminating two target faces (% correct) in the absence of flankers is plotted as a function of the emotional separation between the faces (emotional units). Error bars indicate  $\pm 68\%$  and  $\pm 95\%$  confidence intervals for logistic curves fit to the data. For each subject,  $THR_{75}$  was defined as the smallest emotional separation for which performance was at 75% correct or better;  $THR_{75}$  was used as the fixed separation between the target faces in the main experiment. When performance at the  $THR_{75}$  emotional separation was

retested with flankers present, performance was significantly impaired in every subject (shaded bars; error bars are  $\pm 1$  SEM, all  $p < 0.001$ ), indicating that the flankers were effective at crowding the central faces.

**Figure 3:** A crowded face influences the perceived set ensemble expression. **a)** Subjects judged which set was more disgusted; along the abscissa are the relative average expressions of the flankers of the two sets (excluding the central faces; positive values indicate the right set was more disgusted), and on the y axis is the proportion of trials on which the subject responded 'right'. For the data plotted in square symbols (fit with a dashed line), the left and right central faces were the same. For the open and filled points, the right target was more disgusted than the left, and vice versa, respectively. As shown by the displacement of the black and gray curves away from the dashed baseline curve, the relative expressions of the central faces substantially influenced the perceived average expressions of the sets ( $p < 0.001$  for JF, NS, & SM;  $p = 0.002$  for TH). **b)** After inverting the central faces, which disrupts configural face processing while leaving low-level image statistics intact, the central faces no longer influenced the perceived set ensembles. (JF:  $p = 0.73$ ; NS:  $p = 0.86$ ; SM:  $p = 0.32$ , TH:  $p = 0.41$ ). **c)** Likewise, scrambling the configuration of the facial features while leaving them upright eliminated the influence of the central faces on the perceived expressions of the surrounding sets (JF:  $p = 0.24$ ; NS:  $p = 0.66$ ; SM:  $p = 0.22$ , TH:  $p = 0.16$ ).

**Figure 4:** Inverted and scrambled controls with single face discrimination equated. In two subjects, we repeated the inverted and scrambled control experiments (as well as the main

experiment with upright central faces), this time separately computing  $THR_{75}$ , the threshold emotional separation, for the inverted and scrambled faces when presented in isolation. **a)** For upright central faces, the results from these subjects replicated those of the main experiment: the expressions of the crowded central faces exerted a significant influence on the perceived expressions of the surrounding sets, as indicated by a horizontal displacement between the black and gray solid curves (JF:  $p = 0.001$ ; JS:  $p = 0.002$ ). **b)** Even with the central inverted faces were separated on the emotional continuum (Fig. 1a) such that they were as discriminable as upright faces, they nonetheless failed to influence subjects' perception of the average set expressions (JF:  $p = 0.64$ ; JS:  $p = 0.29$ ). **c)** After equating the discriminability of isolated scrambled faces with that of upright faces, the scrambled faces still failed to influence the perceived expressions of the surrounding sets (JF:  $p = 0.84$ ; JS:  $p = 0.15$ ). **d)** We also tracked subjects' eye movements during these control runs. Example plots are shown for horizontal gaze position over the course of one run for each subject. The subjects were experienced psychophysical observers, and their gaze did not deviate more than  $\pm 2.5$  deg. from fixation at any time during the runs (the stimulus arrays were at 16.5 deg. eccentricity).

Figure 5: Simulated influence of incomplete crowding in the main experiment. On each iteration of a Monte Carlo simulation, we produced simulated data based on the hypothesis that subjects responded according to a comparison of the central faces when crowding was incomplete, and had no information about the central faces on the remaining trials (see Methods). We fit the data from each iteration with a standard



logistic function (**a**), and plotted the resulting parameter estimates in a 2-dimensional parameter space (**b**). We compared the parameter estimates from subjects' empirical data (solid black dots in panel **b**) with those from the simulated data by fitting a 2-dimensional Gaussian to the simulated parameter estimates (**c**). Subjects' empirical data differed significantly from the simulated distributions, indicating that failed crowding on a portion of trials cannot explain the effect we found in the main experiment.

Figure 6: Dual task experiment. Subjects viewed stimuli similar to those in the main experiment, this time judging both the relative average expressions of the two sets, and also whether the central faces were upright or inverted. We selected only the trials in which the subject responded incorrectly to the upright/inverted question for further analysis; on these trials, subjects could not individually access the central faces sufficiently to identify their orientations. The left and right flanking sets always had the same average expression, so any overall difference in the reported relative expressions of the sets was due to an influence of the central faces. **a**) When central faces were upright but incorrectly reported as inverted, they nonetheless had a significant influence on the perceived average expressions of the surrounding sets (GM:  $z = 2.04$ ,  $p = 0.042$ ; JF:  $z = 3.10$ ,  $p = 0.002$ ; TS:  $z = 3.30$ ,  $p = 0.002$ ). **b**) On the other hand, inverted faces that were misreported as upright had no effect on the perceived set expressions (all  $p$  values  $\geq 0.38$ ). Error bars are  $\pm 1$  SEM.

Figure 1

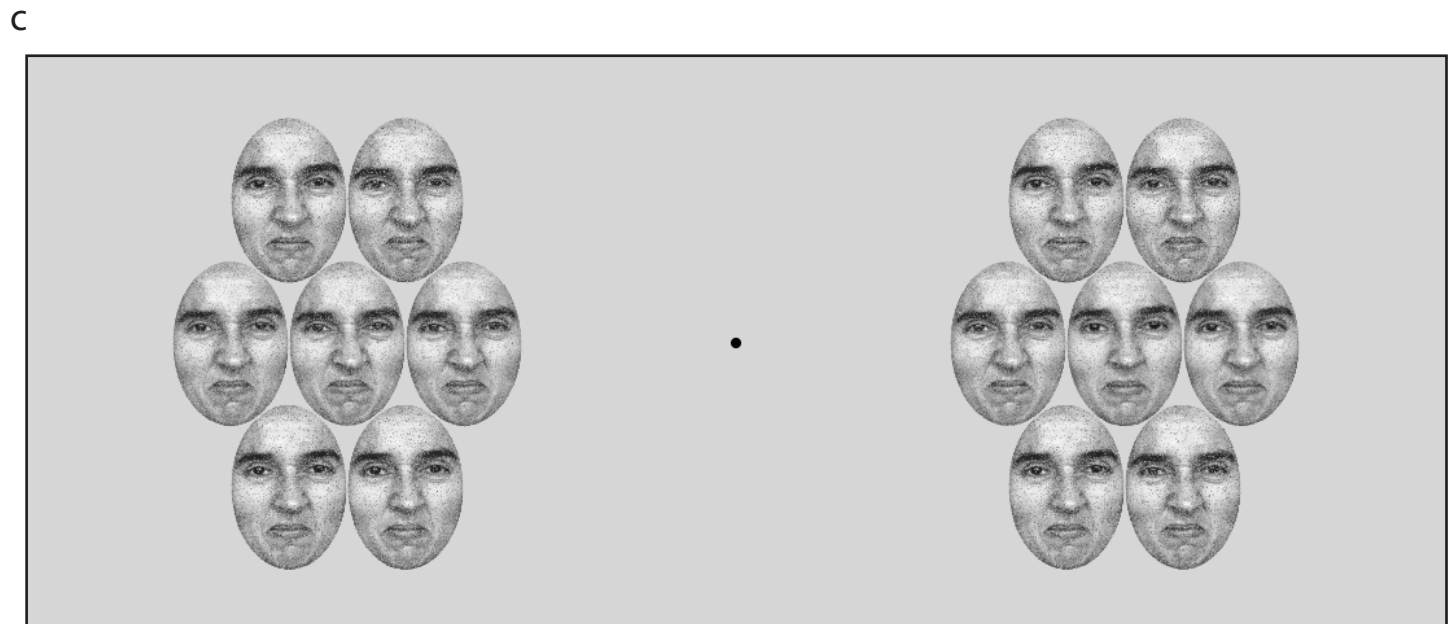
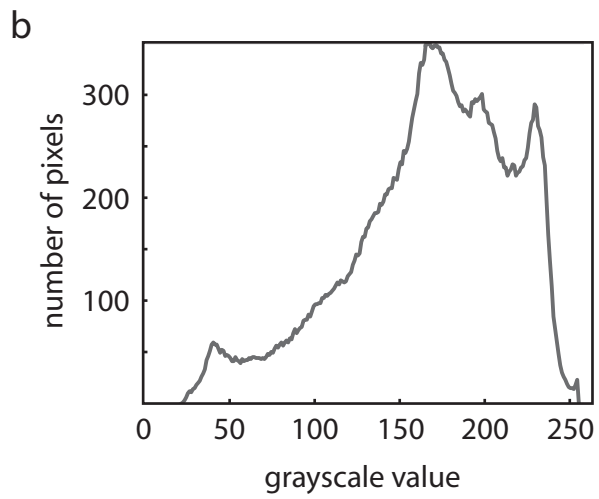
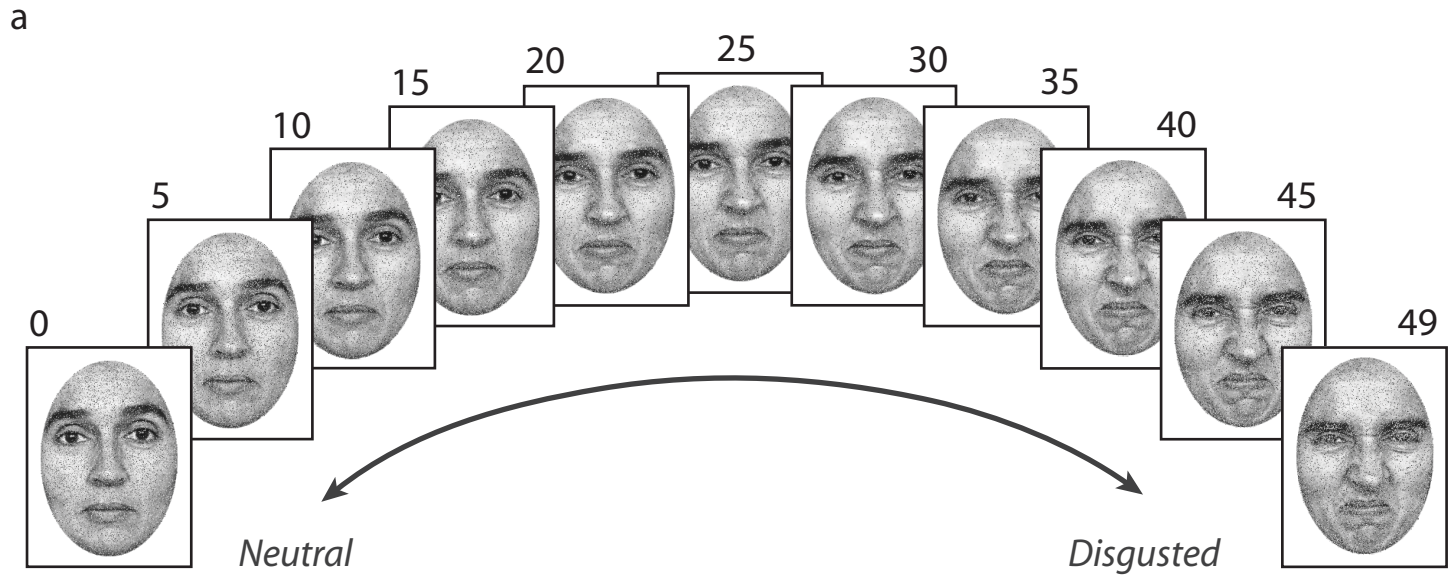


Figure 2

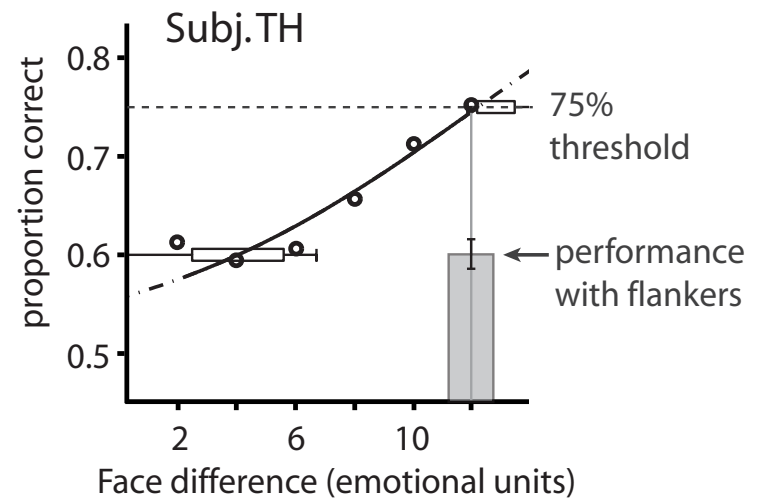
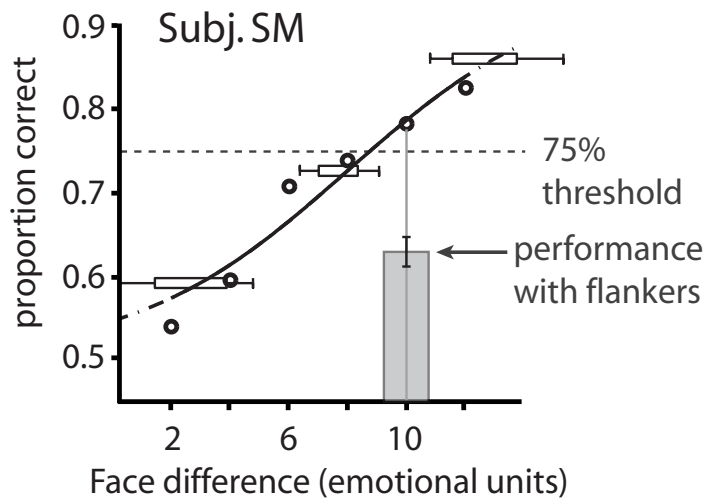
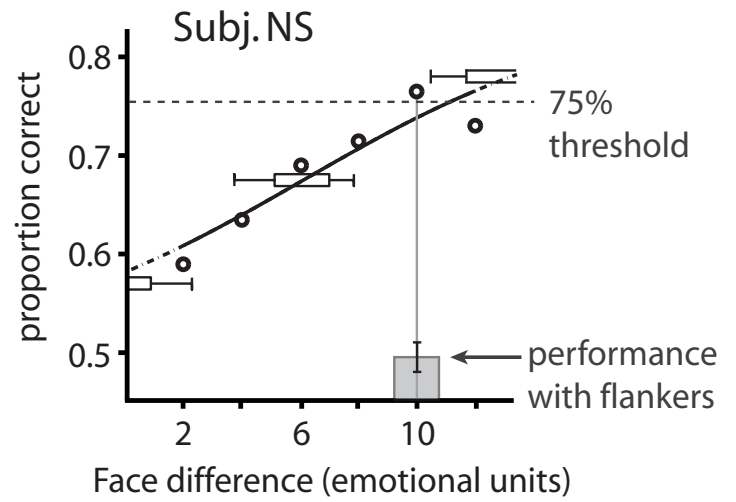
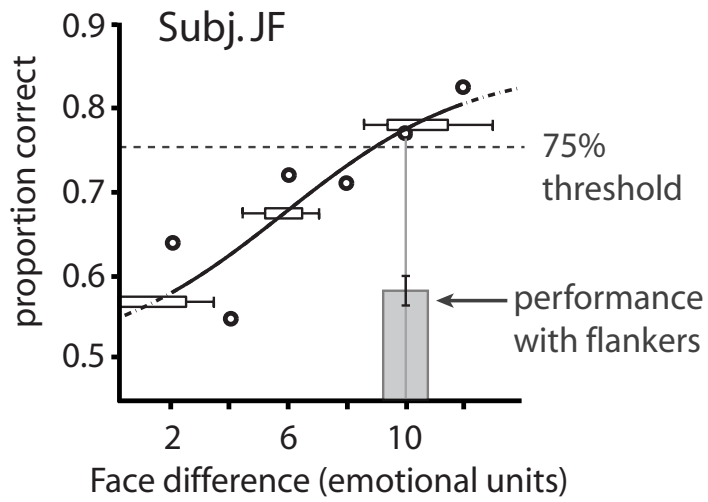


Figure 3

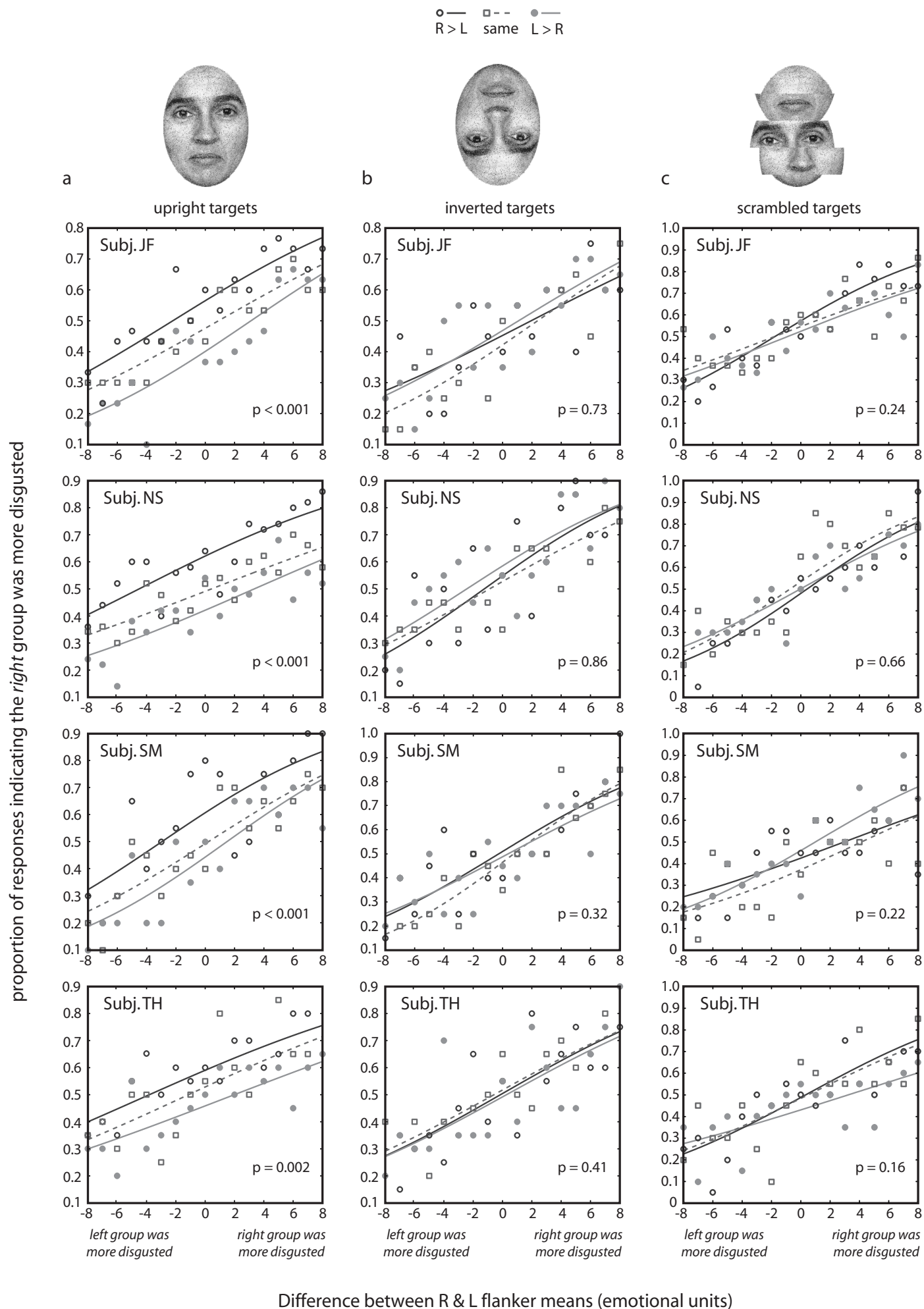


Figure 4

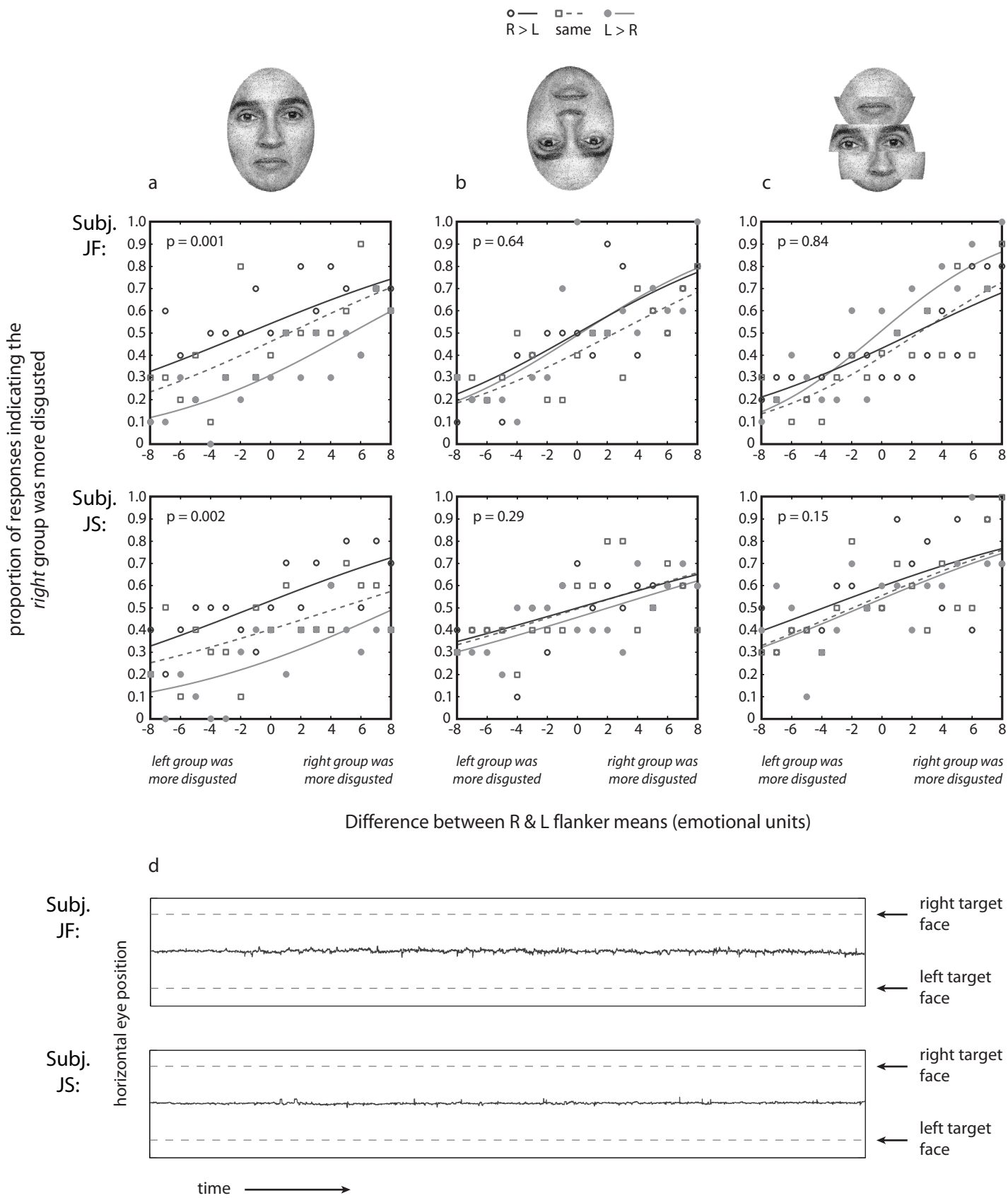
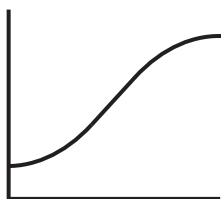


Figure 5

a



Logistic model:

$$y = \frac{1}{1 + e^{-a(x - b)}}$$

a: slope parameter

b: threshold parameter

b

left target > right target

right target > left target

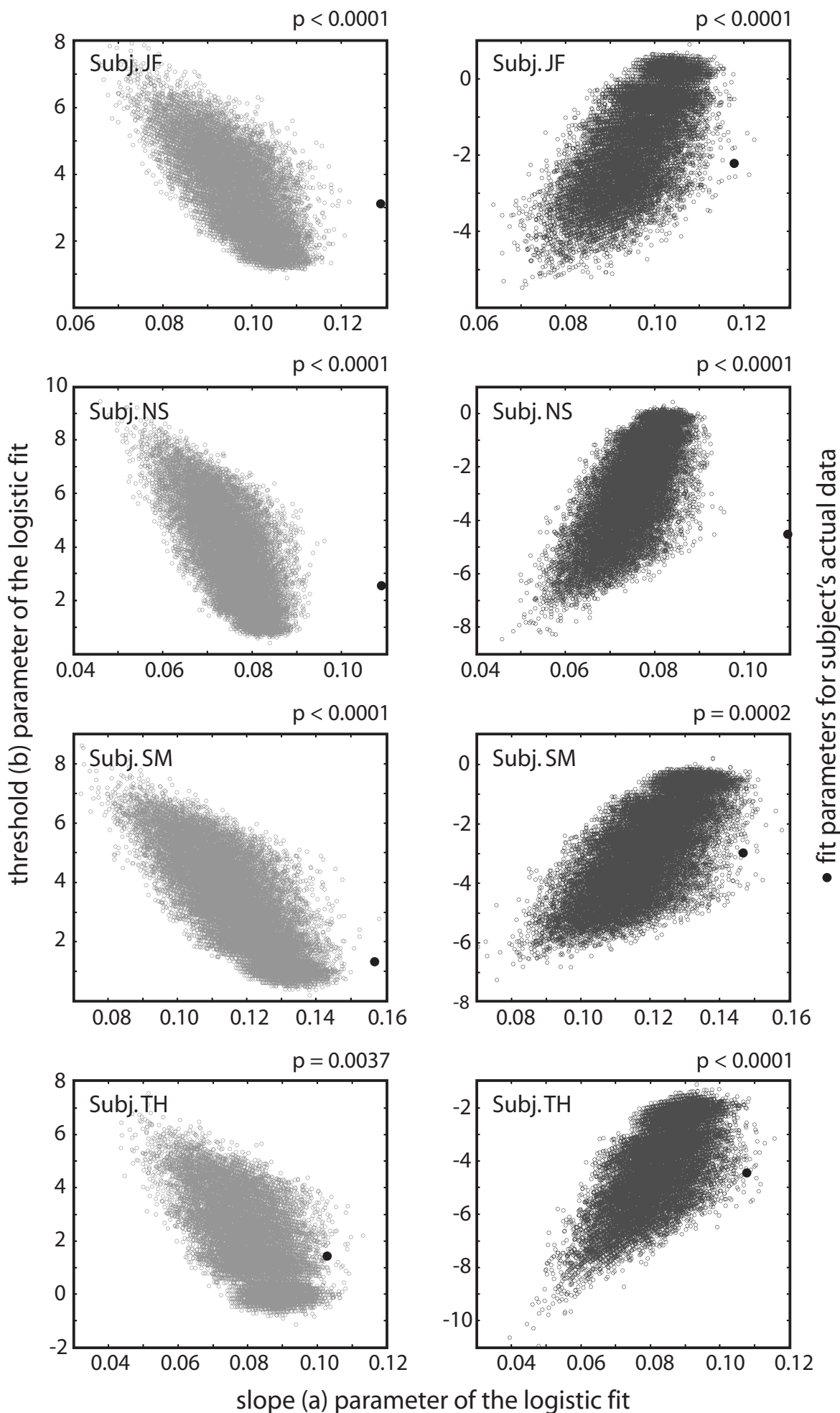


Figure 6

